A Verbal Protocol Analysis of a C-Test

Mohammad Rahimi
and
Mahboobeh Saadat

Shiraz University

E-mail Addresses: mrahimy@gmail.com
mahsaadat@ gmail.co

Abstract

The C-test is widely known as a test of overall language proficiency. Most of the evidence in this regard has been obtained through correlational studies. Nonetheless, construct validity of the C-test is just partially established. Moreover, such studies do not reveal anything about the mental processes going on in the mind of the testees. Verbal protocol analysis has been recommended as an important tool to validate the C-test. A C-test consisting of 5 texts with 100 deletions was given to a sample of 26 Iranian English seniors, and subsequently a retrospective verbal protocol analysis was carried out to learn what happened in the mind of the testees while they were restoring the test items. The results of the study showed that the subjects used 13 different strategies, consisting of both bottom-up and top-down processes. However, the use of different strategies varied as a function of both the types of items in the C-test as well as the proficiency level of the subjects. The results of the study suggest construct validity of the C-test as a test of overall language proficiency.

Key words: C-test, construct validity, language proficiency, reduced redundancy, retrospection, verbal proto.

Introduction

Tests of reduced redundancy have been widely appreciated for being highly valid and eminently authentic. These tests reflect the sociolinguistic-integrative approach to language testing according to which knowledge of a language necessarily requires the ability to function when there is reduced redundancy through the use of what Oller (1979) calls an *expectancy grammar*. In fact, as Feldmann and Stemmer (1987, p. 255) state, "Comprehension of input leads us to form certain expectations about what will come next, be it the next letter, the next word, or the next sentence." Klein-Braley (1997) believes that the concept of reduced redundancy can serve as a good criterion to measure the learner's language proficiency.

The cloze, widely used as a test of overall language proficiency, is an example of tests of reduced redundancy. This test consists of a passage in which every nth word--usually, every fifth, sixth or seventh word--is deleted. The results of many studies have lent support to the validity of cloze as a measure of overall language proficiency by establishing high correlation between the scores of the subjects on this test and those obtained from discrete-point proficiency tests such as TOEFL and UCLA placement test (Chappelle and Abraham, 1990; Oller 1988; Alderson, 1979, 1980; Darnell, 1968 to name a few). None the less, cloze, as Klein-Braley and Raatz (1985) and Klein-Braley (1997) state, suffers from some rather serious shortcomings mainly pertinent to the deletion and scoring procedures employed, reliability and validity of the test, as well as the fact that the use of a single text may make the test biased.

To eliminate the above-mentioned drawbacks, Klein-Braley and Raatz (1985) offered the C-test technique. The C-test has been widely used and praised as a valid test of overall language proficiency. Numerous studies have found high correlations between the C-test and other

integrative tests, say, the cloze test and dictation, and other tests of language proficiency (Jafarpur, 2001; Klein-Braley, 1997; Mochizuki, 1994; Katona, 1992; Neghishi, 1987). They have all been indicative of the empirical validity of the C-test. For instance, Jafarpur (2001) has 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. Inasmuch as the scores show high reliability and concurrent validity, he concludes that the C-test is advantageous over cloze. Klein-Braley (1997) has shown 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. Accordingly, she is convinced that the C-test is the best representative of reduced redundancy tests of general language proficiency. Mochizuki (1994) has demonstrated that the C-test highly correlates with two language proficiency tests--STEP and CELT. He concludes that the C-test seems to be a promising means of assessing overall language proficiency. Having found a high correlation between the C-test scores and those of a language proficiency test in the case of Hungarian subjects, Dornyei and Katona (1992) come to the conclusion that the C-test is a highly valid and reliable integrative instrument for measuring the overall language proficiency. Moreover, they consider it a better measure of general language proficiency than the cloze test. Finally, Neghishi (1987) observes a high correlation between the C-test scores and the scores obtained from a language proficiency test--ELBA.

In addition, Klein-Braley (1985) produces various types of evidence in support of the C-test as a measure of general language proficiency. For instance, she claims that processing the C-test requires, at least, some of the mechanisms involved in normal language processing inasmuch as type-token ratio and mean sentence length, two popular indices of text difficulty and readability, can predict C-test difficulty as well. Hastings (2002) demonstrates that:

> A C-test measures the ability to apply and integrate contextual, semantic, syntactic, morphological, lexical, and orthographic information and knowledge pertaining to a particular written language. Furthermore, the processing that is required for a successful C-test performance seems comparable to natural language processing in both length and complexity, and may in fact have much in common with natural language performance.

However, he admits that his study, being merely an exploratory error analysis of the C-test, fails to definitely answer what a C-test measures. Sigott (2002) disputes the claims that underestimate the C-test prevalently as a test of lower-order skills. The results of his study suggest that the individual test taker's characteristics and those of the individual C-test passage determine whether high-level processing is triggered by an item or not. He argues that the facility index at text level and the word class to which an item belongs are not reliable predictors of high-level processing since a significant number of the subjects engaged in high-level processing to restore both easy and difficult items from all four classes under study.

