Journal of Academic and Applied Studies (JAAS)
Vol. 2(1) Jan 2012, pp. 8-31
Available online @ www.academians.org
ISSN1925-931X

A Comparative Study of the C-Test and the NC-Test with Iranian EFL Students

Farideh. Hosseini[1], Nasser. Hassanzadeh[2] , Kamal. Shayegh[3]

[1]*Department of ELT, Azerbaijan University of Tarbiat Moallem, Tabriz, Iran*

[2]*Department of ELT, Tabriz Branch, Islamic Azad University, Tabriz, Iran*

[3]*Department of ELT and General Linguistics, Ahar Branch, Islamic Azad University, Ahar, Iran*

Abstract
This research aimed at comparing the C-Test with the NC-Test in measuring the overall English language proficiency of Iranian EFL students. It also investigated the test-taking strategies based on Sasaki's categorization of cloze test-taking strategies utilizing the semi-structured individual interviews to realize how these target students responded to each kind of the tests. Both kinds of the tests using identical texts consisted of 5 small texts with 100 deleted words. A total number of 40 Iranian EFL students participated in this study. Each participant took one kind of the cloze test and subsequently a semi-structured individual interview was conducted to 4 volunteer students to learn what happened in the mind of the testees while restoring the test items. The results of this research showed that there were not any statistically significant differences at the 0.05 level of significance between the C-Test and the NC-Test in measuring the English language proficiency of the students in both high- and low-language-ability groups. Hence, the NC-Test may be used as an alternative to the C-Test in measuring the English language proficiency of EFL students. The findings also affirmed that using third-word deletion technique influenced the discrimination power of the NC-Test. It also came out that both C-Test and NC-Test could consistently classify the subjects in their appropriate proficiency levels − elementary or advanced. The cloze test-taking strategies frequently used by volunteer subjects were Within Clause; Across Clause, Within Sentence; Extratextual; Guessing; and Missing.

Key words: C-Test, NC-Test, test-taking strategies, language proficiency, discrimination power.

I. Introduction

A. Cloze Test

Cloze test is a well-known and widely-utilized integrative language test.Taylor in 1953 defined cloze procedure as "any passage of appropriate length and difficulty with every nth word deleted" (cited in Farhady et al., 1994, p.279). Farhady et al. (1994, p. 283) defined cloze as "a passage of appropriate difficulty (determined by readability formulas), and of appropriate length (220-250 words) with every seventh word deleted". Farhady (1996, p. 38) contended that "cloze is a passage of about 240 words with every seventh word deleted providing that the first and the last sentences are left intact". There are many definitions for the 'cloze procedure' by different researchers and scholars, but all lead to a unitary definition that it is a passage in which some words are deleted intentionally, systematically or randomly, that requires students to fill in the blanks by guessing the missing words in the text.

Alderson and Klein-Braley (cited in Klein-Braley, 1997; p. 59-60) state that empirical findings show that the cloze is technically problematic with regard to issues like:

1. The deletion rates used in classical cloze tests are too high. If 50 items are demanded in order to ensure a reasonable level of reliability, then extremely long texts are necessary.

2. If one uses only one text, as is standard practice in cloze testing, then one cannot assume that this is representative sample of the language. In addition, item bias is possible as a result of text content.

3. The two different scoring methods for cloze test are problematical. If the exact method is used, cloze tests are too difficult even for competent adult native speakers (of. Klein-Braley, 1982). But the acceptable scoring method is far from objective and is extremely time-consuming. Even more time-consuming is the clozentropy method advocated by Darnell (1968).

4. It seems intuitively reasonable that an adult native speaker should make very high scores on a test intended for learners of that language. This does not happen with cloze tests (of. Klein-Braley, 1982).

5. Calculation of reliability coefficients using item statistics (e.g. KR-20) is theoretically unsound since this statistical approach assumes item independence. Getting one item right should not depend on getting another item right. But items in cloze tests are obviously textually interdependent.

6. The high reliability and validity coefficients found in many of the research studies were partly a result of extremely heterogeneous subject groups involved in the investigations. For the highly homogeneous groups of the Duisburg placement procedures, low reliability and validity coefficients were more typical.

Besides, other researchers "see cloze as only measuring the ability to make localized connections in the text" (Alderson, 1983; Bachman, 1985; cited in Storey, 1997, p. 214). Alderson (1983) demonstrates that cloze is not a method that always guarantees equivalent results because different starting points or deletion rates on a given text as well as the degree of text difficulty influence the results and affect reliability and validity coefficients.

Raatz and Klein-Braley in 1981 introduced the C-Test as a modified form of the cloze test in view of all the criticisms made against the cloze procedure. Klein-Braley (1997, p. 63) stated that "the C-Test was an attempt to retain the positive aspects of cloze tests but to remedy their technical defects."

B. The C-Test

"The C-Test" invented by Raatz and Klein-Braley in 1981 (cited in Rungruangthum, 2005) is based on the same principle as the cloze test. The letter C actually infers the relationship between the C-Test and the cloze procedure. C-Tests, like the classic cloze tests, are an operationalization of the principle of reduced redundancy testing (Klein-Braley, 1997).

The C-Test consists of four or five short texts on a variety of topics in each of which the first and the last sentences are left intact, then the C-principle (or the rule of two) is applied: the second half of every second word is deleted, beginning with the second word of the second sentence as can be seen in Appendix A. The students' task is to restore the missing parts (Klein-Braley, 1997).

The original C-Test was constructed as a way of testing English language proficiency besides using cloze tests. However, in the C-Test, the second half of each word must be deleted if the deleted word contains an even number of letters. If a word has an odd number of letters, the 'larger' half is omitted. Numbers, proper names, abbreviations, and one-letter words such as 'I' are ignored in the counting. In the conventional C-Test each text will have either 20 or 25 blanks. Only entirely correct restorations are counted as correct , i.e., spelling problems are considered errors. The time needed to administer a C-Test composed of five texts with 20 blanks each is less than 30 minutes, and scoring takes around 1-2 minutes per text and participant, that is, the testees would have roughly five minutes to answer each text, so that a test with five parts would take twenty five minutes to complete. Klein-Braley and Raatz (1984; p. 136) proposed the following criteria for making the C-Test:

1. The C-Test should be much shorter and should have at least 100 items.
2. The deletion rates and the starting points of deletion should be fixed.
3. Only exact word scoring method should be employed.
4. The C-Test should contain various passages.
5. The words affected by the deletion should be a representative sample of the text.
6. Adult educated native speakers should make perfect scores on the C-Test.
7. The test should have high reliability and validity.

On one hand, some researchers have found the C-Test a highly integrative, reliable and valid measure of overall language ability and more effective and more reliable than the traditional cloze tests in assessing the students' language proficiency and easy to construct and to score **(**Klein-Braley and Raatz, 1984; Klein-Braley, 1997; Dörnyei and Katona, 1992; Connelly, 1997; Babaii and Ansary, 2001; Eckes and Grotjahn, 2006; Weir, 1990**)**.

In the original study of the C-Test, Raatz and Klein-Braley (1981; cited in Rungruangthum, 2005) surveyed the use of English and German C-Tests to find out whether the C-Test could be a substitute in weighing the target language. The subjects of the study were divided into two groups. The first group consisted of English native-speaking school children and non-native speakers of English. These students were requested to take the English C-Test. The second group taking the German C-Test consisted of German native-speaking school children at the third grade level and non-native speakers of German. The results expressed that the C-Test had great reliability and validity in measuring the target language of the native and non-native testees.

