

# Discovering a Hidden Facet of the Impact of Deletion Order On the Comprehension of C-tests

Saeed Mehrpour

Department of Foreign Languages and Linguistics, Shiraz University

Shiraz, 71964-85115, Iran

Tel: 98-711-613-4509

E-mail: smehrpur@rose.shirazu.ac.ir

Received: March 14, 2012

Accepted: April 5, 2012

Published: June 15, 2012

doi:10.5430/wjel.v2n2p36

URL: <http://dx.doi.org/10.5430/wjel.v2n2p36>

## Abstract

Cloze test is widely known as an effective technique in testing reading comprehension. However, since the validity of cloze test was questioned by some language experts, a modification of cloze procedure, called C-test, was then proposed. The present study was meant to explore the effects of the position of letters deleted in a C-test on the comprehension of the test; namely, the restoration of the deleted letters. In a standard C-test, the first half of every second word is deleted but in this study an adapted version of a C-test was employed in which the second half of every other word was deleted. A group of high intermediate students (N = 50) learning English at the Iran Language Institute (ILI) participated in this study. The two different versions of the test (the adapted version and the standard version) were randomly administered to the participants of the study so that some students received the standard version and the others the adapted version. Independent samples t-tests were run to analyze the data and the analysis of the data revealed that students taking the standard version of the test had significantly performed better on the test than those taking the adapted version. This indicated that the order of the deletion of the letters in a C-test would affect the comprehension of the text based on which the C-test was prepared. Further analysis of the data revealed that, interestingly enough, five of the highest scores obtained on the two versions of the test belonged to five students who had taken the adapted version of the test. The results of an interview with these students showed that they were used to reading extra material, especially English newspapers and English short stories a lot and that this had helped them increase their reading proficiency to a great extent compared to the other students, especially those who had taken the standard version of the test.

**Keywords:** Cloze procedure, C-test, Deleted letters, Restoration, Reading comprehension

## 1. Introduction

Teaching and testing reading comprehension have special standings in the field of language teaching and testing both in EFL and ESL situations. Since different language experts have different ideas about what reading is and how it can be tested, tests of reading comprehension take different forms and evaluate different kinds of abilities.

One of the techniques which is used in testing reading comprehension is known as the cloze procedure. This procedure was first introduced as a measure of readability of a text by Taylor (1953) and then was employed by him and others as a measure of reading ability of English native speakers. The scope of the application of cloze testing was then extended so that it was used as a measure of ESL/EFL proficiency by other researches (Jafarpur, 1995). Since the validity of cloze testing was questioned by such language experts as Klein-Braley and Raatz (1984), a modification of cloze procedure, known as C-test, was then proposed. After the emergence of the C-test, a number of studies have been conducted to improve the psychometric features of the test, i.e. their reliability and validity. For instance, Eckes and Grotjahn (2006), employing 843 participants from four independent samples, compared the results of TestDaf (Test of German as a Foreign Language) with those of a German C-test and came to the conclusion that the C-test used was a highly reliable, unidimensional instrument which measured the same general dimension as TestDaf did. In a relatively recent study, Lee-Ellis (2009), using Rasch analysis developed a C-test which demonstrated excellent reliability and validity indexes. The researcher also claimed that the test has the potential to be used as a quick indicator of proficiency. Alderson (2002) contends that C-test is not a proper measure of language proficiency and his rationale is that there are no aural and spoken elements involved in C-test taking, so such tests do not measure oral-aural skills. However, in a recent study, Baghaei et al. (2009) compared conversational C-test passages with written C-test passages and concluded that the conversational

C-test passages were better measures of reading, listening and composite scores of reading and listening than written C-test passages.

Jafarpur (1995, p.194) states, "C-testing involves four or five thematically distinct segments of connected discourse in which the second half of every second word, beginning from word two in sentence two is deleted, and the testees receive credit for exact restoration of deleted letters". The C-test was actually meant to operationalize "reduced redundancy principle" (RRP), that is, testing examinees' language ability when linguistic message is distorted by noise or some other kinds of interference such as deleted words from sentences or deleted letters from words (Spolsky, 1973). In general, disrupting the connectivity of a text affects its processability negatively (Grotjahn, 2002).