The validity of the C-test, as a test of overall language proficiency, however, has been criticized based on the results of a series of studies. For instance, Jaafarpur (1995), emphasizing that there is nothing unique about the Rule-of Two, refutes the claims made on the C-test. Particularly, he shows that "The Rule-of Two produces a sizable number of nonfunctioning items" (p. 97). In addition, he convincingly claims that the C-test is not able to make discrimination among the examinees of different proficiency levels. Sigott (1995) comments that the C-test items are sensitive to aspects of vocabulary, syntactic competence, and sentence level grammar. Having reviewed different evidence pertinent to what the

C-test measures, Chapelle (1994) reaches no definite result as to whether the C-test is a valid test of overall language proficiency or not. Hood (1990) claims that the C-test scores are more indicative of general reading skill than general language skill. In fact, he does not find any evidence showing the supremacy of the C-test over the cloze test. Kamimoto (1992) believes that the C-test tends to measure the subjects' vocabulary and grammatical competence and hence its processing occurs at the micro level. He relates this fact to the deletion procedure employed in the design of the C-test. Stemmer (1991) demonstrates that different results may be obtained from the C-test depending on individual text characteristics. Furthermore, the fact that function words are restored more successfully than content words and that text understanding rarely exceeds the proposition border convince Stemmer to presume that the current form of the C-test does not tap general language proficiency. Similarly, Cleary (1988) asserts that the C-test fails to appropriately measure the general language proficiency. Cohen *et. al* (1984), similar to Kamimoto (1992), believe that C-test processing is more at the micro level. They posit that due to the type of deletion procedure employed in the C-test, the testee pays more attention to such aspects as vocabulary and grammar than higher levels of language. Singleton and Little (cited in Chapelle 1994) consider the C-test responses as a source of evidence showing second language lexical development and processing.

The latter group of studies calls into question the validity of the C-test as a test of general language proficiency. Even the correlational studies that prove the empirical validity of the C-test cannot guarantee its construct validity because as Kamimoto (1992, p. 69) states,

> This method of statistical analysis gives no access to what really goes on in the students' minds when they take a C-test. Correlational studies only show an outcome of what has already taken place and prevent us from knowing whether students resort to either integrative skills or discrete-point skills..... In short, studies only on correlational studies are not sufficient for the purpose of an inquiry into what a C-test measures.

In addition, Grotjahn (1986) mentions three reasons why correlational studies are inadequate for construct validation of the tests. Firstly, he presumes, construct validation of tests is only partially established with the help of other tests. Indeed, he believes in the circularity of this approach. Second, the validation of the tests through correlational studies does not tell us anything about the mental processes going on in the mind of the learner. Finally, he contends that the results of such studies heavily depend on the number and type of variables included in the study.

To determine what exactly a measure taps, some scholars have suggested verbal protocol analysis--introspective and retrospective techniques--which help researchers figure out what is really going on in the mind of the testee while taking the test. As Grotjahn (1986, p. 162) remarks, "... in validating (language) tests we also have to analyze the mental processes in the test-taking subject" (p. 162). Green (1998, p.7) states, "The fundamental underlying assumption for protocol analysis is that information that is heeded as a task is being carried out is represented in a limited capacity short-term memory, and may be reported following an instruction to either talk aloud or think aloud." Similarly, Ericsson and Simon ((1984)) maintain that introspective and retrospective reports may tap some of the testee's cognitive processes.

Indeed, Babaii and Ansary (2001) in a retrospective analysis of the C-test found that the learners used four major types of cues with varying frequencies to restore the items in the C-test: automatic processing, lexical adjacency, sentential cues, and top-down cues. They came to the conclusion that C-testing is a reliable and valid procedure mirroring the reduced redundancy principle.

Feldmann and Stemmer (1987), through think-aloud protocols and retrospective interviews 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. According to Adams and Collins (1979), bottom-up strategies are adopted when the information in the text is novel or does not fit the learner's ongoing hypotheses about the content of the text; top-down processing helps the reader to resolve ambiguities or to make a choice between alternative interpretations of the data. Feldmann and Stemmer (1987) identified different strategies adopted 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 enumerate 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.

Storey (1997) is of the opinion that one may find varying degrees of construct validity for different items in a discourse cloze, another measure of reduced-redundancy: "If the item is able to generate processes identified in a theoretical model of the reading process it can be shown to have a good level of construct validity. If alternative processes, irrelevant to the underlying construct, are generated, the validity of the item is called into question" (p. 227). He continues, "If test items generate other processes, then they are not testing what they are designed to test, in other words, they lack construct validity" (p. 226). In his study, Storey noticed that the subjects analyzed the rhetorical structure of the text more deeply for restoring deleted discourse markers. Hence, he concluded that the construct validity of such items was established. However, when the

subjects used a variety of surface matching to restore the deleted cohesive ties, he called the validity of such items into question.

