On the comparability of C-test and Cloze: A verbal protocol approach

Mohammad Salehi *, Hemaseh Bagheri Sanjareh

Sharif University of Technology, Tehran, Iran

1

Abstract

Research on C-test and cloze test has given us few accounts of the real mental processes test-takers are involved in (Rahimi & Saadat, 2005). This will in fact reveal the true nature of what these two tests measure since "in validating (language) tests we also have to analyze the mental processes in the test-taking subject" (Grotjahn, 1986, p. 162). The purpose of this study, therefore, was to explore and compare the extent to which the C-test and cloze test, as measures of the reduced redundancy principle, tap macro-level and micro-level strategies. In doing so, a C-test and a cloze test developed by Jafarpour (1999) and (1995) respectively were used with 9 engineering-major BS students. They all participated in introspective methods of think-aloud and retrospective interviews during and after the test administration phase. The analyses revealed that C-test and cloze test trigger both types of strategies although macro-level strategies were less elicited compared to their micro-level in the former and the other way round in the latter.

Key words: The C-test; The cloze test; The reduced redundancy principle; Verbal protocol analysis; Micro/Macro-level processing, Think-aloud.

* Corresponding author: Tel.: +98 021 66164732.

 Email address: m_salehi@sharif.ir

1. **Introduction**

C-test and Cloze test, being based on the same construct theory of reduced redundancy principle (RRP), have been long contrasted and compared (e.g. Klein-Braley, 1997; Jafarpour, 1995; Eckes & Grotjahn, 2006). This is also owing to their widespread use in numerous institutions round the world as well as their ease of development, administration and scoring. Therefore, these two tests and the reliabilities and validities of their scores have been investigated from different aspects and a myriad of methods. This study seeks to examine the validity of these tests employing the think-aloud verbal protocol approach (VPA). The findings would further illustrate the nature of the tests in question and how VPA can replace quantitative investigations by a differently insightful attitude to test validation.

2. **Review of the literature**

Being initially proposed as a readability measure, i.e. tapping the reading ability of English natives, by Wilson Taylor (Jafarpour, 1995), cloze test appeared as one of the most relied-upon tests in language testing via Spolsky's 'post-modern' trend (Klein-Braley & Raatz, 1984). In his viewpoint, "knowing a language involves the ability to understand a distorted message, to accept a message with reduced redundancy" (Spolsky, 1968, p. 10). By RRP, which is considered as one of the features of natural languages, he meant that a distorted message could still be comprehended even though a good proportion of its words or linguistic elements were omitted. Differential performance of natives and non-natives on tests of RRP led Spolsky to conclude that it could be a promising measure in language testing. Klein-Braley and Raatz (1984) argue that the underlying assumption is that when presented with a mutilated text, learners employ their competence to reconstruct the text and that their score is reflective of this abstract construct. Based on RRP, three testing techniques are proposed: cloze, clozenthropy and the noise test (Spolsky, 1969).

The procedure of a cloze test development initially included the systematic deletion of words in a text, i.e. every nth word, to provide the test-taker with a random sample of language. This was apparently appealing to the proponents of holistic approach to language acquisition and assessment. In terms of scoring methods, two types of exact and acceptable were applied in which in the former the exact word in the original text was required whereas any syntactically

and semantically correct word gained a point in the latter (Klein-Braley, 1997). Yet, cloze test is subject to the following flaws (Klein-Braley & Raatz, 1984; Klein-Braley, 1997):

1. Systematic *n*th word deletion does not necessarily guarantee the production of a random sample of the elements of the text.

2. When applied to the same text, different deletion rates and starting points generate tests which may differ substantially in difficulty, reliability and validity.

3. Cloze tests are likely to have deficient reliability and validity coefficients in particular for homogenous samples.

4. The two common scoring methods are problematic in that scorers do not reach a consensus on the acceptability of the responses offered in 'acceptable scoring method', while the use of 'exact scoring method' renders these tests highly difficult and frustrating.

5. To ascertain a satisfactory level of reliability, extremely long texts are required due to the fact that deletion rates used in classical cloze tests are too high.

6. If one single text is used, as is the standard principal in cloze testing, the test developer cannot ensure that it is a representative sample of the language. Besides, item bias is very likely in this regard.

7. Adult native speaker barely achieve a perfect score.

8. The use of KR-20 together with other measures of internal consistency is possible in that the test items are unlikely to be independent of each other.

