

# C-test vs. Multiple-choice Cloze Test as Tests of Reading Comprehension in Iranian EFL Context: Learners' Perspective

Parviz Ajideh<sup>1</sup> & Sorayya Mozaffarzadeh<sup>2</sup>

<sup>1</sup> Department of English Language and Literature, University of Tabriz, Iran

<sup>2</sup> Department of English Language and Literature, University of Tabriz, Iran

Correspondence: Sorayya Mozaffarzadeh, Department of English Language and Literature, University of Tabriz, Iran. E-mail: soraya.mozafarzadeh@yahoo.com

Received: August 17, 2012 Accepted: September 25, 2012 Online Published: September 26, 2012

doi:10.5539/elt.v5n11p143 URL: <http://dx.doi.org/10.5539/elt.v5n11p143>

## Abstract

Cloze tests have been widely used for measuring reading comprehension since their introducing to the testing world by Taylor in 1953. But in 1982, Klein-Braley criticized cloze procedure mostly for their deletion and scoring problems. They introduced their newly developed testing procedure, C-test, which was an evolved form of cloze tests without their deficiencies (Klein-Braley, 1982 cited in Baghaei, 2008). After that, the effectiveness of C-test and cloze test became the main interest of the scientists in the field of language testing. The present study aims to compare the results of multiple-choice cloze test with those of C-test as measures of reading comprehension. To this end, one traditional C-test and one fixed ratio (n=7) multiple-choice cloze test were prepared from reading passages with similar readability level. The subjects of the study were 27 female EFL advanced learners. The results of the study revealed that multiple-choice cloze is a better measure of reading comprehension. Through a retrospective study which was done at the end of the tests, the students' impressions and opinions about tests and their own performance were recorded and taken into consideration. The implications of the findings and suggestions for more studies are discussed within a foreign language testing context.

**Keywords:** C-test, cloze test, reading comprehension

## 1. Introduction

### 1.1 Testing Reading Comprehension

The most common of the four skills tested is reading. Testing reading seems to be easy; however, reading as a receptive skill does not usually manifest itself directly in overt behavior. How to assess reading ability in the EFL context in a best way has interested language testing researchers for a long time. Harris (1968) puts across the idea that the same general types of tests which were used to test the reading ability of the Native English have the same effectiveness with the foreign learners of the language. In English as a foreign/second language, reading comprehension tests include a series of related items that are based on the same reading passage (Lee, 2004). These items can be posed after a passage as traditional comprehension questions multiple-choice, short-answer, cloze or c-test which are embedded in the passage itself (Klein-Braley, 1985). So, as Alderson (2000) argues, the selected text and test methods are so effective in testing reading comprehension.

Most of the studies on reading tests (e.g. Phakiti 2003; Atai & Soleimany 2009) show that the choice of text has a marked effect on the test scores. 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).

### 1.2 Basic Methods for Testing Reading

Two areas of applied linguistic theories – reading and testing – come together when testers design a test of reading ability. In such cases, the test designer decides what s/he wants to test, that is, what s/he means by reading ability and finds a means for testing it. Hughes (2003) discusses that in order to elicit reliable behavior from the candidate and have highly reliable scoring, the test designer must consider which ability s/he is interested to measure, so s/he writes the items on the basis of his/her aims. Alderson (2000) points out that there is no ‘best method’ for testing reading comprehension and no single test method to fulfill all the purposes of test. He believes that most of times, different methods act as complementary to each other. As aforementioned, different methods have been used for testing reading ability but here we discuss only the methods which are related to our study. Among significant methods of testing comprehension, one can refer to discrete-point (multiple-choice) and integrative (cloze & C-test) tests.

#### 1.2.1 Multiple-choice Method

Multiple-choice questions are common devices for testing students’ reading comprehension. The candidate provides evidence of his/her successful reading by choosing one out of a number of alternatives. Despite popularity of multiple-choice method, their value and validity are under question. Kobayashi (2002) and Alderson (2000) argue that despite these tests popularity as tests format for assessing reading comprehension in a second/foreign language, they have a significant drawback in that test takers can guess the right answer without fully understanding the reading passage, and thus test validity is questionable. Alderson (2000) writes that the popularity of multiple-choice method is at the expense of validity and “it would be naïve to assume that because a method is widely used it is therefore ‘valid’” (p. 204).