Some verbal protocol analyses carried out in the case of the C-test have cast doubt on its construct validity as a test of general language proficiency. Grotjahn (1986) maintains, "The C-test is very economical and, above all, a highly reliable measurement instrument. However, what it measures, i.e., its construct validity, is in my opinion thus far not very clear" (p. 161). Chapelle and Abraham (1990) believe that C-testing is mostly a measure of grammatical competence rather than textual competence. Finally, Cohen *et al.* ((1984)) and Kamimoto (1992) insist that the cognitive processes in the case of the C-test are more at the micro level than the macro level. They believe that due to the deletion procedure, the learner uses the lexical and grammatical processes to provide the response to the test.

As the results of the aforementioned studies reveal, construct validity of the C-test, examined through verbal protocol analysis, is still in a state of indeterminacy. As such, this paper is a further attempt at investigating the construct validity of the C-test through the analysis of the processes going on in the mind of the testees retrospectively.

Method

Subjects

The subjects of the study were 26 Iranian English seniors taking a course in language testing with the first researcher. They were native speakers of Persian and enjoyed different levels of proficiency in English. They were in their twenties and of both sexes, 18 females and 15 males.

Instruments

The instrument utilized in the study was a C-test consisting of 5 texts with 100 deletions. To construct the test, six short passages with a variety of interesting subjects and different levels of difficulty, judged by the Flesch Reading Ease readability scale (Microsoft Word, (1995), were selected from Rakhshanfar and Jahrudi (n.d.). The texts were arranged from the easiest to the most difficult one as recommended by Klein-Braley and Raatz (1985). The difficulty levels of the texts were 94, 93, 88, 86, 84, and 68, respectively. The first and the last sentences of the texts were left intact. Starting from the second word of the second sentence, half of the letters of every second word were deleted. Each text yielded 20 items, so there were 120 items on the whole.

As suggested by Klein-Braley (1997), the test was subsequently given to a control group of 6 EFL teachers. Their scores on the test ranged between a low of 112 and a high of 120, thus over 90% correct on average. Furthermore, it was piloted with a group of subjects similar to the target group. In other words, the 6 texts along with Shiraz University Placement Test were given to 25 Iranian English majors. The correlation between the C-test and the proficiency test was 0.69. The reliability coefficient obtained for the C-test scores as measured through K-R 21 was 0.88. KR-21 is generally considered not to be suitable for estimating the reliability of tests of reduced redundancy, because the items in such tests, unlike multiple-choice ones, are not independent. Yet, Brown (2002), based on a series of studies, claims that K-R 21 only underestimates the reliability of tests with dependent items.

Finally, based on the results of the item analysis, text five which contained a higher number of mal-functioning items, as compared with other texts, was ultimately removed from the test so that the final version

consisted of five texts with 100 deletions. A copy of this C-test will appear in the Appendix.

Procedure

The administration of the C-test and the subsequent interviews were conducted by the first researcher with whom the subjects were taking a course. In order to gain further information about how to efficiently conduct the verbal protocol analysis, before carrying out the main verbal protocol analysis, the C-test was given to another group of English majors and then a retrospective pilot study was done with 7 of them to see what kinds of explanation they might put forward. As such, the researcher could get some idea of how to elicit information from the target subjects.

As for the target group, one session before administering the main C-test, another C-test was given to them to complete in class as a warm up. Then, in the same session, they were asked to say why they had provided each of the responses. Whenever, they failed to do so, the instructor tired to help them by encouraging remarks to continue their explanations so that they finally verbalized the strategy they used. The next session, the main C-test was given to the subjects and they were asked to be attentive to the way they reconstructed the texts.

The retrospection started from the afternoon of the very day the subjects took the test and continued for two days so that all the subjects would take part in it and report the strategies they used for all the items as far as they could. It was carried out in the subjects' native language-- Persian--so that the subjects could explicate exactly what had happened in their mind. It was thought that conducting the explanation in English might have made the task either impossible or very difficult for them to do in some cases. Each subject was given his/her own paper and, starting from the first blank, was asked to say why he/she had given a particular response. Their explanations were tape recorded for later analysis.

Data Analysis

The recorded explanations were transcribed and then coded based on the type of reason(s) the subjects had mentioned for their responses. Codings were singly done by both researchers for each case and were then compared so that a consensus was made on any disagreement. In addition, the frequencies of the different types of reasons were obtained and were then changed into percentages to determine the most frequent ones. To determine if there was any difference between the type and the percentage of the strategies used by the subjects with different levels of proficiency, the top and bottom 25% were identified as the high and low ability groups, respectively. In addition, the percentage of the strategies used in each individual text was obtained once for all the subjects and next for the two proficiency groups.

Results and discussion

The following strategies were detected in the subjects' explanations. They are presented through the classification proposed by Feldmann and Stemmer (1987).