Klein-Braley (1997) empirically compared the C-Test with a group of other reduced redundancy tests – classical cloze test, cloze elide test, multiple-choice cloze test, and standard dictation. She showed that the C-Test highly correlates with other tests of reduced redundancy and a language proficiency test- DELTA, the Duisburg English Language Test for Advanced Students. She found that the C-Test is the best representative of reduced redundancy tests of general language proficiency. Also, Eckes and Grotjahn (2006) examined the hypothesis that C-tests evaluate general language proficiency. They conveyed that the C-tests measured the same general dimension as reading, listening, writing, and speaking. Eckes and Grotjahn (2006, p.316) also concluded that "C-tests demand the integration of both skills and knowledge: a core ability in all kinds of language use." Eckes and Grotjahn (2006, p.

291) add that "The importance of measuring general language proficiency primarily rests on the utility of the resulting test scores in applied settings". Eckes and Grotjahn (2006) found clear evidence that the C-test is a highly reliable and unidimensional measure of general language proficiency. They found that "lexis and grammar are important components of general language proficiency as measured by C-tests" (p. 316).

Dörnyei and Katona (1992) validated a C-Test against four different language tests including an oral interview, and a TOEIC (the Test of English for International Communication). They reported that their C-Test was a random and representative sample of the original text. They found a high correlation between the C-Test scores and scores of a language proficiency test in the case of Hungarian EFL students. Dörnyei and Katona (1992, p.203) pointed out that "The C-Test proved to be a highly integrative and versatile measuring instrument, working well in samples of various difficulty and homogeneity levels… our conclusion about the C-Test is that not only is it a reliable and valid measure of general language proficiency, but it is also one of the most efficient language testing instruments in terms of the ratio between resources invested and measurement accuracy obtained." Their investigation was conducted in order to validate this type of language test for the EFL students. The subjects of the study were 102 Hungarian university students and 53 Hungarian secondary students. These students were then requested to take the C-Test. The results of this study show that the C-Test is suitable to measure language proficiency of non-native students for whom English is not the first language, although this C-Test was reported to be too difficult for Hungarian students at the secondary level. Dörnyei and Katona (1992, p. 198) found that "in samples of high- vs. low-proficiency examinees content and structure (or function) words behaved very differently in relation to a composite measure of general language proficiency consisting of four different language tests. In the first group, for which the C-Test was fairly easy, the content words were a better measure of general language proficiency than structure words; in the second group, for which the C-Test was fairly difficult, the opposite tendency showed up." Nevertheless, they conveyed that the C-Test was less efficient in testing grammar. Moreover, they consider it a better measure of general language proficiency than the cloze test.

Connelly (1997) supported using the C-Test to measure the general language proficiency of high-level students studying English as a foreign language. His study examined the English C-Test with non-native postgraduate students studying at the Asian Institute of Technology (AIT) in Bangkok, Thailand. The C-Test with 100-deleted items was administered to EFL postgraduate students from six different countries: Thailand, Vietnam, Taiwan, Indonesia, Japan, and Cambodia. The results of this investigation express that the C-Test is highly reliable and has concurrent validity in assessing the language proficiency of English within EFL contexts.

In his study of a validation of the C-Test with Iranian EFL learners, Rouhani (2008) came to the conclusion that the obtained reliability estimates confirmed earlier reports of high reliability coefficients associated with the C-Test. He used a content/function word analysis to investigate the content validity of the C-Test. The C-principle showed a satisfactory method of sampling the linguistic elements in the text; therefore, he claimed that the C-Test enjoys content validity. Rouhani (2008, p. 173-174) adds that "As far as criterion-related validity is concerned, the C-Test scores correlated fairly highly with those of the MTELP. Not only that, but the C-Test's correlation coefficient with the MTELP was higher than with the grammar, vocabulary, and reading comprehension tests. This is to be considered further evidence in favor of the claims that C-Tests measure general language

ability. The C-Test also was capable of fulfilling many of the requirements of a suitable test in terms of construct validity."

Hastings (2002a; cited in Rahimi and Saadat, 2005) reports that a C-Test measures the ability to use and integrate contextual, semantic, syntactic, morphological, lexical, and orthographic information and knowledge related to a particular written language. Moreover, the processing that is required for a successful C-Test performance seems comparable to natural language processing in both length and complexity, and may have much in common with natural language performance. However, he accepts that his study is just an exploratory error analysis of the C-Test and fails to answer definitely what a C-Test measures.

A very significant feature of C-tests is that they are very 'economical' measurement instruments specifically about the time factor which is a very essential criterion for test designers and teachers. 'Objectivity and fast scoring' and the 'high reliability' are two other factors which are met in this format of cloze tests.

On the other hand, some other studies report problems in using the C-Test in measuring the proficiency in the target language. For example, the C-Test does not assess language abilities beyond the sentence level (Sigott and Köberl, 1993), and seems to measure the intelligence quotient (IQ) or spelling ability rather than general language skills (Jafapur, 1995). Some C-Test items, especially the functional words, were reported to have low discrimination power (Cleary, 1988; Jafapur, 1999; Wolter, 2002) and lack of validity (Grotjahn, 1986; Jafapur, 1995). Dörnyei and Katona (1992) add that the C-Test is too difficult for EFL secondary students at the secondary level. Hastings (2002b; cited in Eckes and Grotjahn, 2006, p.292) concluded that "the value of C-Testing as a measure of global proficiency in second language has been demonstrated too many times to be open to dispute." In more specific terms, Hughes (2003, p. 195) referred to the puzzle-like nature of the C-Test as a disadvantage: it is harder to read than a cloze passage, and correct responses can often be found in the surrounding text. Thus the candidate who adopts the right puzzle-solving strategy may be at an advantage over a candidate of similar foreign language ability. Along the same lines, Weir (1990) believes that the face validity of the procedure is low as it is irritating for students to have to process heavily mutilated texts.

Kamimoto (1992) developed two versions of a C-test: the "original" C-test and the "tailored" C-test. He developed the "original" C-test by mutilating the second half of every second word beginning with the second word of the second sentence. The "tailored" C-test was developed by tailoring based on the item analysis on the original C-test. He administered the two versions of the C-Test to two homogeneous groups of junior EFL students. The tailored C-test represented higher item facility and discrimination indexes than those of the original C-test. Kamimoto (1992) believes that the C-Test tends to measure the subjects' vocabulary and grammatical competence and its processing occurs at the micro-level. He relates this fact to the deletion procedure used in the design of the C-Test.

Stemmer (1991) expresses that different results may be obtained from the C-Test depending on individual text characteristics. Moreover, the fact that function words are restored more successfully than content words and that text understanding rarely exceeds the proposition border persuade Stemmer to convey that the current form of the C-Test does not tap general language proficiency.

Jafarpur (1995) found that the rule of two is not a proper tool to obtain a representative sample of the basic elements of a text. He adds that "The Rule-of-Two produces a sizeable number of nonfunctioning items" (p. 197). He was able to show that different deletion starts and deletion ratios produce different tests with different results –which he interpreted as suggestive of the invalidity of the procedure . In addition, he claims that the C-Test is not able to make discrimination among the examinees of different proficiency levels. Moreover, his analysis of his subjects' answers to 10 attitudinal questions after taking the C-Test led him to the conclusion that "C-Tests do not possess face validity" (p. 209). The subjects on the whole believed that the C-Test is more of an IQ test or a test of spelling than a test of overall language ability. They believed it is more like a puzzle and is basically good for children.