Born out of the deficiencies attributed to cloze, C-testing has been claimed to remedy the defects of its predecessor (Klein-Braley, 1997), " The C in the name C-Test was chosen specifically as an abbreviation of the word 'cloze' in order to indicate the relationship between the two test procedures" (p. 63). Klein-Braley and Raatz proposed it as a means of measuring general proficiency more accurately through replacing the one-long-text procedure of cloze text with several shorter passages in order to eliminate the impact of topic familiarity as well as controlling the content bias. Moreover, in C-test deletions are at the word level whereas they are at the text level in the cloze. C-test normally comprises five short authentic passages in which the first and the last sentences remain intact while beginning from word two of sentence two, the second half of every other word is mutilated. Nevertheless, this 'rule-of two' is not applied to

one-letter words. In case of having an odd number of letters, the larger half of the word is deleted. To establish objectivity, only exact scoring should be applied. Another improvement, as Lee-Ellis (2009) asserts, is its practicality, ease of development and quick simple administration and scoring in that, as one of the reasons, it abolishes the time-consuming process of deletion rate.

What is most at issue with respect to C-testing is what it really measures. Several studies have targeted the construct validity of C-test, the results of which have been inconsistent. Correlating strongly with grammatically-based tests, C-tests, according to Chappelle and Abraham (1990), mainly tap the grammatical than textual competence which can be in part due to shorter texts and the presence of clue in the immediate environment. To Jafarpour (1995, p. 209), "C-testing does not achieve the claims made on its behalf" in that the findings of his study question the validity of C-tests revealing the same shortcomings stemming from its ancestor, the cloze test. Similar unclear construct validity has been reported by Grotjahn (1986).

Nonetheless, some other studies undertaken in this area have displayed that C-test is advantageous over cloze, correlating highly with other tests of general proficiency and having satisfactory reliability and validity (e.g. Dornyei & Katona, 1992; Eckes & Grotjahn, 2006; Klein-Braley & Raatz, 1984; Klein-Braley, 1985; Klein-Braley, 1997). Working on Hebrew C-test, Cohen, Segal and Weiss (1984) found it a highly reliable measure which had, also, a remarkably good correlation employing confirmatory factor analysis and Rasch modeling measurement . Based on the results, they concluded that C-test unidimentionally measures the multicompositional nature of general proficiency along with a considerably high reliability. Reliability and validity of C-testing have been empirically supported as well by Babaii and Ansary (2001), who, relying on the findings of verbal protocol analysis (hereafter VPA), argued that C-test tap both micro- and macro-level features of language although micro-level cues appear more frequently. In the same vein, Rahimi and Saadat (2005, p. 78) maintained that, based on retrospective measures, "subjects tend to use the bottom-up strategies quite more frequently than top-down ones for restoring the item" which to a considerable extent relies on the content of the text, the mutilated lexical items and subjects' proficiency levels. It is, nevertheless, noteworthy that test tasks can play a crucial role in triggering more macro-level processing as

confirmed by Babaii and Jalali Moghaddam (2006). Using two sets of C- tests with the same content but different formats along with employing verbal protocol analysis, they assert that texts which are syntactically complex, and also in which the number of missing letters in mutilated words are deleted, hindering access to textual clues, produce more difficult tasks. These two factors are consequently crucial in encouraging more macro-level strategies.

However, blemished by lack of face validity, C-test is criticized for not looking decent enough as an appropriate tool for language proficiency. Majority of subjects in Jafarpour's (1995) study, for instance, expressed negative views about the appearance of this test in the provided questionnaire. Although the importance of face validity, as a 'test appeal' to the researcher and test-takers, has been seriously questioned (Bachman, 1990), this apparent drawback might cause the subjects not to take the test seriously. By the same token, "they are less likely to perform to their best of their abilities. This has obvious implications for test validity. If test attitudes interfere with test performance this may result in unwarranted inferences being drawn from test scores" (Elder, Iwashita, & McNamara, 2002, p. 350). Similarly, Bachman (1990) concludes that test appearance must be considered crucial in test use.

3.   Methodology

3.1. Participants

The participants in this study were 9 students majoring in different engineering fields at Sharif University of Technology. All of these students were Persian native speakers and they were in the age range of 18 and 22 years old. They were attending two intact classes (10, 8 students in class one and two respectively) and their homogeneity in terms of language proficiency was established through their scores on the C-test and cloze passage administered as research instruments. Moreover, prior to the study, these students had received at least six years of formal English instruction. Therefore, the sample appears to have homogeneity with respect to age, English language background, L1 background and educational level. The total sample participated both in the test administration phase of the study and the think-aloud verbal protocols.