1. S*tructural analysis*

    a. Syntactic analysis

The subject analyses the syntactic structure of the sentence to retrieve a word. For instance, in the case of "*Lions a-- found....*" the subjects usually indicated, "I have used "*are*" because *"Lions"* is plural and the previous sentence is in the simple present tense.

    b. *Formal indicators*

The subject uses a formal syntactic indicator to guess the missing word. For instance, in the case of "*the grass----- of Afr---*" a subject said, "I

wrote the word *Africa* because the word starts with a capital letter. Usually, the names of countries and cities start with a capital letter.

2. Adding letters/syllables to the item beginning

The subject guesses the missing word just by getting help from the undeleted part of the word and adding some letters to it. For example, in the case of *"in Euro---- zoos"* a subject said, "I guessed it was *European* because of the first four letters. It was quite clear."

3. Using past situations

The subject has already seen the word, so just by noting the beginning of the word he/she can retrieve the whole word. For instance, in the case of *"the grass---- of ..."* a subject said, "I had seen the word *grassland* many times, so I easily wrote it." As for some other items, the subject uses the word before or after the item and since he/she has already seen the same two words together, he/she guesses the missing word. For instance, in the case of *"So-- people ..."* a subject said, "In many cases, I have seen the word *people* preceded by *some* before."

4. Translation to mother tongue (translation of immediately following or preceding words)

For instance, in the case of *"In win--- its ... ."* a subject said, "I guessed the word with regard to the meaning of the sentence. It means /dar zemestan/ (in winter).

5. Using the co-text, preceding/following sentence(s) (including the introductory and the   final sentence of the text)

For instance, in the case of *"There a-- no wi-- lions i- Europe, b-- there a--- captive li--- in ...."* A subject said, "I read the sentence preceding and following the missing word and since they seemed to be in contrast, I wrote *but*."

6. Using mother tongue meaning equivalent

The subject, translating the whole sentence, guesses what specific word should be used in his/her mother tongue. Then he/she looks for its equivalent in the target language. For example, in the case of *"... that set-- - low ov-- the ar-- ... ."* a subject said, "Reading the whole sentence, I guessed it should be something like */ruye/ (over)*, so I wrote the word *over*."

7. Using the general meaning of the text

The subject uses the general meaning and idea of the text to restore a missing word. For instance, in the case of *"there a-- captive li--- in ... ."* a subject said, "I guessed the word should be *lion*, since the whole text is about lions."

8. Using external help (other C-test texts, introductory or final part of the text)

The subject retrieves the missing word because he/she has seen the same word in previous texts or in some other tests. For instance, in *"t-- black s- - ... ."* a subject said, "I saw *the* with *black* in the previous text, so I guessed it should be *the* here, too."

9. Using reference

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

The subject guesses the word because he/she has already seen the same word in the text. For example, in the case of *"So-- other people fa--- ... ."* a subject said, "I wrote the word *faint* here because it was used in the first sentence of the text."

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

The subject retrieves the missing word by referring to a lexical item which is related to it and is mentioned in the same text. For instance in the case of *"... in Euro---- zoos."* a subject said, "I wrote *European* since in the previous sentence I had seen the word *Europe*."

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

For example, in the case of "*So-- people fa--- if th-- ... ."* a subject said, "I guessed it would be the word *they* because it refers to the word *people* in the same sentence."

10. Using inference

    a) *Inferring from the meaning of a lexical item/a phrase*

The subject retrieves the missing word by inferring from a lexical item/a phrase mentioned in the text. For instance, in the case of "*... but Ger--- soup ... "* a subject said, "I guessed it should be *German* because in the previous sentence we had the word *Chinese* referring to a country, so I guessed here we must have the name of a country."

    b) *Inferring from the meaning of a sentence*

For instance, in the case of *"... while oth--- like ... ."* a subject said, "Reading the previous sentence and this sentence, I got that the writer is contrasting two groups of people, so since we had *some people* in the previous sentence, I guessed it must be *others*."

11. Juxtaposition

The subject restores the missing word because of its co-occurrence with the preceding and/or the following word (he/she has seen such a combination before). For instance, in *"And th---, as though--- resting o- ... "* a subject said, "I chose *though* because I usually see it coming with *as*."

12. Using background knowledge

The subject uses his/her background knowledge to restore a missing word. For example, in the case of *"So-- people fa-- in crow---."* a subject said, "I guessed it should be *crowds* because usually people faint in busy places."

13. No strategy (automatic processing)

The subject cannot explain why he/she has written a particular item. For instance, in the case of *"... people fa--- if th-- ... "* a subject said, "I don't know why I've written *they*. I just guessed it should be *they*."

In all, 13 strategies were discerned. Feldmann and Stemmer (1987) divide the strategies used to retrieve the C-test items into two groups of bottom-up and top-down strategies. However, it is not possible to draw a clear demarcation line between the two types of strategies. That is, the difference between the two types is a matter of degree, rather than type. Thus, we can say that *background knowledge* is much closer to the top-down end of the continuum, whereas *adding letters/syllables to the item beginning* is closer to the bottom-up end of the continuum. Other strategies, such as *looking for external help*, can be put nearly in the middle of the continuum.