In another study, Jafarpur (1999) pre-tested a C-Test consisting 5 texts and 126 items with 146 subjects. Based on a classical item analysis, he removed unsatisfactory items and developed a C-Test version with 100 items and tried it with 60 other subjects. The results indicated that classical item analysis does not improve the psychometric and statistical characteristics of the C-Test. Moreover, Jafarpur (2002; cited in Rouhani, 2008) compared the performance of a C-test and a cloze test against a standardized criterion measure. He found a high correlation between the scores on the C-Test and the English Placement Test and a relatively high correlation between the C-Test and the cloze. Since the scores show high reliability and concurrent validity, he expresses that the C-Test is advantageous over cloze. The results showed that the deletions in a C-Test demonstrate a more comprehensive coverage of different language elements than the cloze test. But he concluded that: (a) the C-Test is not an easily constructed, automatically reliable and valid measure of language competence, (b) the use of the 'rule of two' does not ensure acceptable discrimination power for all items, (c) scoring does not suggest any advantage over the cloze.

C. The NC-Test

Some researchers report that the C-Test is too difficult and less effective for non-native students in the lower levels studying a target language such as English (Dörnyei and Katona, 1992; Cleary, 1988; Connelly, 1997).

Consequently, "Thongsa-nga (1998) adopted the original C-Test to make it suitable for Thai students studying English as a foreign language. Imitating the C-Test construction, Thongsa-nga (1998) proposed "the New C-Test (the NC-Test) by deleting the second half of every third word in order to provide more clues for the non-native test takers"(cited in Rungruangthum, 2005, p. 5) and the students' task is to fill in the deleted parts as can be seen in Appendix B.

According to the investigation of Thongsa-nga (1998; cited in Rungruanthum, 2005), the NC-Test is employed as a proficiency test for non-native Thai students at a secondary school level. The investigation of Thongsa-nga (1998; cited in Rungruanthum, 2005, p.20) "examined the effect of different starting points in the NC-Tests and students' attitudes towards the measurement using these language tests. In this study, the three forms of the NC-Test, with third, fourth, and fifth starting points were administered to 97 Mathayom Suksa six students at Srakaew School. These participants were also requested to answer a research questionnaire about what skills the NC-Test measured-- vocabulary, grammar, reading comprehension, or English language proficiency. Her findings reveal that the NC-Test with the third starting point is the most reliable form for measuring the English language

proficiency of these Thai Mathayomsuksa six students; nonetheless, the majority of these students reported that the NC-Test seemed to measure vocabulary and reading skills. Thangsa-nga (1998) also adds that the different starting points had an influence on the discrimination power of these three forms of the NC-Tests."

D. Test-Taking Strategies

The need for more research on the test-taking process becomes obvious when Messick (1989, p.54) stresses that "individuals performed the same task in different ways and even the same individual might perform in a different manner across items or on different occasions….".

In addition to the construction of new language tests, language teachers should investigate the students' test-taking strategies in order to validate the language test and to examine what language abilities the test can measure (Cohen, 1994, 1998). In order to recognize the test-taking strategies employed by the target students, investigations can be done by observation, performance analysis, questionnaires, and interviews (Cohen, 1994). Test-taking strategies can be defined as "the processes that the test takers make use of in order to produce acceptable answers to questions and tasks, as well as the perceptions that they have about these questions and tasks before, during, and after responding to them" (Cohen, 1998, p. 216). Moreover, the perceptions of language tests and test-taking strategies of the students with high- or low-language ability are different (Sasaki, 2000).

According to Tsagari (1994; cited in Jamil, Abd. Aziz, and Abdul Razak, 2010, p. 107) "since the open-ended questions required students to produce their own answers and use their productive skills, this led them to get into the text to find the most accurate and appropriate information. Leaving the question and returning to it later was another strategy adopted in the open-ended questions. And in Tsagari's (1994) study her analysis revealed that test-takers tended to employ more than one mental processing strategy depending on the individual. This could be seen when one of the test-takers in her study expressed "different strategies or combinations of strategies can be applied in each question in order to obtain a correct answer" (Tsagari, 1994, p.48)". Cohen and Upton (2006) conveyed that more than one strategy is used to respond to a question which they call, strategy "naturally grouped together" (p.48). These "grouped together" strategies were realized in a multiple-choice test and it is interesting to survey whether it is similar in the open-ended format.

Analyzing retrospective verbal protocols of 32 C-Test takers, Babaii and Ansary (2001) found four major types of cues with various frequencies used by the participants to restore the items in the C-Test: automatic processing, lexical adjacency , sentential cues , and top-down cues . They reported that the test takers fully employed macro-level cues to restore the items and concluded that the C-Test taps various aspects of language proficiency to varying degrees and, it is a reliable and valid procedure expressing the reduced redundancy principle. Nonetheless, they maintained that their subjects mostly relied on their grammatical judgments to restore the items.

Feldmann and Stemmer (1987) through concurrent think-aloud methods (i.e. simultaneous introspection according to which the subjects report what is going on in their minds and express the kinds of decisions they make and the kinds of strategies they use while carrying out a task) and retrospective interviews (i.e. recall data in that the test-takers report what went on in their minds and represent how a task or activity was carried out after it occurred) showed that what a C-Test would measure seemed to vary according to the deletion in the test. They found that the subjects used bottom-up and top-down processing depending on the item that was deleted and their own level of proficiency and that a skilled reader would use

both strategies. Feldmann and Stemmer (1987) identified different strategies employed by the subjects while taking the C-Test. They primarily attempted to put these strategies on a continuum ranging from bottom-up to top-down strategies. However, they finally admitted that it was not possible to unambiguously put the strategies used by the learners on such a continuum and that in some cases they even failed to make a clear distinction between a bottom-up and a top-down strategy. These researchers describe some of the strategies used by the subjects as follows: recall by structural analysis, by adding letters/syllables to the item beginning, by repetition, by search for meaning, by looking for external help, by substitution, and recall of past situations.

Also Sasaki (2000) surveyed test-taking strategies by using verbal reports to see how the students responded to the language tests. Sasaki (2000) studied the effects of cultural schemata on students' test-taking processes for cloze tests. The students responded to either a culturally familiar or unfamiliar fixed-ratio cloze passages. The subjects were 60 Japanese EFL students with the equivalent English reading proficiency level who were divided into two groups; each group was required to complete culturally familiar or culturally unfamiliar cloze passages (Sasaki, 2000). The students were asked to report their test-taking strategies to be categorized based on the modified cloze test-taking strategies categorization of Bachman (1985). The results show that "the students who read the modified, culturally familiar, version of the text demonstrated correct understanding of the key terms more often, tried to solve more items and generally understood the text better. This resulted in a better test performance than that of the students who read the original text" (Sasaki, 2000, p. 107). Sasaki (2000, p. 85) adds that "These results also support the claim that cloze tests can measure higher-order processing abilities". Students taking the culturally familiar cloze passage use these three categories of test-taking strategies: 'Within Clause', 'Across Clause, Within Sentence' and 'Extratextual' more frequently than the students reading the unfamiliar cloze passage (Sasaki, 2000). Sasaki (2000, p. 95) categorized test-taking strategies required for successful performance in cloze tests as following:

(1) Within Clause: The examinee uses information provided only by the clause in which an item appears.

(2) Across Clause, Within Sentence: The examinee uses information provided by a broader context than the clause in which an item appears, but a narrower context than the orthographic sentence.

(3) Across Sentences, Within Paragraph: The examinee uses information provided by the broader context of the orthographic paragraph containing an item.

(4) Across Paragraphs, Within Text: The examinee uses information provided by the context of the entire text.

(5) Extra-textual: The examinee uses information that is not provided by the text itself, but which is assumed to be included in the examinees' world knowledge.

(6) Guessing: The examinee guesses at the answer.