#### 1.2.2 Cloze Test

What is a cloze test? A standard cloze test is a passage with blanks of standard length replacing certain deleted words which students are required to complete by filling in the correct words or their equivalents. The first and the last sentences left intact to provide the examinee with some context. Cloze tests have probably been the most popular kind of tests (Farhady, 1986). Although the idea originated in the early fifties, cloze tests were not utilized as testing instruments until the late sixties and early seventies. The term ‘cloze procedure’ was first developed by Wilson Taylor in 1953 which seems to be a spelling corruption of the word “close”. He explains that the term cloze derived from the Gestalt psychology concept of ‘closure’ (Oller, 1979). The origin of the cloze procedure suggests that at least one of the skills required to ‘cloze’ the gaps created by deleted words is not a language skill at all, but rather a kind of non-verbal reasoning skill, known in Gestalt psychology as ‘closure’ (Mc Kamey, 2006). It describes a tendency that humans have to complete a familiar but not-quite-finished pattern (Lu, 2006). Farhady (1986, p. 30) writes

*In the cloze procedure, the closures are created by deleting certain words from a passage. The examinee, then, is required to fill in the blanks with appropriate words on the basis of contextual clues provided in the passage.*

Cloze procedure is one of the major test forms which makes use of Spolsky’s idea of reduced redundancy (Spolsky, 1969). Spolsky believes that the knowledge of language requires the ability to function 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). Cloze tests reduce natural linguistic redundancies and require the examinee to rely upon organizational constraints to fill the blanks and infer meaning (Mousavi, 1999). In Spolsky’s idea, the more developed competence the learner has, the more able he is in making use of the clues provided by the text to restore a greater number of missed items.

Taylor was the first to study cloze procedure in 1953 for its effectiveness as an instrument for determining the readability level of the texts and then as a device of assessing reading comprehension. During the 1970s, cloze tests began to be used as a measurement of overall L2 proficiency (Ahluwalia, 1992, cited in Lu, 2006). Today, cloze tests are widely used in some places (such as Iran & China) as part of some large-scale language tests (such as TOEFL & IELTS). After Taylor’s introduction of cloze procedure<sup>1</sup>, different types of cloze tests were developed including traditional cloze and discourse cloze tests.

In traditional cloze testing or fixed-ratio method or standard cloze, every n<sup>th</sup> word of a passage is removed and replaced by a standard-length blank space (Oller, 1979). Usually, no word is omitted either in the first or the last sentence of the passage to provide the examinee with some context. This kind of deletion is called random deletion because it deletes every n<sup>th</sup> word consistently, so that all classes and types of words have an equal

chance of being deleted. It is believed that this type of deleting provide an actual sampling of real-life language (Oller, 1979). This type of cloze has widely been focused upon since it retains the original concept of the term cloze itself (Sadeghi, 2010).

A modification of cloze procedure introduced by Bachman in 1985 was discourse cloze or rational cloze used to measure specific linguistic abilities in reading assessment, for example, grammatical features (Lee, 2008). It involves deletion of special words from the passage to include a development of sensitivity to the operation of lexical items in the discourse. In this type of cloze, a specific type of word is deleted according to a linguistic principle, such as nouns, verbs, adjectives, etc. (Lu, 2006). Students engaged in discourse cloze should go back and forth across a developing discourse drawing information from it as a whole and interpret it. It is difficult to complete the text by relying only on syntactic or semantic clues and text-level Knowledge is also needed (Mousavi, 1999). Yamashite (2003) believes that the use of a rational deletion is useful for the reading researchers who desire to measure globule comprehension ability because they require text level understanding while fixed-ratio cloze needs the understanding of clause level and extra-textual knowledge. Alderson (2000) clearly differentiates between these two types of format by calling the rational cloze 'gap-filling tests' and confining the term 'cloze' only to random cloze. He emphasizes that "all other gap-filling tests should not be called 'cloze tests' since they measure different things" (p. 208). However, Bachman (1990) views rational deletion as simply another type of cloze procedure.