Storey (1997) believes that when the restoration of the deleted item requires reference to material outside the sentence providing the immediate context for the item, it is done at the macro level. In contrast,

when the restoration is done within the immediate context, it is done at the micro level.

Taking into account the two types of categorization proposed by Feldmann and Stemmer (1987) and Storey (1997), the strategies used by the subjects in the present study were divided into two groups: top-down strategies and bottom-up strategies. Thus, *syntactic analysis, using formal indicators, adding letters/syllables to the item beginning, translation to mother tongue, using mother tongue equivalent,* fall within the bottom-up category, whereas *using past situations, using co-text, using the general meaning of the text, using external help, inferring from the meaning of a lexical item/a phrase, inferring from the meaning of a sentence, background knowledge, and juxtaposition* fall within the other category. Besides, strategies related to *reference*, i.e., *referring to the same lexical item, referring to a morphologically related item,* and *substituting a pronoun for a lexical item* may fall within either category depending on whether the item referred to occurs within the same sentence or the preceding sentences. The farther the deleted item is from the item referred to, the closer the strategy used by the subject tends to be to the top-down end of the continuum and vice versa.

As the strategies mentioned earlier indicate, the testees employed different types of strategies in completing the C-test. In fact, this supports the claims made on the construct validity of the C-test as a test of overall language proficiency. The results of the present study are in line with those of Feldmann and Stemmer (1987) and Babaii and Ansary (2001). As Babaii and Ansary (2001) say, "... to the extent that the C-test triggers both macro- and micro-aspects of the language, it confirms well to the principle of reduced redundancy which fundamentally emphasizes that both a global and a local knowledge are required to supply the missing elements in a distorted linguistic message" (p. 216) .

It is also noteworthy to see how frequently any one of these strategies has been used by the subjects. Table 1 shows the percentages of the strategies used by the subjects in this study.

Table 1
Percent strategies used by the subjects

Strategy	Percentage
Translation to mother tongue	26.3
Syntactic analysis	20.3
Adding letters or syllables to item beginning	14.3
Referring to the same lexical item	10.4
Inferring from the meaning of the same lexical item/phrase	9.6
Juxtaposition	4.3
No strategy identifiable	3.5
Using the general meaning of the text	2.8
Inferring from the meaning of a sentence	2
Substituting a pronoun for a lexical item	1.5
Formal indicators	1
Mother tongue equivalent	1
Background knowledge	0.9
Past situations	0.8
External help	0.8
Referring to a morphologically related item	0.5
Using co-text, preceding and/or following sentence(s) (including the introductory and the final sentence of the text)	0.4
Top-down **Bottom-up**	**22.3** **74.2***

*The sum of the percentages of top-down and bottom-up strategies in this table and the following tables excludes "*no strategy identifiable*".

As it is evident from the table, the highest percentage belongs to *translation to mother tongue* (26.3%), which is a bottom-up strategy. The next highest percentage, too, belongs to another bottom-up strategy, i.e., *syntactic analysis* (20.3%). In fact, it can be said that about half of the strategies used by the subjects fall within these two categories. As such, these results are in line with those of Chapelle and Abraham (1990) who claim that C-testing most likely results in tests of more grammatical and less textual competence. The next strategy used by the subjects is *reference to the same lexical item* (10.4%), which is, as mentioned before, a middle-of-the-roader strategy. (Of course, it was found that in 6% of the cases the item referred to was in the sentences other than the sentence in which the deleted item appeared and in 4.4% of the cases in the same sentence). The lowest percentage is that of *using co-text ...* (0.4). On the whole, it was found that 74.2% of the strategies were bottom-up, 22.3% top-down, and 3.5% no strategy. Thus, although both types of processes were employed by the subjects and this confirms the claims made on the C-test as a measure of general language proficiency, it seems that the testees did not complete the C-test as a whole but acted on the individual items independent of each other. This finding is in line with those of Cohen *et al.* ((1984)) and Kamimoto (1992) who state that in C-testing, processing is more at micro-level than macro-level. However, the results are in contrast with those of Dornyei and Katona (1992) who claim that the C-test is quite integrative and the aspect which is less efficiently measured in the C-test is grammar.

Feldmann and Stemmer (1987) believe what a C-test measures seems to vary according to the deletions in the test. Storey (1997), too, holds that the items on a C-test have varying degrees of construct validity. Accordingly, it may be assumed that different texts in the C-test may yield quite different results. In order to verify the above assumption in the case of the C-test utilized in this study, the analysis done earlier for the whole

C-test was done for each individual text of the test. Table 2 illustrates the results.