(7) Missing: The examinee does not/cannot say anything about his or her test-taking processes, or does not answer the item.

II. **Material and Methods**

A. Participants

The target subjects of the present study were 40 Iranian EFL students (20 elementary level students considered as low-language-ability students and 20 advanced level students considered as high-language-ability students) taking a general English course at the Noor-e-Eram Language Institute in Tabriz in the form of six intact classes- three elementary classes and three advanced classes- based on the placement of the institute were selected. They were native speakers of Turkish and enjoyed different levels of proficiency in English. They were female students aged between 15 to 24.

B. Instruments

To construct the tests, five short passages validated by Rahimi and Saadat (2005) to a sample of 26 Iranian English seniors were selected. The texts were of different levels of difficulty as judged by the Flesch Reading Ease readability scale (Microsoft Word, 1995) and a group of five EFL instructors. They varied in difficulty with Flesch Reading Ease values of 94, 93, 88, 86, and 68, respectively. The texts were arranged from the easiest to the most difficult one.

C. Procedure

The selected five texts were used to construct the C-Test, and the NC-Test. Each prepared test comprised of 100 deleted items, fulfilling the recommended minimum number of mutilations (Klein-Braley, 1997). The first and last sentences of each passage were left intact providing appropriate context to set the subjects gap-filling mechanism in motion. Each passage contained 20 items of second-word or third-word deletion. The instructions were given in English along with a short English example of that kind of cloze test given to the subjects and its restored answer. The time limit for test completion was set at 30 minutes plus 3 minutes of instruction. The subjects were ensured that misspelling and guessing will not have any penalties and it was emphasized that the blanks are for deleted parts of words.

The tests were administered to the 40 participants at the start of their English classes. In fact test implementation occurred 6 times in six intact classes. Both C-test and NC-Test papers were circulated randomly among students. We had equal replications for our two kinds of cloze tests- 20 subjects per kind (Steel and Torrie, 1980) leading to a balanced design (Hatch and Lazarton, 1991) as presented in Table 1. Four voluntary students from both advanced and elementary language ability groups were subsequently requested to report what test-taking strategies they used while taking each test.

TABLE 1

NUMBER OF PARTICIPANTS

	Test Type	
Group	**C-Test**	**NC-Test**
High	10	10
Low	10	10
Total	20	20

To avoid any subjective judgment, assessment items in cloze tests were considered as objective as possible by implementing the 'exact word method' of scoring.

a. Piloting the C-Test and the New C-Test

The C-Test and the NC-Test were piloted with 8 elementary and advanced Iranian EFL students at the Noor-e-Eram Language Institute in Tabriz. The main purpose was to see whether these cloze tests were too difficult or too easy for the target students. The pilot students were divided into two subgroups in equal proportion to take the C-Test and the NC-Test. In addition, the pilot students were asked to fill in all of the mutilated parts. According to the results of the pilot study, none of the texts were too difficult or too easy.

The reliability coefficient and the consistency of the language proficiency tests were estimated by employing the Kuder-Richardson formulas. The acceptable reliability coefficient of the proficiency tests should be .80 or higher to identify good assessment (Bachman, 1990; Cohen, 1994).

TABLE 2

RELIABILITY OF TESTS USING THE K-R 21

Test Type	Mean	Std. Deviation	Reliability
C-Test	68.75	24.4455	0.9737
NC-Test	65.25	19.0010	0.9466

b. Interview

In the present study, individual interviews were employed to find out the test-taking strategies used by Iranian EFL students. Generally, the interview can be divided in two types: an individual interview and a group interview. The individual interview is appropriate for investigating individual performance in spite of being time-consuming (O'Malley and Chamot, 1990).

Semi-structured interview has been defined by Bryman (2001, p.110) as " a context in which the interview has a series of questions that are in the general form of an interview schedule but is able to vary the sequence of questions…also, the interviewer usually has some latitude to ask further questions in response to what are seen as significant replies". Semi-structured interviewing is guided only in the sense that some form of interview guide, such as the questions mentioned below is prepared in advance, and provides a framework for the interview.

Based on the investigation of Cohen (1984), semi-structured interview questions were constructed carefully and then were piloted to see whether the interview questions were useful in finding out the test-taking strategies of the target students. Moreover, these interview questions focused on the students' cloze test-taking strategies. Three questions constructed about cloze test-taking strategies are as follows:

1. What strategies did you use while answering the C-Test or the NC-Test?

2. Did you read the whole texts, or only parts of the texts, or jump around?

3. Did you use other strategies to find the answers in the cloze tests?

Each individual interview took approximately 2-5 minutes. The students' responses were collected by recording with their permission. Afterwards, the interview data were transcribed and tabulated to see what categories the students employed in filling in the mutilated parts.

III. Results and Analysis

A. Results of the Original C-Test and the NC-Test

The prerequisite for testing the means is the homogeneity of the variances of the scores of the C-Test and the NC-Test within high and low groups, 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. Taking into account the 0.417 significance level in high group and the 0.852 in low group which are larger than 0.05, the homogeneity of the variances in both groups is approved. Then, we utilized the independent sample t-test to compare the mean scores of the C-Test and the NC-Test in both high and low proficiency groups. Taking into account the 0.242 significance level in high group and the 0.307 in low group which is larger than 0.05, the homogeneity of the mean scores in both groups is approved. As a result, the mean scores of the C-Test and the NC-Test in both high-language-ability groups and low-language-ability groups are not significantly different as can be seen in Table 3 and Table 4.

Hence, the C-Test and the NC-Test can be most likely used interchangeably to measure the English language proficiency of the students which is consistent with the findings of Rungruangthum (2005) because of the fact that there were no statistically significant differences between the mean scores on the original C-Test and the NC-Test for both high and low proficiency groups at the 0.05 level of significance (Sig. > 0.05).

TABLE 3

COMPARISON OF THE MEAN SCORES ON THE C-TEST AND THE NC-TEST WITHIN THE HIGH-LANGUAGE-ABILITY GROUPS

Group	**Test-Type**	**N**	**Mean**	**Std. Deviation**	**Levene's Test for Equality of Variances**		**t-test for Equality of Means**		
					F	**Sig.**	**t**	**df**	**Sig.**
high	**C-Test**	10	76.20	20.94	.691	.417	1.210	18	.242
	NC-Test	10	63.60	25.40					

TABLE 4

COMPARISON OF THE MEAN SCORES ON THE C-TEST AND THE NC-TEST WITHIN THE LOW-LANGUAGE-ABILITY GROUPS

Group	Test-Type	N	Mean	Std. Deviation	Levene's Test for Equality of Variances		t-test for Equality of Means		
					F	Sig.	t	df	Sig.
low	C-Test	10	41.00	17.25	.036	.852	1.050	18	.307
	NC-Test	10	33.30	15.48					

The findings of this study confirm the results of some other previous researches indicating that some C-Test items especially the functional words had low discrimination power (Cleary, 1988; Jafarpur, 1999; Wolter, 2002). Hence, the type of the deleted words (content words and functional words) has an influence on the test difficulty (Klein-Braley, 1997). Weir (1990) states that the C-Test offers more chances of guessing because of the second-half or the second-part deletion. The present study indicates that the reliability of the original C-Test (Rc = 0.9737) calculated by Kuder-Richardson 21 formula was higher than the reliability of the NC-Test (Rnc = 0.9466) with third-word deletion. This is inconsistent with the findings of Thongsa-nga (1998; cited in Rungruanthum, 2005) reporting that the reliability of the NC-Test with third-word deletion was higher than the other forms of the C-Tests.