The advent of the cloze test and its adapted versions were based on psychological principles fundamental to learning. And the initial deletion of the letters in blank positions of a C-test has psycholinguistic reasons. The reader of a passage, in his attempts to restore the missing elements of the passage, makes conscious use of syntactic, semantic, and pragmatic clues found in the passage (Farhady *et al.* 1994).

Another equally important factor in the comprehension of a cloze passage is the prior knowledge of the reader. The influence of prior knowledge on comprehension of a text is based on a psychological theory known as schema theory. According to this theory, the knowledge already stored in memory, in the form of abstract structures called "schemata", interferes with the interpretation of new information. If new information is incomplete, the reader makes inferences on the basis of the selected schema in order to fill in the missing parts. To put it differently, schema theory assumes that readers use a process of semantic constructivity to create meaning from a written or spoken text, which itself has little or no meaning (Chastain, 1988). Nearly the same processes involved in the comprehension of a mutilated passage such as a cloze passage or a C-test can be observed in a non-linguistic task like doing cross-word puzzles. In doing a cross-word puzzle, the positions of missing letters of the words in the total context of the puzzle, whether they are at the beginning of the words or at the end of them, affect the restoration of them. In like manner, in a standard C-test the initial letters of the words whose second parts are deleted help the reader restore the remaining letters of the words.

## 2. Literature Review

The use of adapted versions of C-tests is not something new and examples of the employment of such versions are cited in the literature. In a study carried out by Cleary (1988), he addressed the problem of poor discrimination of C-tests by means of an experimental C-test variant in which grammatically marked items were deleted to the left rather than to the usual right. The outcomes of the research indicated that the discrimination of the C-test could be boosted by left-hand deletion.

Ikeguchi (1998) conducted a study to determine the ability of the C-tests to discriminate the English levels of students in Japan using two types of C-tests: one type consisting of short passages using different texts, and another type constructed from only one long narrative passage. The results of the study revealed that the former type of C-tests can discriminate levels of English proficiency among Japanese learners of English better than the latter type.

Jafarpur (1995) studied the effect of changing the deletion ratio on the C-test. He developed 20 C-test versions of the same text by using different deletion starts and deletion ratios and found that changing the deletion start and/or deletion ratio produced different C-tests, which tapped different abilities.

Jafarpur (1999b) admitting the shortcomings of his previous study (1995), and questioning "the rule of two", that is the deletion of every second word in a C-test, conducted a new study to shed more light on the issue. He administered 5 versions of a C-test with 3 different deletion starts and a cloze test to 340 Iranians majoring in English as a foreign language. The subjects were divided to 5 groups and each group took one C-test version and the cloze test. The results showed that all the 5 C-test versions were able to discriminate among subjects. He also concluded that not only there is nothing unique about the "rule of two" but also this rule produces a considerable number of nonfunctioning items.

In another study, Heidari (1999) compared a left-hand C-test (having 79 items) with a standard one (having 64 items) and found that the C-test with left hand deletion enhanced the discrimination power of the C-test. However, he did not make any mention of how efficiently his C-test could differentiate between subjects with different levels of proficiency.

Farhady and Jamali (2006) investigated the effect of using various deletion procedures on the C-test. They constructed 5 versions of a C-test from a single text by deleting the first half of every second/third/fourth/fifth, and sixth word. They also developed 5 other versions from the same text by deleting the second half of every second/third/fourth/fifth, and sixth word. The C-tests were then administered to 482 senior EFL students and the results indicated that the different versions of the C-test measured different underlying abilities.

In still another study, Rahmani Anaraki (2007) explored the effects of changing the deletion direction and deletion ratio

on the validity and reliability of the C-test. He administered a test of TOEFL along with two versions of a C-test with a change in deletion ratio (from 2 to 3) and deletion direction (from right to left) to a group of 324 English majors and found that changing the deletion ratio from 2 to 3 could result in an easier C-test with a higher discrimination power and a better criterion-related validity. On the other hand, he concluded that changing the deletion direction (from right to left) could result in a more difficult C-test with a lower discrimination power and a low concurrent validity – a C-test which might be a completely different test measuring a different construct. However, changes in deletion technique did not result in any change in reliability of the C-test.

Taking a look at the studies reported above, one can recognize that all of them have been concerned with quantitative analysis of the data obtained from the administration of different versions of C-tests and no study has been undertaken to analyze such data qualitatively. In fact, the present study was conducted to fill such a noticeable gap in the extant literature by delving into the way the testees reacted to two different versions of a C-test through conducting interviews with them in a follow-up study.