Table 2
Percent strategies used by the subjects in individual texts of the C-test

Strategy	Text 1	Text 2	Text 3	Text 4	Text 5
Translation to mother tongue	23	27.3	24.6	31	27
Syntactic analysis	26	25.6	13.8	22.5	12.7
Adding letters or syllables to item beginning	17.9	10.3	17.2	12.9	12.9
Referring to the same lexical item	4.7	16.6	9.8	9.9	12
Inferring from the meaning of the same lexical item/phrase	4.5	2.5	21	7.5	13.3
Juxtaposition	4.7	2.5	4	4	5.8
No strategy identifiable	1.8	6.7	2.6	4	2.6
Using the general meaning of the text	5	2.5	1.5	0.15	4.3
Inferring from the meaning of a sentence	1	0.6	1.7	1	4.5
Substituting a pronoun for a lexical item	1.5	3.4	0.8	2	-
Formal indicators	3	-	-	0.15	1.5
Mother tongue equivalent	1.5	0.4	1.1	1.2	0.7
Background knowledge	2.6	1	-	-	0.2
Past situations	0.3	0.1	1.7	0.6	1
External help	-	-	0.04	3.5	0.4
Referring to a morphologically related item	1.7	-	-	-	0.4
Using co-text, preceding/following sentence(s) (including the introductory and the final sentence of the text).	0.8	0.5	0.16	0.05	0.7
Bottom-up **Top-down**	**79** **21**	**73.1** **26.9**	**65.3** **34.7**	**76.2** **23.8**	**65** **34.9**

Translation to mother tongue and *syntactic analysis* are strategies used most frequently in texts one, two, and four.

This, however, does not stand true for texts 3 and 5. In these two texts, although the strategy most frequently used is *translation to mother tongue*, which is a bottom-up strategy, the next highest strategy is *inference from a lexical item/phrase*, which is a top-down one. Interestingly enough, as Table 2 indicates, in these two texts the percentage of the overall top-down strategies is higher than the other ones (34.7 in text 3 and 34.9 in text 5).

As mentioned earlier, the difference observed in the percentage of using different strategies in particular and top-down and bottom-up strategies in general might be due to the nature of the texts and the deleted items. A scrutiny of the deleted items shows that the reason cannot be related to whether the deleted items are content words or function words, because in almost all the texts about 90% of the deleted items are content words. The difference may not be attributed to the readability of the texts, either, since presumably as we proceed from text one to text five the difficulty level of the texts increases.

However, since *inferring from a lexical item/phrase*--which is a top-down strategy--is used much more frequently in these two texts than the other texts, the reason might be the fact that the vocabulary in these two texts has been much easier or the topic has been more familiar to the subjects.

Feldmann and Stemmer (1987) maintain that a skilled reader will activate both top-down and bottom-up processing simultaneously. Yet, the more proficient the subjects are, the more they will be able to use the nature of redundancy of the text. To examine to what extent this idea holds true in the case of the subjects participating in the present study, the top and bottom 25% of the subjects were selected as the high and low proficiency groups. Then, the percentage of the strategies used by either of the two groups on the whole test as well as individual texts was

obtained to see if any difference would be observed. Table 3 shows the results pertaining to the whole test.

Table 3
Percent strategies used by high and low ability groups

Strategy	High	Low
Translation to mother tongue	19.5	34.8
Syntactic analysis	23	17
Adding letters or syllables to item beginning	10.6	14.4
Referring to the same lexical item	12	9.5
Inferring from the meaning of the same lexical item/phrase	13	9.1
Juxtaposition	3.5	3.3
No strategy identifiable	2.2	2.5
Using the general meaning of the text	3.6	2.6
Inferring from the meaning of a sentence	3	1.7
Substituting a pronoun for a lexical item	2.7	1.1
Formal indicators	1.4	0.8
Mother tongue equivalent	1	1.2
Background knowledge	1.4	0.5
Past situations	0.8	0.2
External help	1	0.7
Referring to a morphologically related item	0.8	0.3
Using co-text, preceding/following sentence(s) (including the introductory and the final sentence of the text)	0.5	0.4
Bottom-up	64.5	77.9
Top-down	33.5	19.6

   As Table 3 indicates, in the high ability group the highest percentage is that of *syntactic analysis* (23%), whereas in the low ability group it belongs to *translation to mother tongue* (34.8). The second highest frequent strategy is reversed in the two groups, i.e., *translation to mother tongue* (19.5%) in the high ability group and *syntactic analysis* (17%) in

the low ability group. One justification for this phenomenon can be the fact that since the high ability group are more proficient in all aspects of the language, including grammar, they have attempted to restore the missing items via syntactic analysis, while the low ability group, not being so proficient, have just resorted to the easiest way to restore the items, i.e., translation to their mother tongue. These two strategies, however, are bottom-up processing. Of course, an examination of the third highest strategy and the overall percentage of the strategies used by either group shows that the high ability group tend to use top-down strategies more frequently than the low ability group. The third highest frequent strategy used by the high ability group is *inference from a lexical item/phrase* (13%), which is a top-down strategy and in the low ability group *adding letters/syllables* (14.4%), which is a bottom-up strategy. In addition, a comparison of the overall percentage of the strategies used by the low and high ability groups shows that the percentage of the bottom-up strategies used in the low ability group (77.9%) is higher than that of the high ability group (64.5%). In contrast, the percentage of the top-down strategies used by the high ability group (33.3%) is higher than that of the low ability group (19.6%). These results confirm the idea of Feldmann and Stemmer (1987) indicating that the high ability group tend to use more top-down strategies than the low ability group in restoring the missing items.