### 1.2.3 Multiple-choice Cloze

In fact multiple-choice cloze is the marriage of traditional cloze procedure to multiple-choice test. 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). The answer to this question was positive. The construction of a multiple-choice cloze begins with the construction of a normal cloze test. First an appropriate passage is chosen. The deletion procedure starts from the second sentence. Every  $n^{\text{th}}$  (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.

### 1.3 Cloze Procedure as a Measurement of Reading Comprehension

Since its introduction in 1953, the cloze technique has been used extensively for reading and measurement purposes. One of these purposes is to measure reading comprehension. Traditional reading tests have been criticized because comprehension items are difficult to construct and may misrepresent the author's meaning. Cloze tests in both respects, item construction and avoiding misrepresentation of author's meaning, seem to offer an improved method of measuring reading comprehension (Cranney, 1972).

The use of cloze procedure in testing reading comprehension has given rise to much controversy. Some researchers (e.g. Bachman, 1982) argue that since test takers need to relate various pieces of information from the extra-text environment to fill the blanks, the cloze procedure does not evaluate testee's reading ability. Sadeghi (2008) claims that cloze test is not an appropriate measure of reading comprehension because cloze scores do not reflect the readers' comprehension. He also argues that while other testing methods of reading present the complete text to the reader first and then try to find out if the text has been comprehended, cloze tests appear to be too unfair in that they require the reader to reconstruct something hidden from him/herself, and then to understand the rightly or wrongly reconstructed discourse. However, the effectiveness of fixed-ratio cloze has been supported in L1 research for measuring reading ability, correlating highly with other standardized tests (Hinofotis, 1987, cited in Lee, 2008). Alderson (2000) recommends cloze procedure for reading assessment. Studies by Alderson (2000), Yamashita (2003), Sageghi (2010), Williams, Ari & Santamaria (2011), showed that cloze tests have correlation with other reading test like TOFEL. And finally, Green (2001) claims that the findings of his study provide strong evidence that if cloze tests are designed appropriately, they permit valid assessment of reading comprehension.

### 1.4 C-test

In the late 1970s and early 1980s when the cloze test had become a popular test and well established test of overall language proficiency and reading comprehension, it came under severe attack. In the light of the criticism leveled at cloze test, a modification has been proposed by Klein-Braley and Raatz in 1982. The new testing procedure, called *c-test*, was based on the tenets of cloze test without its deficiencies. In fact the letter C stands for *Cloze* to call to minds the relationship between the two tests (Baghaei, 2008).

The construction of C-test involves a number of short texts (usually five or six) to which the *rule of two* is applied. The reason of including more than one text is to avoid bias from text content. Klein-Braley & Raatz (1984) propose a list of criteria which the new test of reduced redundancy ought to meet:

1. it should use several different texts;
2. it should have at least 100 deletions;
3. adult native speakers should obtain virtually perfect scores;
4. only exact scoring should be possible;
5. the deletions should affect a representative sample of the text;
6. the test should have high reliability and validity.

*The rule of two*: the rule of two or 'C-principle' (Khodadady & Hashemi, 2011) is the defining feature of C-test. According to this rule, the second half of every second word in the text is deleted until the required number of mutilations is reached, leaving the first and last sentence of passage intact to provide enough contexts (Klein-Braley & Raatz, 1984). 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.

The rationale behind the C-test is the reduced redundancy principle. C-tests are claimed to be the best in the family of reduced redundancy tests (cloze, clozentropy, noise test). These tests are developed on the basis of Reduced Redundancy Principle (Spolsky, 1969). This rule suggests that native speakers are able to restore missing or distorted texts by resorting to various textual information and making use of natural redundancy in the text (Khodadady & Hashemi, 2011).