B. Results of Using Third-Word Deletion in the NC-Test

As presented in Table 5, the prerequisite for testing the means is the homogeneity of the variances of the scores within high-language-ability and low-language-ability groups. So, we tested the homogeneity of the variances employing the Levene Test. For the C-Test, the homogeneity of variances is confirmed. For the NC-Test, the significance level of the NC-Test in the Levene Test (0.044) is smaller than 0.05, so the homogeneity of variances is rejected. Therefore, we used the independent sample t-test with moderated degrees of freedom.

Since the significance levels of the independent sample t-test in both kinds of cloze test are smaller than 0.05 (Sig. <0.05), the homogeneity of the mean scores compared between the high and low proficiency groups is rejected. As a consequence, there were statistically significant differences between the high and low groups on the mean scores of two kinds of the cloze test at the 0.05 level of significance (Sig. <0.05), that is, the mean scores of the high proficiency groups are significantly larger than the mean scores of the low proficiency groups in both C-Test and the NC-Test. This indicates that these two kinds of the cloze test could differentiate among Iranian EFL students at different language proficiency levels.

TABLE 5
COMPARISON OF THE MEAN SCORES OF THE C-TEST AND THE NC-TEST BETWEEN THE HIGH- AND LOW-LANGUAGE-ABILITY GROUPS

Test-Type	Group	N	Mean	Std. Deviation	Levene's Test for Equality of Variances		t-test for Equality of Means		
					F	Sig.	t	df	Sig.
C-Test	high	10	76.20	20.94	1.365	.258	4.102	18	.001
	low	10	41.00	17.25					
NC-Test	high	10	63.60	25.40	4.689	.044	3.222	14.876	.006
	low	10	33.30	15.48					

Item Analysis was also used in the present study to analyze all deleted items in the C-test and the NC-Test to find out whether using third-word deletion in the NC-Test had an impact on the test discrimination power. Item analysis consisted of two calculations: "Item Facility (IF)" referring to "the easiness of an item" (Farhady et al., 1994; p. 100) and "Item Discrimination (ID)" indicating "the index which is derived from comparing the difference between the performance of more knowledgeable and less knowledgeable examinees on a particular item" (Farhady et al., 1994; p. 102). Moreover, the acceptable facility values should range from .20 to .80 (Nuttal and Skurnik, 1969; cited in Rungruanthum, 2005) and Jafarpur (2002; cited in Rouhani, 2008) believes that item discrimination indices higher than .20 are acceptable. Therefore, the discrimination indexes more than 0.20 were acceptable in the present study. The summary of the discrimination indexes presented in Table 6 indicates that the number of the C-Test items with higher discrimination power was respectively more than that of the NC-Test items in assessing the general language skills of English.

TABLE 6
ITEM DISCRIMINATION OF THE C-TEST AND THE NC-TEST

ID	C-Test	NC-Test
≥ 0.80	3	1
0.61 - 0.79	2	0
0.41 - 0.60	28	10
0.21 - 0.40	39	54
≤ 0.20	28	35
Total	100	100

In the original C-Test, 28 mutilated items (28%) out of the total number (100) were considered to have low discrimination power. Most of the deleted items were reported to be too easy for these Iranian EFL students since this type of the language test probably provides more chances of guessing as can be seen in Table 7. Although the results of this study reveal that the original C-Test could distinguish among Iranian EFL students with high and low proficiency levels, some items of the C-Test should be modified to make it more appropriate for tapping the English language proficiency of Iranian EFL students.

TABLE 7
THE C-Test ITEMS WITH LOW DISCRIMINATION POWER

Item	**Word**	**IF Total**	**ID**	**Item**	**Word**	**IF Total**	**ID**
3	in	0.9	0.2	83	thin	0.4	0.2
7	animals	0.8	0.2	85	German	0.5	0.2
12	in	0.85	0.1	86	is	0.7	0
37	they	0.75	0.1	87	and	0.55	0.1
38	in	0.95	0.1	89	like	0.5	0.2
57	is	0.85	0.1	91	others	0.45	0.1
72	the	0.65	0.1	92	meat	0.5	0
73	sea	0.7	0	93	if	0.4	0.2
74	there	0.5	0	94	is	0.6	0
77	mountains	0.8	0	95	Many	0.65	0.1
78	great	0.5	0.2	96	like	0.65	0.1
79	of	0.5	0.2	97	fried	0.3	0.2
81	example	0.6	0.2	99	there	0.7	-0.2
82	soup	0.6	0	100	people	0.75	0.1

Considering the NC-Test with third-word deletion, 35 mutilated items (35%) out of the total number (100) were considered to have low discrimination power. Most of the deleted items were reported to be too easy for these Iranian EFL students because Weir (1990) contended that the second half or the second part deletion provides more chances of guessing.

Only three items (44, 54, and 80) were too difficult (IF<0.2) for the Iranian EFL students as shown in Table 8. Therefore, there were 32 items in the NC-Test considered as less suitable for measuring the English language proficiency of the Iranian EFL students. The results of the present study reveal that the NC-Test could differentiate among Iranian EFL students with high and low proficiency levels. The results of the item analysis suggested that utilizing third-word deletion gave the NC-Test items lower discrimination power than the original C-Test in assessing the English language proficiency of Iranian English language students in an EFL context. Hence, the different rate of deletion had an impact on the discrimination power of the NC-Test in the present study.

TABLE 8
THE NC-Test ITEMS WITH LOW DISCRIMINATION POWER

Item	Word	IF Total	ID	Item	Word	IF Total	ID
3	of	0.5	0.2	67	mountains	0.6	0.2
7	no	0.9	0.2	68	the	0.5	0.2
8	in	0.85	0.1	69	And	0.3	0.2
13	animal	1	0	72	great	0.3	0.2
17	When	0.6	0.2	73	light	0.3	0.2
22	are	0.75	0.1	75	green	0.65	0.1
24	news	0.4	0.2	76	with	0.4	0.2
29	people	1	0	79	is	0.7	0.2
31	if	0.5	0.2	80	faraway	0.1*	0.2
39	his	0.5	0.2	81	Chinese	0.75	0.1
40	he	0.7	0.2	83	but	0.5	0.2
41	its	0.6	0.2	88	meat	0.4	0.2
44	foam	0.1*	0.2	89	it	0.5	0.2
49	miles	0.5	0.2	92	fine	0.2	0.2
54	border	0.1*	0.2	93	are	0.6	0
56	empty	0.3	0.2	97	eat	0.6	0.2
58	few	0.3	0.2	98	people	0.7	0
60	sea	0.7	0.2				

Many studies confirm that deleted functional words can be restored easier than deleted content words (Cleary, 1988; Jafarpur, 1999; Klein-Braley, 1997; Weir, 1990; Wolter, 2002). Afterwards, Klein-Braley (1997, p.59) believed that "the actual difficulty of the test constructed from the same text varied according to the proportion of function words deleted". The more content words a text contains, the more difficult the text is for the fill-in activity. The original C-Test consisted of 64 content words and 36 function words while the NC-Test included 67 content words and 33 function words as presented in Table 9. Therefore, the third-word deletion made the NC-Test more difficult than the original C-Test and increased the NC-Test items with lower discrimination power.

TABLE 9
THE NUMBER OF CONTENT AND FUNCTION WORDS
IN THE C-TEST AND THE NC-TEST

Test-Type	Content Words					Function Words					
	Noun	Verb	Adjective	Adverb	Total	Pronoun	Preposition	Conjunction	Article	Negative	Total
C-Test	27	20	13	4	64	14	11	7	4	0	36
NC-Test	28	20	14	5	67	12	8	9	4	0	33

As a consequence, the findings of this study indicate that different deletion techniques influence each kind of cloze test used in this study especially its discrimination power. That is, the results of this study confirm some previous research findings showing that changing the deletion rates and the starting points affects on language tests (Jafarpur, 1995).