In order to determine if there is any difference between the performance of the high and the low ability groups on individual texts of the test, the percentage of the strategies used by either group on each text was determined. Table 4 shows the results.

Table 4
Percent strategies used by high and low ability groups in individual texts of the C-test

Strategy	T1 high	T1 low	T2 high	T2 low	T3 high	T3 low	T4 high	T4 low	T5 high	T5 low
Translation to mother tongue	14.9	25.6	20.3	37.5	19.7	28	24.4	44.6	12	38.3
Syntactic analysis	27.4	21	31.9	18	14.3	15.3	21.8	20.7	14.8	9.8
Adding letters or syllables to item beginning	14.9	19.8	5.8	14.8	16.3	12.7	16	9.1	10	15.8
Referring to the same lexical item	6.5	1.7	18.8	14.8	8.8	12.7	11.8	8.3	14.1	9.8
Inferring from the meaning of the same lexical item/phrase	7.1	6.4	5.8	1.6	26.5	21.2	8.4	7.4	17.4	9
Juxtaposition	4.2	3.5	1.4	3.9	2.7	0.8	2.5	0.8	6.7	7.5
No strategy identifiable	3	1.2	4.3	3.1	1.4	4.2	0.8	3.3	1.3	0.8
Using the general meaning of the text	5.4	7	1.4	3.1	1.4	0.8	-	-	10	2.3
Inferring from the meaning of a sentence	1.2	1.7	1.4	1.6	2.7	0.8	1.7	1.7	8	4.5
Substituting a pronoun for a lexical item	2.4	1.7	5	2.3	2	0.8	4.2	0.8	-	-
Formal indicators	4.2	2.3	-	-	0.7	-	-	-	2	1.5
Mother tongue equivalent	0.6	2.9	0.7	0.8	1.4	0.8	2.5	0.8	-	0.8
Background knowledge	5.4	2.3	1.4	-	-	-	-	-	-	-
Past situation	0.6	-	0.7	-	0.7	0.8	0.8	-	1.3	-
External help	-	-	-	-	0.7	-	4.2	2.5	-	0.8
Referring to a morphologically related item	2.4	1.7	-	-	-	-	0.8	-	0.7	-
Using co-text, preceding /following sentence +(s) (including the introductory and the final sentence of the text)	-	1.2	0.7	-	0.7	0.8	-	-	1.3	-
Bottom-up **Top-down**	**70.3** **26.7**	**79** **20**	**68.1** **27.6**	**86** **10.9**	**59.1** **39.5**	**65.5** **29.9**	**73.2** **26**	**82** **14.7**	**49.5** **49.2**	**80.5** **19**

The performance of the two groups of the subjects shown in Table 4 indicates those texts 1, 2, and 4 follow more or less the same pattern observed in all other cases, i.e., the strategy most frequently used in each case for both high and low ability groups is a bottom-up strategy-- *translation to mother tongue* or *syntactic analysis*--and in all cases the percentage of the top-down strategies used by the high ability group is higher than that of the low ability group. None the less, in the case of texts three and five things are different. As the table illustrates, in both cases the strategy most frequently used by the high ability group is a top-down strategy, i.e., *inference from a lexical item/phrase*. Likewise, the overall percentage of the top-down strategies used by the high ability group in the case of these two texts is much higher than the other texts (39.5% for text three and 49.2% for text five). In contrast to the other four texts, text five is the only case where the percentage of the bottom-up strategies used by the high ability group does not greatly exceed that of the top-down strategies (49.5% and 49.2%, respectively). These findings support the idea proposed by Feldmann and Stemmer (1987) that what a C-test measures varies according to the deletions in the test and the point made by Storey (1997) that items have varying degrees of construct validity.

Conclusions

The results of this study show that the subjects have used 13 different strategies, consisting of both bottom-up and top-down processes. Although the subjects tended to use the bottom-up strategies quite more frequently than the top-down ones for restoring the items, this pattern was found not to prevail throughout the five texts included in the C-test. In other words, depending on the content of the text and the deleted lexical items, the type of strategy used by the subjects and its percentage varied.

It was also noticed that the type and the percentage of the strategies used by the subjects with different proficiency levels on the whole test as well as individual texts were, to some extent, different.

All in all, the results of the study are indicative of the construct validity of the C-test as a test of overall language proficiency. This, none the less, does not mean that all aspects of language are measured equally through a C-test; it all depends on the texts included in the test and the proficiency level of the subjects taking it. In fact, the subjects' knowledge of lower levels of language such as vocabulary and syntax are engaged more while they are restoring the test items.