3.2. Instrumentation

The instruments employed in this study were the C-test and the cloze passage developed and utilized by Jafarpour (1999) and (1995) respectively.

2.2.1. The C-test

Observing the instructions by Klein-Braley (1997), Jafarpour (1999) constructed the C-test whose five selected texts were all authentic and self-contained and that varied in terms of subject matter. The texts have Flesch Reading Ease readability values of 89, 75, 71, 70 and 56, respectively with total number of 126 mutilations.

2.2.2. Cloze test

Jafarpour (1995) constructed the test deleting every sixth word which resulted in a total of 25 deletions. He also left the first two sentences and the last one intact.

2.3. Procedure

The C-test and cloze test were administered to all the subjects individually. Prior to test administration, they received a full explanation about the procedure and what the researchers expected. For more preparation, they were also asked to verbalize their thoughts on an example C-test. This was conducted since "in validating (language) tests we also have to analyze the mental processes in the test-taking subject" (Grotjahn, 1986, p. 162). Since Storey (1997) attributed lack of verbalization of some parts in his study to the use of second language, the think-aloud and retrospective protocols of the current research were obtained in the subjects' mother tongue so that they would face less difficulty and constraints verbalizing their thoughts.

2.4. Verbal protocol analysis

Following the guidelines provided by Camps (2003) and Ericsson and Simon (1984), verbal protocol reports were collected from 9 students individually. The think-aloud protocols were obtained concurrently, that is, the subjects verbalized whatever they thought all through the test-taking procedure. The researchers tried to make the interview sessions as least mediated and

intrusive as possible. Only when the test-takers remained silent they were reminded to verbalize their thoughts. In the meantime, students' statements were recorded and transcribed for analyses which embraced the coding phase whereby the strategies proposed by Feldman and Stemmer (1987, as cited in Rahimi & Saadat, 2005) were coded in the transcribed data.

4. **Results**

Feldmann and Stemmer (ibid) have proposed 13 strategies categorized under two broad methods of top-down and bottom-up processes. However, since making a clear-cut distinction between these two types of strategies, due to the fuzzy nature of them, is not possible, it is more plausible to consider them as two extremes of a continuum. Storey (1997) asserts that macro strategies are those employing materials outside the sentence to restore the mutilated word, whereas micro strategies require reference to the immediate context, to the original text itself. Thus, taking the classifications made by these two scholars into account, the micro strategies are as follows:

1. Structural analysis

a) Syntactic analysis

To retrieve the word, the subject analyzes the sentence syntactically. For instance, in the case of "never fini- - - - a", the subject stated, " I chose *finishes* because of never and also the sentence lacks a verb, the subject in the previous sentence is a person and the tense is present simple".

b) Formal indicators

To guess the word, the test-taker employs a formal syntactic indicator. For instance, in the case of "the cof- - - of Br- -" The subject would choose *Brazil* since it starts with a capital letter. However, this strategy is not included in this study in that none of the participants used it.

2. Adding letters/syllables to the item beginning

The subject completes the mutilated word considering the first undeleted half. For instance, when faced with "his appoint- - - - - -" a subject said it was *appointment* because of the first seven letters. He also counted the number of gaps to ensure his answer was right.

3. Translation to mother language

For example, in case of, "always lea- - - a lit- - - food", the subject translated the phrase into his mother language /meghdari food baghi migozarad/ and when made sure that it matched the meaning of the sentence, he filled the gaps.

4. Using mother tongue meaning equivalent

Employing this strategy, the subject translates the entire sentence, and then attempts to fill the gap by finding a suitable equivalent for that word in the target language. For instance, in "build da- - and ho- - -", a subject said "reading the sentence to the end, I guess it should be something like /sad va khane/ (*dams and houses*) so I wrote them".

5. Using reference

a. Retrieving the word by referring to the same lexical item repeated before.

Since the subject has already seen the word in the test, he/she is able to retrieve it. For instance, in the case of "the che- - - - consists" a subject said, "I wrote the word *checkup* because it was use in the previous sentence".

b. Retrieving the item because it is morphologically related to another item in the text.

The subject fills the gaps by alluding to a related lexical item in the same text. For example, in the case of "…in As- - - houses", a subject justified his choice for *Asian* by referring to the word *Asia* in the next sentence.

c. Retrieving the missing word by substituting a pronoun for a lexical item.