In conclusion, the findings of the present study reveal that using third-word deletion in the NC-Test resulted in tests with lower discrimination power.

C. Cloze Test-Taking Strategies Used by Volunteer Students

The interview data obtained from the volunteer students (N=4) was tabulated and analyzed based on the latest categorization framework of cloze test-taking strategies represented by Sasaki (2000). These findings indicate that the volunteer students said that they used the Extratextual and Missing strategies most. Moreover, the high proficiency level students in this study could answer more items correctly than the low proficiency level students. The high-language-ability students utilized the Within Clause; Across Paragraphs, Within Text; and Guessing test-taking strategies more frequently than the low-language-ability students. While the low proficiency level students mostly employed the Across Sentences, Within Paragraph strategy for responding the C-Test and the NC-Test.

The test-taking strategies used by the volunteer C-Test taking subjects in both high and low proficiency groups are expressed in Table 10. Both high and low proficiency students said that they often used Within Clause; Across Clause, Within Sentence; Across Paragraphs, Within Text; Extratextual; Guessing; and Missing while taking the C-Test. Moreover, neither used Across Sentences, Within Paragraph strategy. Therefore, the C-Test probably measured the overall language proficiency of English.

TABLE 10
STRATEGIES USED BY C-TEST TAKERS

1. Within Clause	**High:** "I thought of the meaning of the word." **Low:** "I used the clues before the deleted words."
2. Across Clause, Within Sentence	**High:** "I read for main idea and used the contextual clues." **Low:** "used contextual clues and grammatical structure."
3. Across Sentences, Within Paragraph	**High:** –––– **Low:** ––––
4. Across Paragraphs, Within Text	**High:** "I read every sentence in the paragraph before answering." **Low:** "I read all the text before filling in the blanks."
5. Extratextual	**High:** "In some parts I used the general knowledge to fill in the deleted words." **Low:** "I had previous knowledge."
6. Guessing	**High:** "I counted the number of the deleted letters then I guessed." **Low:** "I guessed."
7. Missing	**High:** "If I didn't know the answer, I would leave it blank." **Low:** "When I didn't know the answer, I would leave it blank."

The test-taking strategies employed by the volunteer NC-Test taking students in both high- and low-language-ability groups are presented in Table 11. Both high and low proficiency students reported that they frequently used Within Clause; Across Clause, Within Sentence; Across Sentences, Within Paragraph; Extratextual; and Missing while taking the NC-Test. Moreover, the high-language-ability subject said that she guessed the answer by counting the number of the deleted letters. Additionally, neither used Across Paragraphs, Within Text strategy. Therefore, the NC-Test possibly measured the English language proficiency of the volunteer students.

TABLE 11
STRATEGIES USED BY NC-TEST TAKERS

1. Within Clause	**High:** "I found part of speech of the word and the meaning of the word." **Low:** "Thinking of the meaning of the word"
2. Across Clause, Within Sentence	**High:** "I used contextual clues and grammatical structure and I analyzed the deleted word to see what type of word it is." **Low:** "reading for main idea and using contextual clues"
3. Across Sentences, Within Paragraph	**High:** "I read some parts of the passage." **Low:** "Reading some parts of the passage."
4. Across Paragraphs, Within Text	**High:** ——— **Low:** ———
5. Extratextual	**High:** "I used general knowledge to fill in the deleted words." **Low:** "I had a previous knowledge."
6. Guessing	**High:** "I counted the number of deleted letters and then I guessed." **Low:** ———
7. Missing	**High:** "If I didn't know the answer, I'd leave it blank." **Low:** "If I didn't know the answer, I would leave it blank, I would ignore it."

Considering the original C-Test, this study indicates that the volunteer high- and low-language-ability students mostly used the Within Clause; Across Clause, Within Sentence; Across Paragraphs, Within Text; Extratextual; Guessing; and Missing strategies while answering the C-Test. When the passage is familiar to the students' past experience, they could restore test items easier than unfamiliar passages. Moreover, the students' responses indicated that the C-Test in this study probably measured the overall language proficiency of the target students since all of them utilized the Across Paragraphs, Within Text and Extratextual strategies which are beyond the sentence level. These findings support many previous studies (Babaii and Ansari, 2001; Connelly, 1997; Dörney and Katona, 1992; Klein-Braley, 1997) indicating that the C-Test could measure the proficiency of the target language. And these findings do not accord with the findings of Rungruangthum (2005) believing that the C-Test seems suitable for measuring specific language ability, such as grammar and vocabulary.

While taking the NC-Test, the volunteer high and low proficiency students reported that they frequently used the Within Clause; Across Clause, Within Sentence; Across Sentences, Within Paragraph; Extratextual; and Missing strategies and only a high-language-ability student used the Guessing strategy. Moreover, the students' responses revealed that the NC-

Test in this study could probably measure the overall language proficiency of the volunteer students since all of them used the Extratextual test-taking strategy which is beyond the sentence level. As a result, this study is in agreement with Thongsa-nga's (1998; cited in Rungruanthum, 2005) results showing that the NC-Test with third-word deletion is appropriate for assessing the English language proficiency of twelfth graders while it is inconsistent with the claims of Rungruangthum (2005) stating that the NC-Test seems suitable for measuring specific language ability, such as grammar and vocabulary.

Moreover, the results of the test-taking strategies for the C-Test and the NC-Test report that both the high and low proficiency level students frequently utilized the Within Clause; Extratextual; and Missing strategies. Therefore, these two kinds of cloze test seem to be able to measure the overall language proficiency of the Iranian EFL students. As a result, the findings of the present study are in agreement with many other studies (Babaii and Ansari, 2001; Connelly, 1997; Dörney and Katöna, 1992; Klein-Braley, 1997) claiming that the C-Test was reliable in assessing the proficiency of the second and foreign language learners, and the C-Test was easy to score and construct. Conversely, some previous studies (Jafarpur, 1995; Sigott and Köberl, 1993) show that the C-Test is considered to be less suitable in measuring English language proficiency of EFL university students. However, some test items in these two language tests with low discrimination power were too easy or too difficult due to the type of the deleted words. The functional words were easier to restore than the content words (Jafarpur, 1999; Klein-Braley, 1997; Wolter, 2002). Language teachers who intend to make use of these two kinds of cloze test should take into account the word selection and the deletion techniques which are appropriate for students' language proficiency level.

IV. Conclusion

The results of this study indicate that the C-Test and the NC-Test can be most likely used interchangeably to measure the English language proficiency of the students due to the fact that there were no statistically significant differences between the mean scores on the original C-Test and the NC-Test for both high and low proficiency groups at the 0.05 level of significance (Sig. > 0.05).

In this research, the third-word deletion was used in the NC-Test to provide more clues for the EFL students. The findings of this study reveal that using third-word deletion technique affects on the discrimination power of the tests. Considering the comparisons between the C-Test and the NC-Test, third-word deletion decreased the discrimination power of the NC-Test. The results indicate that both C-Test and NC-Test could distinguish among the high and low proficiency level students. Another factor that influenced the two kinds of cloze test in this research was the type of deleted words. For instance, deleted function words could be restored by using only linguistic or grammatical competence.

Finally, we came to the conclusion that these two kinds of cloze test could measure the English language proficiency of the students within an EFL situation regarding the test-taking strategies. The interview data in this research indicates that the students in these kinds of cloze test utilized at least one of these strategies which are beyond the sentence level: the Across Paragraphs, Within Text; and the Extratextual strategies. Almost all of the students used the Extratextual strategy due to the fact that the subjects of some the texts were familiar to them.