However, as Grotjahn (1986) states, when reporting retrospectively especially in delayed cases, the subject may convey information that is not related, in one way or another, to the real corresponding activity carried out in his/her mind. In other words, the subject may give some explanation for providing a particular response, but the real cognitive processes carried out in his/her mind might be something quite different. Specifically, since the data for the present study were elicited within two days after the subjects had taken the test, some of the subjects, specially those who were interviewed after a longer lapse of time between the administration of the test and the retrospection might have forgotten why they had provided a particular response and the strategy they reported was not exactly the one they used while doing the test. As such, introspective and/or retrospective verbal protocol analyses with shorter time lapse are needed to verify the results reported here.

References

Adams, M. J., Collins, A. (1970). A schema-theoretic view of reading. In Freedle, R. O. (Ed.), *New Directions in Discourse Processing.* (pp. 1-22). *Norwood* , N. J. Ablex.

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

Brown, J. D. (2002). Do cloze tests work? Or, is it just an illusion? *Second Language Studies,* Vol. 21, pp. 79-125.

Chapelle, C.A., (1994). Are C-tests valid measures for L2 vocabulary research? *Second Language Research*, Vol.10, pp. 157-187.

Chapelle, C., Abraham, R., (1990). Cloze method: What difference does it make? *Language Testing,* Vol. 7, pp. 121- 146.

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

Cohen, A.D., Segal, M. and Wiss, R., (1984). The C-test in Hebrew. *Language Testing,* Vol. 1, pp. 221- 225.

Dornyei, Z., Katona, L., Validation of the C-test amongst Hungarian EFL learners. *Language Testing,* Vol. 9, pp.187-206.

Ericsson, K., Simon, H., (1984). Protocol Analysis: Verbal Reports as Data. Cambridge: Cambridge University Press.

Feldmann, U., Stemmer, B., (1987). Thin_ aloud a_ retrospective da_ in c-te_ taking: diffe_ languages-diff_ learners-sa_ approaches? In Faerch, C., Kasper, C., (Eds.*), Introspection in Second Language Research*. (pp. 251-267). Multilingual Matters, Clevedon, Green, A., (1998). Verbal Protocol Analysis in Language Testing Research. Cambridge: Cambridge University Press.

Grotjahn, R., 1986. Test validation and cognitive psychology: some methodological considerations. *Language Testing,* Vol. 3, pp. 159-185.

Hastings, A. J., (2002). Error analysis of an English C-test: Evidence for integrated processing. In Grotjahn, R. (ed.), Der C-Test. Theoretische Grundlagen und praktische Anwendungen, (Vol. 4). AKS, Bochum, (pp. 53-66).

Hood, M., (1990). The C-test: a viable alternative to the use of the cloze procedure in testing? In: Arena, L., (ed.), *Language Proficiency*. (pp. 173-189.), New York. Plenum Press.

Jafarpur, A., (1995). Is C-Test superior to cloze? *Language Testing*, Vol.12, pp.194-216.

--------, (2001). A comparative study of a C-test and a cloze test. In: Grotjahn, R., (Ed.), Der C-Test. Theoretische Grundlagen und praktische Anwendungen, (Vol. 4). AKS, Bochum, (pp. 21-41).

Kamimoto, T. (1992). An inquiry into what a C-Test measures. *Fukuoka Women's Junior College Studies, ,* Vol. 44, pp. 67-79.

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

----------, (1997). C-tests in the context of reduced redundancy testing: an appraisal. *Language testing,* Vol. 14, pp. 47-84.

Klein-Braley, C., Raatz, U., (1985). A survey of research on the C-test. *Language Testing,* Vol. 1, pp. 134-146.

Microsoft word, (1985)-95. Microsoft Word, Arabic Edition, Version 3.1. Microsoft Corp.

Mochizuki, A., (1994). Four kinds of Texts, their reliability and validity. *JALT Journal*, Vol.16,   pp. 41-54.

Oller, J.W. Jr. (1979). Language Tests at School. Longman Group Ltd., London.

Rakhshanfar, M.R., Jahrudi, H., (n.d.). Selected English Reading Books. Khajeh Nasir Technical College, Tehran.

Sigott, G., (1995). The C-test: some factors of difficulty. *Arbeiten aus Anglistik und Amerikanistik,* Vol. 20, pp. 43-53.

---------, (2002). High-level processes in C-Test taking? In: Grotjahn, R., (Ed.), Der C-Test. *Theoretische Grundlagen und praktische Anwendungen*,(Vol. 4). *AKS, Bochum,* (pp.67-82).

Singleton, D., Little, D., (1991). The second language lexicon: some evidence from university-level learners of French and German. *Second Language Research*, Vol.7, pp. 62-81.

Stemmer, B. (1991) What's on a C-test taker's mind? Mental processes in C-test taking. University of Dr. N. Brockmeyer, Bochum.

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

Appendix (The C-test)

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, 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.