For instance, when faced with "th- - cut", a subject instantly said "I'm sure it is *they* because it refers to beavers in the previous sentence."

Since these three types of "reference" strategy might be employed for the items in the same sentence or preceding and following sentences, they can be included for both micro and macro strategies. This highly relies on the location of the referred item; the nearer the item, it is more

likely to be marked as micro and vice-versa. In this article, therefore, this strategy was generally labeled as "reference (micro)" and "reference (macro)".

The following strategies are among the macro ones:

6. Using past situations

Noting the beginning letters, the subject is able to restore the whole word in that he/she has encountered that expression or phrase in other texts. For instance, in the case of, "thought –f himself –s being" a participant stated, "it is *as being*, I have already heard *as being* a lot".

7. Using the co-text, preceding/following sentence(s)

This strategy helps the students to restore the deleted parts by establishing relations between clauses or sentences. For example, "I_ European coun_ _ _ _ _, even tho_ _ _ shoes some_ _ _ _ _ become ve_ _ muddy, th_ _ is n_ _ done". The subject said, "I chose *even though* because of the sentence structure, that is whenever we have even though, it shows the contrast".

8. Using the general meaning of the text

Based on the general meaning and idea of the text, the subject seeks to restore the mutilated word. Faced with a sentence like: "Before ente_ _ _ _ a ho_ _ _ in so_ _ Asian coun_ _ _ _ _, it i_ good man_ _ _ _ to ta_ _ of yo_ _ shoes". One of the subjects verbalized as follows: "*before entering a house in some Asian countries*,.. at the beginning in the topic sentence it was mentioned that the paragraph focuses on the manners of different countries then it started talking about Asian countries then I guessed that it was going to talk about one of their manners, when I saw shoes and off I realized that it concerns their habit of taking off the shoes…*so it is good manners to take off your shoes*".

9. Using external help

The subject fills the gaps because he/she has already encountered the same word in previous text or in other tests. This strategy was excluded from investigation in the current text due to the fact that none of the learners employed it.

Mohammad Salehi *, Hemaseh Bagheri Sanjareh (2013). On the comparability of C-test and Cloze: A verbal protocol approach. *English for Specific Purposes World, 14*(39).

10. Using background knowledge

The background information the learner has about the topic helps him/her to retrieve the word. For instance, in the case of, "Many o_ these te_ _ _ are do_ _ by mac_ _ _ _", a subject stated, "the last word is *machines* because in hospitals and labs, the tests are done by machines".

11. Using inference

Two types of inference are considered; one is inferring from a lexical item and the other is from a sentence. Since each participant used a few instances of this strategy, both types were counted as one. For example, in the case of, "The che_ _ _ _ consists o_ a ser_ _ _ of te_ _ _ for t_ _ eyes, he_ _ _, lungs, a_ _ so o_". One of the subjects said, "The checkup consists of a series of tests for the *eyes*. Ok, *eyes* and something else plus *lungs*, ok so it could be *heart* because they are all organs of the body".

12. Juxtaposition

Due to the co-occurrence of the words and familiar combination of a set of words in a phrase, the subject can restore the missing parts. In case of "Then th_ _ cut t_ _ trees in_ _ pieces", a test-taker stated, "they *cut the trees into pieces* ... cut into pieces...it's very familiar, have seen it a lot".

Table 1. Micro and Macro Strategies Employed in both C-test and Cloze

	Strategies	Percentages	
		C-test	Cloze
Micro	Syntactic analysis	16.4%	31.6%
Strategies	Adding letters	24.6%	0%
	Translation to mother language	7.2%	7.4%
	Mother language equivalent	2.5%	0.9%
	Using reference	5.8%	7.9%
	Total	**56.5%**	**47.8%**
Macro	Using past situations	4.8%	3.8%
Strategies	Using general meaning	19%	17.5%
	Using reference	1%	5.2%
	Using inference	8.9%	7.4%
	Juxtaposition	4.1%	1.3%
	Using background information	1.8%	2%
	Using co-text	6.6%	14.3%
	Total	**45.2%**	**51.5%**

As displayed in Table 1, micro-strategies used in C-test outnumber those of cloze. On the other hand, cloze test elicited more macro-strategies comparatively. In C-test, the micro-strategy applied most was *adding letters* where subjects counted the number of gaps and highly relied on the first half of the mutilated words to restore them. This strategy was mostly relied upon probably because in many cases it was too much of a guidance making it considerably easy for

test-takers to guess the words. This strategy, however, was not employed at all in cloze since the structure of the test, obviated the need for it.