Redundancy is a concept developed as a part of the statistical theory of communication. According to this theory, a message carries information to the extent that it causes a reduction of uncertainties in communication by eliminating certain possibilities. In natural language, more units are used than are theoretically necessary i.e. natural languages are redundant (Spolsky, 1969). Spolsky argues that messages in normal language can be understood without leading to break-down in communication even though a good proportion of them are omitted or masked. Therefore a learner can complete a mutilated text (damaged message) by using the information in the context as in real language communication. In a study Babaii and Ansari (2000), explore whether or not the C-test serves a valid operationalization of the reduced redundancy as it is claimed. The findings showed that C-testing is a reliable and valid procedure that 'mirrors' the reduced redundancy principle.

An immediately appealing feature of C-tests is that they are very economical measurement instruments. They are easy both to design and score and several different texts can be used to make a complete test, which are shorter and contain more deletions than cloze tests. The second advantage of them is that students find it less frustrating than cloze tests (Dörnyei & Katona, 1992). The other feature of C-tests is that they allow highly objective administration and scoring, and generally show high reliability (Eckes & Rudiger, 2006). The other advantage of C-test over cloze is the use of different passages so as to eliminate text specificity and test bias (Baghaei, 2008).

Researchers also report some problems with C-tests. One of them is that test takers with high reading comprehension ability may score very low on the C-test because of a lack of productive skills in the language. Another important problem is the question of what C-test actually measures which has not yet been resolved.

The present study aims to compare the learners' performance on cloze test vs. C-test as measures of reading comprehension. So through this study, the authors try to answer the following research question:

1. Is there any difference between advanced subjects' performance on the C-test and their scores on the cloze test as measures of reading comprehension?

## **2. Methodology**

### *2.1 Subjects*

This study was conducted in Iran Language Institute (ILI), Tabriz Branch. The subjects of the study were 27 female EFL learners with advanced level proficiency. The proficiency levels of the subjects were determined by placement tests of the institute. Most of them were from Tabriz and were speakers of Azeri. The only opportunity for learners to communicate in English is formal classroom interaction. They have no or little opportunity for informal interaction outside the classroom. They have to speak English in the classroom, and they are not allowed to use Azeri or Persian in the classroom.

## 2.2 Data Collection Procedure

### 2.2.1 Data Collection Instruments

Two tests were used in this study. The first test was standard multiple-choice cloze test. The cloze test was served as a counterpart device to be compared with the second test, C-test. The difficulty levels of the tests were in accordance with the proficiency level of the subjects. They were calculated by using Flesch Readability Formula. Also the subjects' opinions about their performance on the tests were recorded after administering the tests.

### 2.2.2 Test Preparation

For preparing the tests, different texts were studied and finally two appropriate ones were selected. 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 cloze test and C-test were 46.3 and 41.4 respectively.

The standard multiple-choice cloze test was developed out of the passages taken from authentic sources (TOFEL) by using a 7<sup>th</sup> deletion random cloze test. The first and second sentences were left intact to yield what Oller and Jonz (1994) call lead-in and lead-out, and the deletions began with the 7<sup>th</sup> word of the second sentence. The difficulty levels of the text used for constructing cloze test was 41.3. The constructed tests yielded 38 items.

The prepared test was piloted two times among the learners similar to the samples. At the first time the test was piloted in the traditional form referred to as the free-repose cloze test. The most frequent incorrect responses written in the pilot was used to construct the distractors of the multiple-choice cloze. The distractors had the same part of speech. In the second time piloting, the malfunctioned distractors were recognized and replaced with suitable ones. Furthermore, the problems with the appearance of the test like font size and spacing were obviated and the required time for test completion was estimated. In the main administration 30 minutes were allocated for each test.

The reason for using multiple-choice cloze was that it provided the possibility of objective scoring. If we wanted to use traditional cloze test we had to score the items by using exact word method requiring the testee to provide the original words deleted from the passage which made the test extremely difficult and frustrating. Therefore, multiple-choice cloze test was used to have objective scores to be compared with scores of C-tests.

The C-test like cloze test was constructed out of the passages extracted from authentic source with approximate difficulty level to the cloze test. The difficulty level of the used text was 43.6. The C-test was constructed by using *rule of two* referred to it as the principle of traditional C-tests developed by Klein-Braley in 1982. That is, it was developed by deleting second half of every second word. If a word had only one letter it was ignored in counting the words and if a word had an odd number of letters, the larger half was removed. The first and last sentences of the texts were left intact and the deletion began from the second word of the second sentence. The constructed C-test had 112 items. The test also was piloted to detect the probable problems with the tests like typographical errors and also to estimate the length of the time needed to complete the tests.