The present research has of course suffered some limitations due to administrative problems. The major limitations of this research were that it was gender-based and confined merely to

the subjects of one institute. Another limitation to the present research was that it used the language proficiency placement of the institute instead of administering a criterion measure.

Acknowledgement: This paper is extracted from M.A. thesis "A Comparative Study of Iranian EFL Students' Performance between the C-Test and the NC-Test AND between the MC-Test and the NMC-Test" written by Farideh Hosseini with special thanks to my supervisor Dr. Salahshoor.

References:

Alderson, J. C. (1983). The cloze procedure and proficiency in English as a foreign language. In J. W. Oller, Jr., (Ed.), *Issues in language testing research* (pp. 205- 217). Rowley, MA: Newbury House.

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

Bachman, L.F. (1985). Performance on cloze tests with fixed ratio and rational deletions. *TESOL Quarterly*, 19 (3), 535-556.

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

Bryman, A. (2001). *Social Research Methods*. Oxford: Oxford University Press.

Cleary, C. (1988). The C-Test in English: Left-hand deletions. *RELC Journal*, 19 (2), 26- 38.

Cohen, A. D. (1984). On taking language tests: what the students report. *Language Testing*, 1, 70-81.

Cohen, A. D. (1994). *Assessing language ability in the classroom* (2nd ed.). Boston: Heinle & Heinle Publishers.

Cohen, A. D. (1998). Strategies and processes in test taking and SLA. In Bachman, L. F. & Cohen, A. D. (Ed). *Interfaces between second language acquisition and language testing research*. (pp. 90-111). Cambridge: Cambridge University Press.

Cohen, A. D. & Upton, T. A. (2006). Strategies in responding to the new TOEFL reading tasks. *TOEFL Monograph Series*. ETS TOEFL.

Connelly, M. (1997). Using C-Test in English with postgraduate students. *English for Specific Purposes*, 16, 139-150.

Dörnyei, Z., & Katona, L. (1992). Validation of the C-Test amongst Hungarian EFL learners. *Language Testing*, 9 (2), 187-206.

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

Farhady, H. (1996). Varieties of cloze procedure in EFL education. *Roshd FLT Journal*, 12 (44), 30-40.

Farhady, H., Jafarpur, A., & Birjandi, P. (1994). *Testing language skills: From theory to practice*. Tehran, Iran: SAMT Publications.

Feldmann, U., & Stemmer, B. (1987). Thin_ aloud a_ retrospective da_ in c-te_ taking: diffe_ languages-diff_ learners-sa_ approaches? In C. Faerch & C. Kasper (Eds.), *Introspection in Second Language Research* (pp. 251-267). Multilingual Matters, Clevedon.

Grotjahn, R. (1986). Test validation and cognitive psychology: Some methodological considerations. *Language Testing*, 3, 159-185.

Hastings, A. (2002a). Error analysis of an English C-Test: Evidence for integrated processing. In R. Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen [The C-Test: theoretical foundations and practical applications]* (Vol. 4, pp. 53-66). Bochum: AKS-Verlag.

Hastings, A. (2002b). In defense of C-testing. In R. Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen [The C-Test: theoretical foundations and practical applications]* (Vol. 4, pp. 11-25). Bochum: AKS-Verlag.
Hatch, E., & Lazarton, A. (1991). *The Research Manual: Design and Statistics for Applied Linguistics*. Rowley, MA: Newbury House Publishers.
Hughes, A. (2003). *Testing for language teachers (2nd ed.)*. Cambridge: Cambridge University Press.

Jamil, A., Abd. Aziz, M. S., & Abdul Razak, N. (2010). The utilization of test-taking strategies among female students in a tertiary institution. *GEMA Online™ Journal of Language Studies*, 10 (3), 105-125.
Jafapur, A. (1995). Is C-Test superior to cloze? *Language Testing*, 12 (2), 194-216.
Jafarpur, A. (1997). *An introduction to language testing*. Shiraz: Shiraz University Press.
Jafarpur, A. (1999). Can the C-test be improved with classical item analysis? *System*, 27, 79-89.
Jafarpur, A. (2002). A comparative study of a C-Test and a cloze test. In R. Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen [The C-Test: theoretical foundations and practical applications] (Vol. 4, pp. 31-51).* Bochum: AKS-Verlag.
Kamimoto, T. (1992). An inquiry into what a C-Test measures. *Fukuoka Women's Junior College Studies*, 44, 67-79.
Klein-Braley, C. (1997). C-Test 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 (2), 134-146.
Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement*. (3rd ed). (pp. 13-103). New York: Macmillan.
Nuttal, D. L., & Skurnik, L. S. (1969). Examination and item analysis manual. England and Wales: National Foundation for Educational Research.
O' Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.
Raatz, U., & Klein-Braley, C. (1981). The C-Test--a modification of the cloze procedure. In T. Culhane, C. Klein-Braley & D. K. Stevenson (Eds.), *Practice and problem in language testing* (pp. 113-138). Colchester: University of Essex.
Rahimi, M., & Saadat, M. (2005). A verbal protocol analysis of a C-Test. *IJAL*, 8 (2), 55-85.
Rouhani, M. (2008). Another look at the C-Test: A validation study with Iranian EFL learners. *The Asian EFL Journal*, 10(1), 154-180.
Rungruangthum, M. (2005). A comparative study between the C-Test and the NC-Test and between the MC-Test and the NMC-Test using identical texts. M.A. Dissertation. Mahidol University of Thailand.
Sasaki, M. (2000). Effects of cultural schemata on students' test-taking processes for cloze tests: A multiple data source approach. *Language Testing*, 17, 85-114.
Sigott, G., & Köberl, J. (1993). Validating the X-Test. *Language Testing*, 14, 53-58.
Steel, R. G. D., & Torrie, J. H. (1980). *Principles and procedures of statistics: A biometrical approach*. (2nd ed.). New York: Mc Graw- Hill.
Stemmer, B. (1991). *What's on a C-Test taker's mind: Mental processes in C-Test taking*. Bochum: Brockmeyer.
Storey, P. (1997). Examining the test-taking process: A cognitive perspective on the discourse cloze test. *Language Testing*, 14 (2), 214-231.
Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 30, 413-453.

Thongsa-nga, T. (1998). *A validation of the NC- Test for Mattayom six students*. Master's Thesis, The National Institute of Development Administration, Bangkok, Thailand.
Tsagari, C. (1994). *Method effect on testing reading comprehension: How far can we go?* Unpublished M.A. thesis. University of Lancaster, U.K.
Weir, C. (1990). *Communicative language testing*. Hemel Hempstead: Prentice Hall.

Wolter, B. (2002). Assessing proficiency through word associations: Is there still hope? *System*, 30, 315-329.

Appendix A.

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.

Example:

America's oldest cities, such as Boston, San Francisco, and New York, were built before the days of the automobile. At th_ _ time, ever_ _ _ _ had t_ live cl_ _ _ to t_ _ city cen_ _ _. But t_ _ coming o_ the c_ _ changed a_ that. Mo_ _ and mo_ _ of th_ _ _ who co_ _ _ afford t_ buy th_ _ _ own ho_ _ moved o_ to t_ _ new sub_ _ _ _, where they could forget the noise, dirt, and crime of the city.

Answers:

America's oldest cities, such as Boston, San Francisco, and New York, were built before the days of the automobile. At that time, everyone had to live close to the city center. But the coming of the car changed at that. More and more of those who could afford to buy their own house moved on to the new suburbs, where they could forget the noise, dirt, and crime of the city.