Syntactic analysis is the most used micro-strategy in cloze coming second in C-test. It has also attained the highest percentage among all strategies both micro and macro ones. This is indicative of the significance this strategy has for the test-takers and the crucial role it plays in this regard. With respect to macro-strategies, resorting to general meaning while filling the gaps was by far the most popular strategy in both C-test and cloze possessing a slightly higher percentage in C-test. This reflects that comprehending the text and the message it attempts to convey enables the subjects to form meaningful associations between the clauses, sentences and even paragraphs.

5.   Discussion

The findings of C-test in this study lent support to those of Babaii and Ansary (2001) as well as Rahimi and Saadat (2005) in that C-test taking involves both micro and macro processes even though micro-level processes comparatively appear more frequently. However, unlike the two cited studies, this research embraced a VPA on a cloze test too, rendering these tests comparable in terms of the processes they involve and consequently their construct validity. Cloze test was also shown to require both micro and macro processes although, in this case, macro processes outnumbered. This is well in line with the principle of reduced redundancy stressing the role of both global and local knowledge to restore mutilated words in a distorted text (Babaii & Ansary, 2001).

The results do not thoroughly concord with Dornyei and Katona's (1992) claims who argued that the vulnerability of C-test lied in testing of grammar since, as displayed in the table, the strategy of syntactic analysis was employed to a notable degree in both of these tests. Moreover, in sharp contrast to Jafarpour (1995, p. 209) who concluded that "C-tests are irritating and unacceptable to the subjects", the retrospective protocols obtained from subjects after taking each test revealed otherwise. Some of their excerpts about C-tests are as follows:

"I: Does this type of test irritate the test taker?

S1: I personally enjoyed it very much coz when I could fill the rest of the words it gave me a good feeling and also energy to do the rest, no I wasn't irritated".

"S2: was interesting, C-test was interesting coz it's the first time I've seen such a test

I: Was it only interesting because you've never seen such a test?

S2: yes and also it was interesting for me that how the words came to my mind just by seeing the beginning letters of them coz you don't have any presupposition but by seeing some letters you remember the whole word".

S5: "Not only should you know enough vocabulary but you should have also heard the common expressions and phrases to be able to answer well. And also you should have a general idea and concept of that topic and text. That is you more need the knowledge of voc than grammar….was like a game..but this one[cloze] is irritating".

Though majority of subjects approved of the test, they considered the first half of the words as too much of a help making it very easy for the test-taker to restore the word, as can be noted below:

I: In C-test, to fill the gaps did you rely mostly on the beginning letters of the gaps or on the meaning of the whole sentence and paragraph?

"Well, all these are involved while answering..but for some words, yes perhaps be coz they are constantly repeated or emphasized in our text books and we are very familiar with them so as soon as we see the beginning letters we can guess the whole word".

"In c-test the beginning of the missed words itself helped to guess the word"

This is in fact in close conformity with the principles if reduced redundancy as stated by Spolsky (1968, 1969).

Moreover, when asked about cloze test and its comparability with C-test, they had contradictory ideas in that some stated that cloze was a harder test because they had to fill the gaps requiring a productive skill whereas C-test was easier, more 'passive' providing the initial letters and limiting their choices:

"Cloze is harder than C-test though be coz of containing mutilated words, it can be irritating sometimes. C-test was easier because the beginning letters were given"

On the other hand, some were under the impression that the very same characteristics of C-test made it harder and more challenging since in cloze test they could fill the gaps with any word they thought was suitable:

"Cloze is easier than the other one coz you can put any word you like".

Almost all students admitted, as also supported by the results in the table above, that background information and topic familiarity was helpful and one of the much relied upon strategies in both tests.

6. **Conclusion**

It can be concluded that contrary to what is claimed (Klein-Braley & Raatz, 1984; Klein-Braley, 1997) C-test and cloze are not completely distinct from each other though it may not seem so on the surface. Moreover, C-test holds satisfactory face validity from the viewpoint of majority of subjects in this study.

Despite the fact that a certain amount of criticism has been leveled at C-test and cloze test in terms of validity (Grotjahn, 1986; Jafarpour, 1995), these tests are appropriate for the purposes they are purported for. Taking into account that both these two tests engage to slight differing degrees both high-level and low level mental processes and linguistic knowledge, employing a combination of two as a language test can reveal a more comprehensive picture of the learners' linguistic ability.