### 2.2.3 Test Administration

The final versions of tests were administered with one week interval. To control the observer effect, the tests were administered by the subjects' own teacher and in their own class time. The allocated time for completing the tests was 30 minutes. Of course, if the time of the class permitted, we extended the allocated time because our intention was covering of all of the items by the subjects.

### 2.2.4 Test Scoring

The used scoring methods for both of the cloze test and C-test were exact word method. This method was objective so that obtained scores were reliable. The multiple-cloze 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 this study we decided to tolerate the spelling problems in the C-test which did not change the meaning and part of speech of the words. If these two happened, the written word would not have scored.

## 3. Findings and Discussion

### 3.1 Descriptive Statistics

After scoring, for the ease of comparison all the scores were calculated out of 100. Descriptive statistics for both cloze test and C-test are represented in Table 1.

Table 1. Descriptive statistics for subjects' performance on the cloze test and C-test

	N	Minimum	Maximum	Mean	Std. Deviation
Cloze test	27	42.10	84.21	65.7237	10.82079
C-test	27	28.57	76.78	49.8300	12.20135
Valid N (listwise)	27				

SD= Standard Deviation

As table one presents, the mean score of the subjects in cloze test (65.72) is higher than the mean score of them in the C-test (49.83).

### 3.2 The Result of Retrospective Study

Furthermore a retrospective study was done at the end of the tests. What follows are some students' perspectives on the above mentioned tests:

1. The multiple-choice cloze is easier to take than the C-test.
2. The higher number of deletions in the C-test makes the process of comprehension difficult.
3. If C-test is assumed to be reading comprehension test, more time will be needed.
4. One of the reasons of failing to complete the deletion is that there is more than one word which begins with the same letters. When we want to find the appropriate word we cannot choose among words.
5. The subjects claimed that there is a chance of more than 50% guessing probability in both of the tests.

The phenomenon which happens in number four is explained in psycholinguistics by **cohort model of lexical access** (Marslen-Wilson and Tyler 1980, cited in Fernández & Cairns (2010)). According to this model, a word's **cohort** consists of lexical items that share an initial sequence of phoneme (letters in written form). Lexical entries that match the stimulus phonologically are activated. After receiving the first syllable or letters of a word, all the lexical entries in its cohort will be activated. To deactivate the other words, which mismatch the stimulus, the remaining letters of the word should be received; however in the C-test processing, only the first half of the word is presented and the testee should find the remaining part by using the context. Therefore, in completing deletions, the cohort of the word intervenes in the process of finding words. The testee should recognize the suitable word on the basis of context, but the subjects in the study believed the existence of more deletions in the C-test prevent them from understanding the context.

## 4. Conclusion and Implications

The aim of this study was to answer the question whether there is any difference between advanced subjects' performance on the C-test and their scores on the cloze test as measures of reading comprehension. The general conclusion and answer which can be drawn from the findings of this study is that despite the widely held view that c-test works better than the cloze test, it was shown that subjects performed better on the cloze test as a measure of reading comprehension. The result of the retrospective study confirmed the findings of the experimental method of data analysis.

### Reference

- Alderson, J. C. (1990). Testing reading comprehension skills (Part two): Getting students to talk about taking a reading test (A Pilot Study). *Reading in a Foreign Language*, 7(1), 465-503.
- Alderson, J. C. (2000). *Assessing reading*. Cambridge, Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511732935>
- 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. (2000). The C-test: a valid operationalization of reduced redundancy principle? *Elsevier Science Ltd.*, 29, 209-219.
- Bachman, L. F. (1982). The trait structure of cloze test scores. *TESOL Quarterly*, 16(1), 61-68. <http://dx.doi.org/10.2307/3586563>
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Baghaei, P. (2008). The effects of the rhetorical organization of texts on the C-test construction: A Rasch Model