Now, try to complete the texts below. Marks will be taken off for spelling errors.

Text 1

The lion is called the king of beasts. Lions a_ _ found liv_ _ _ wild i_ the grass_ _ _ _ _ of Afr_ _ _. They hu_ _ smaller ani_ _ _ _ and fe_ _ on th_ _. There a_ _ no wi_ _ lions i_ Europe, b_ _ there a_ _ captive li_ _ _ in Euro_ _ _ _ zoos. T_ _ male li_ _ is a beau_ _ _ _ _ animal. Ro_ _ _ his head he has a ring of long hair called a mane. When the lion is young, the hair of his mane is yellow. When he is old, the hair is sometimes black. The female lion, or lioness, does not have a mane. Lions are dangerous animals. A lion can kill a man.

Text 2

People faint when the normal blood supply to the brain is suddenly cut down. This c_ _ happen i_ they a_ _ surprised o_ shocked b_ sudden ne_ _ or b_ something th_ _ see. So_ _ people fa_ _ _ if th_ _ see oth_ _ _ hurt. So_ _ people fa_ _ _ in cro_ _ _. Others fa_ _ _ if th_ _ are i_ a room th_ _ is h_ _ and stuffy. If a person faints while standing, lay him down. If his face is pale, lift his feet. If he is sitting down when he faints, place his head between his knees. Loosen any tight clothing that might keep him from breathing easily. If possible, place a cold, wet cloth on his forehead.

Text 3

The Black Sea gets its name from the color of its water. In win_ _ _ its co_ _ _ is ve_ _ dark. Th_ _ is cau_ _ _ by fo_ _ that set_ _ _ _ low ov_ _ the ar_ _ and c_ _ off sunl_ _ _ _. The Bl_ _ _ Sea i_ 748 mi_ _ _ from ea_ _ to we_ _; it i_ 374 mi_ _ _ from no_ _ _ to so_ _ _. Four countries- Russia, Romania, Bulgari, and Turkey- border the sea. Several large rivers empty into it; the Danub, Dnieper, Don, Bug, and Kuban are a few. The deepest part of the sea is in its south central region. Many ports line the sea. Grain, lumber and sugar are the main exports that pass through these ports. Fishing is good in the Black Sea and supports many of the people on its coasts.

Text 4

We have just climbed out of a spaceship onto the surface of the moon. Behind u_ is t_ _ ship, ha_ _ in t_ _ sunlight a_ _ half i_ deep sha_ _ _. A few mi_ _ _ ahead i_ a wall o_ mountains towe_ _ _ _ against t_ _ black s_ _. And th_ _ _, as tho_ _ _ resting o_ the moun_ _ _ _ _, is a gr_ _ _ ball o_ light beaut_ _ _ _ _ _ colored in blue and green and brown with a patch of dazzling white at the top. It is our own faraway world- the earth. We take a step and rise like prize jumpers- up, float, and down again. Hopping carefully, we explore the valleys, the sloping crater walls, the shadowy crater floors. Not a sound can be heard there is no air to carry sound, no wind; there are no smells, no plants, and no animals. There is nothing but rock and dust, blinding sunlight and cold black shadows.

Text 5

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 i_ 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.

Appendix B.

The NC-Test

In each of the texts below, the second part of every third 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.

Example:

America's oldest cities, such as Boston, San Francisco, and New York, were built before the days of the automobile. At that ti_ _, everyone had t_ live close t_ the city cen_ _ _. But the com_ _ _ of the c_ _ changed at th_ _. More and mo_ _ of those w_ _ could afford t_ buy their o_ _ house moved o_ to the n_ _ suburbs, where they could forget the noise, dirt, and crime of the city.

Answers:

America's oldest cities, such as Boston, San Francisco, and New York, were built before the days of the automobile. At that time, everyone had to live close to the city center. But the coming of the car changed at that. More and more of those who could afford to buy their own house moved on to the new suburbs, where they could forget the noise, dirt, and crime of the city.

Now, try to complete the texts below. Marks will be taken off for spelling errors.

Text 1

The lion is called the king of beasts. Lions are fo_ _ _ living wild i_ the grasslands o_ Africa. They hu_ _ smaller animals a_ _ feed on th_ _. There are n_ wild lions i_ Europe, but th_ _ _ are captive li_ _ _ in European zo_ _. The male li_ _ is a beautiful ani_ _ _. Round his he_ _ he has a ri_ _ of long ha_ _ called a mane. Wh_ _ the lion i_ young, the ha_ _ of his ma_ _ is yellow. When he is old, the hair is sometimes black. The female lion, or lioness, does not have a mane. Lions are dangerous animals. A lion can kill a man.

Text 2

People faint when the normal blood supply to the brain is suddenly cut down. This can hap_ _ _ if they a_ _ surprised or sho_ _ _ _ by sudden ne_ _ or by some_ _ _ _ _ they see. So_ _ people faint i_ they see oth_ _ _ hurt. Some peo_ _ _ faint in cro_ _ _. Others faint i_ they are i_ a room that i_ hot and stu_ _ _. If a person fai_ _ _ while standing, l_ _ him down. I_ his face i_ pale, lift h_ _ feet. If h_ is sitting down when he faints, place his head between his knees. Loosen any tight clothing that might keep him from breathing easily. If possible, place a cold, wet cloth on his forehead.

Text 3

The Black Sea gets its name from the color of its water. In winter i_ _ color is ve_ _ dark. This i_ caused by fo_ _ that settles l_ _ over the ar_ _ and cut o_ _ sunlight. The Bl_ _ _ Sea is 748 mi_ _ _ from east t_ west; it i_ 374 miles from no_ _ _ to south. Fo_ _ countries- Russia, Romania, Bulgari, and Turkey- bor_ _ _ the sea. Sev_ _ _ _ large rivers em_ _ _ into it; t_ _ Danub, Dnieper, Don, Bug, and Kuban are a f_ _. The deepest pa_ _ of the s_ _ is in its south central region. Many ports line the sea. Grain, lumber and sugar are the main exports that pass through these ports. Fishing is good in the Black Sea and supports many of the people on its coasts.

Text 4

We have just climbed out of a spaceship onto the surface of the moon. Behind us i_ the ship, ha_ _ in the sunl_ _ _ _ and half i_ deep shadow. A f_ _ miles ahead i_ a wall of moun_ _ _ _ _ towering against t_ _ black sea. A_ _ there, as tho_ _ _ resting on t_ _ mountains, is a gr_ _ _ ball of li_ _ _ beautifully colored i_ blue and gr_ _ _ and brown wi_ _ a patch of dazz_ _ _ _ white at t_ _ top. It i_ our own far_ _ _ _ world- the earth. We take a step and rise like prize jumpers- up, float, and down again. Hopping carefully, we explore the valleys, the sloping crater walls, the shadowy crater floors. Not a sound can be heard there is no air to carry sound, no wind; there are no smells, no plants, and no animals. There is nothing but rock and dust, blinding sunlight and cold black shadows.

Text 5

People in different countries may eat the same food but they prepare it very differently. For example, Chi_ _ _ _ soup is th_ _ and clear, b_ _ German soup i_ thick and he_ _ _. Some people li_ _ raw meat, wh_ _ _ others like me_ _ only if i_ is well-cooked. Ma_ _ people like but_ _ _ fried and fi_ _, but there a_ _ people in In_ _ _ who like i_ melted into o_ _ before they e_ _ it. Many peo_ _ _ in the Ea_ _ like plain boi_ _ _ rice, but in some countries people like theirs made into a sweet pudding.