7. **Implications of the Study**

VPA is highly recommended for language test developers to observe the test-takers' performance and obtain insights into those areas they find too easy or too challenging. This awareness of the mental processes they undergo aids test developers and researchers to construct quality tests which more accurately target the objectives.

With respect to the precautions taken to obtain the most reliable data from VPA, one needs to well familiarize participants with the procedure by having some training in advance. Besides, the recording session must be held as least stressful and intrusive as possible.

8. **(De) limitations of the Study**

Some (de) limitations are pertinent to the current study:

1. Even though gender can exert influence on the choice of test-taking strategies, this was not taken into account.

2. Field of study is another contributing variable which was controlled in this study. The investigation of this variable could have been insightful.

3. Selective choice of participants yielded a homogenous sample, which besides the benefits it enjoyed, deprived the researchers from exploring the impact of proficiency on strategies subjects employed.

4. The results should be treated with circumspection due to paucity of participants, which is in fact one of the inevitable features of qualitative studies (Dornyei, 2007), since this does not allow thorough generalizability.

Mohammad Salehi *, Hemaseh Bagheri Sanjareh (2013). On the comparability of C-test and Cloze: A verbal protocol approach. *English for Specific Purposes World, 14*(39).

Reference

Babaii, E., & Ansary, H. (2001). The C-test: A valid operationalization of reduced redundancy principle? *System, 29*, 209-219.

Babaii, E., & Jalali Moghaddam, M. (2006). On the interplay between test task difficulty and macro-level processing in the C-test. *System, 34*, 586-600.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Chappelle, C., & Abraham, R. (1990). Cloze method: What difference does it make? *Language Testing, 9,* 187-206.

Cohen, A., Segal, M., & Weiss, R. (1984). The C-test in Hebrew. *Language Testing, 1,* 221-225.

Dornyei, Z. (2007). *Research methods in applied linguistics.* Oxford: Oxford University Press.

Dornyei, Z., & Katona, L. (1992). Validation of C-testing among Hungarian EFL learners. *Language Testing, 9*, 187-206.

Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, *23*(3), 290-325.

Elder, C., Iwashita, N., & McNamara, T. (2002). Estimating the difficulty of oral proficiency tasks: What does the test taker have to offer? *Language Testing, 19*(4), 347-368.

Mohammad Salehi *, Hemaseh Bagheri Sanjareh (2013). On the comparability of C-test and Cloze: A verbal protocol approach. *English for Specific Purposes World, 14*(39).

Ericsson, K., & Simon, H. (1984). *Protocol analysis: Verbal reports as data.* Cambridge: Cambridge University Press.

Grotjahn, R. (1986). Test validation and cognitive psychology: Some methodological considerations. *Language Testing, 3,* 159-186.

Jafarpour, A. (1995). Is C-test superior to cloze? *Language Testing, 12,* 194-216.

Jafarpour, A. (1999). Can the C-test be improved with classical item analysis? *System, 27*, 79-89.

Klein-Braley, C. (1985). A cloze-up on the C-Test: A study in the construct validation of authentic tests. *Language Testing, 2,* 76-104.

Klein-Braley, C. (1997). C-tests in the context of reduced redundancy testing: An appraisal. *Language Testing, 14*(1), 47-84.

Klein-Braley, C., & Raatz, U. (1984). A survey of research on the C-test. *Language Testing, 1,* 134- 146.

Lee-Ellis, S. (2009). The development and validation of a Korean C-test using Rasch analysis. *Language Testing, 26*(2), 245-274.

Rahimi, M., & Saadat, M. (2005). A verbal protocol analysis of a C-test. *International Journal of Applied Linguistics, 8*(2), 55-86.

Spolsky, B. (1968). *What does it mean to know a language, or how do you get someone to perform his competence?* Paper presented at the second conference on Problems in Foreign Language Testing, University of South California.

Spolsky, B. (1969). *Reduced redundancy as a language testing tool.* Paper presented to the Language Testing section of the second International Congress of Applied Linguistics, Cambridge, England.

Storey, P. (1997). Examining the test-taking process: A cognitive perspective on the discourse cloze test. *Language Testing, 14*(2), 214-231.

Mohammad Salehi *, Hemaseh Bagheri Sanjareh (2013). On the comparability of C-test and Cloze: A verbal protocol approach. *English for Specific Purposes World, 14*(39).