- study. *Melbourne Papers in Language and Testing*, 13(2), 32-51.
- Cranney, A. G. (1972). The construction of two types of cloze reading tests for college students. *Journal of Literacy Research*, 5(1), 60-64. <http://dx.doi.org/10.1080/10862967209547022>
- Day, R. R. (1994). Selecting a passage for the EFL reading class. *Forum*, 32(1). Retrieved from <http://eca.state.gov/forum/vols/vol32/no1/p20.htm>
- Dörnyei, Z., & Katona, L. (1992). Validation of the C-test amongst Hungarian EFL learners. *Language Testing*, 9, 187-206.
- Eckes, T., & Rudiger, G. (2006). A closer look at the construct validity of C-test. *Language Testing*, 23, 290-325. <http://dx.doi.org/10.1191/0265532206lt330oa>
- Farhady, H. (1986). Fundamental concepts in language testing (5) theories. *Roshd Foreign Language teaching Journal*, 2(4), 28-37.
- Fernández, E. M., & Cairns, H. S. (2010). *Fundamentals of psycholinguistics*. New York, NY: John Wiley & Sons.
- Greene, B. B. (2001). Testing reading comprehension of theoretical discourse with cloze. *Journal of research in Reading*, 24(1), 82-98.
- Harris, D. P. (1968). *Testing English as a second language*. New York, NY: McGraw Hill.
- Hughes, A. (2003). *Testing for language Teachers* (2<sup>nd</sup> ed.). UK: Cambridge University Press.
- Jafarpur, A. (1999). What's magical about the rule of two for constructing C-test? *RELC Journal*, 30(2), 86-100. <http://dx.doi.org/10.1177/003368829903000205>
- Khodadadi, 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. (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., & Rattz, U. (1984). A Survey of research on the C-test. *Language Testing*, 1(2), 134-146.
- Lee, S. H. (2008). Beyond reading and proficiency assessment: The rational cloze procedure as stimulus for integrated reading, writing and teacher-student interaction in ESL. *Elsevier (system)*, 36, 642-660.
- 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. <http://dx.doi.org/10.1191/0265532204lt260oa>
- Lu, G. (2006). Cloze tests and reading strategies in English language taching in China. Unpublished, MA Thesis. University of Western Cape, China.
- McMamey, T. (2006). Getting closer on the cloze: A validity study of the 'rational deletion' method. *Second Language Studies*, 24(2), 114-164.
- Mousavi, S. A. (1999). *A Dictionary of Language Testing* (2<sup>nd</sup> ed.). Tehran: Rahnama Publications.
- Oller, J. W., Jr., & Jonz, Jon. (1994). *Cloze and coherence*. Cranbury, New Jersey: Bucknell University Press.
- Oller, J. W. (1979). *Language tests at school*. New York, NY: Longman.
- 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. (2008). Measuring reading comprehension: The judgmental validity of cloze procedure. *IJAL*, 11(2), 115-132.
- Sadeghi, K. (2010). Cloze validation against IELTS reading paper: Doubts on correlational validation. *Journal of English Teaching and Learning*, 53(217).
- Spolsky, B. (1969). Reduced redundancy as a language testing tool. Presented at Second International Congress of Applied Linguistics, Cambridge, England, September 8-12.
- Williams, R. S., Ari, O., & Santamaria, C. N. (2011). Measuring college students' reading comprehension ability using cloze tests. *Journal of Research in Reading*, 34(2), 215-231. <http://dx.doi.org/10.1111/j.1467-9817.2009.01422.x>

- Yamashta, J. (2003). Process of taking a gap-filling test: Comparison of skilled and less skilled EFL readers. *Language testing*, 20(3), 267-293.
- Zamanian, M., & Heydari, P. (2012). Readability of texts: State of the art. *Theory and Practice in Language Studies*, 2(1), 43-53. <http://dx.doi.org/10.4304/tpls.2.1.43-53>

**Notes**

Note 1. The terms 'cloze procedure' and 'cloze test' are closely related. Sometimes they are used interchangeably. The difference is that cloze procedure is more general, demarking the use of an activity that follows 'cloze procedure' while a cloze test is a specific application of the cloze procedure to the testing situation (McKamey, 2006).