### 3. Objective and Significance of the Study

The present study tries to explore the effects of the positions of the letters deleted in a C-test on the comprehension of the test, namely, the restoration of the deleted letters. In a standard C-test the second half of every second word is deleted but in this study an adapted version of a C-test is employed in which the first half of every second word is deleted.

The main purpose of this study is to see whether the order of deletion of the letters in a C-test affects the comprehension of the test or not. The study holds significance in that it challenges the schema theory and the role of prior knowledge in comprehension of a text through using an adapted version of a C-test. In a standard C-test the first half of the letters in a word whose second half is deleted, functions as known information helping readers to make appropriate guesses based on moving from known to unknown information. However, in the adapted version of the C-test employed in the present study the first half of the letters in a word is deleted which makes it difficult or even impossible for the testees to move from known to unknown information and to make proper guesses at the letters forming the first half of the word.

## 4. Method

### 4.1. Subjects

Two classes of high intermediate English at the Iran Language Institute (ILI) took part in this study. The number of students in each class was 25 (50 subjects in all). This type of sampling was based on the following considerations: 1) the subjects had nearly the same level of proficiency, learning English for two and a half years. 2) They had been taught through the same teaching materials and teaching procedures with teachers having nearly the same level of proficiency and efficiency. 3) To control for the subject variability of the two classes, the means and standard deviations of their scores in the previous term were calculated and compared. The mean and standard deviation of the first class were 77 and 1.948, respectively, and those of the second class were 76.14 and 1.739. Although these two indexes were close enough and could be indicative of the comparability of the two groups, an independent t-test was run to make sure that any possible difference between the two classes was merely due to chance and not any other factor. The results obtained from the t-test proved that the two classes were significantly comparable and represented the same population. Table 1 illustrates the results obtained from the t-test.

<Table 1 about here>

Furthermore, in order to find more support for the comparability of the two groups, they were also given the Oxford Placement Test (OPT) whose result revealed that the two groups enjoyed relatively the same level of proficiency. The following table shows the results obtained from a t-test run on the students' proficiency test scores.

<Table 2 about here>

### 4.2. Instruments

In addition to the proficiency test (OPT) which was used to compare the students' level of proficiency prior to the study and whose reliability based on Cronbach Alpha Coefficient turned out to be 0.89, to conduct the present study, two alternative forms of a C-test were constructed. The text of the C-test was excerpted from Seal (1988, p.64). The passage was a relatively long one and was reduced to a 253 word passage for testing purposes. In fact, an attempt was made to omit parts of the passage which contained redundant information whose deletion would not have affected the coherence of the passage. To construct the tests, two copies of the passage were provided. The first and the last sentence of the passages were left intact and the deletion of the letters applied to words in sentences after and before these two sentences. In one of the C-tests the first half of every second word was deleted and in the other one, the second half of every second word was deleted. In the case of words having an odd number of letters, half of the letters plus one were

deleted. One-letter words such as 'I' and 'a' were ignored and counting began on the next word. In general, the test was a 235 word passage with 100 mutilated words which were viewed as test items and each item was assigned one point.

On the whole, the two tests were entirely identical except for the positions of deleted letters of the words in the tests. Finally the two versions of the test were labeled as standard version (SV) and adapted version (AV). The reliability indexes of the two versions of the test were also calculated again using Cronbach Alpha Coefficient. The index obtained for the standard version was 0.79 and that of the adapted version, 0.91, which are generally viewed as acceptable reliability indices. The reason why the adapted version enjoyed a higher index of reliability can be attributed to the significantly higher standard deviation of the adapted version (17.03) compared to that of the standard version (6.46). Another point which can be claimed with regard to the standard deviation of scores obtained from the adapted version is that due to the higher standard deviation of the scores, this test enjoys higher discrimination power differentiating low, intermediate, and advanced students better.

#### 4.3. Procedure

The constructed C-tests were administered to the subjects in the two classes. Before the administration of the test, a brief explanation was given about the test and the instructions were provided in Farsi. Since cloze tests comprise a sub-test of the subjects' mid-term and final exams in the institute and they were already familiar with cloze tests, the explanation given made them a little test-wise. This would reduce the effect of factors such as strangeness of the test which might otherwise have become an extraneous variable influencing the final results.

The subjects were told that they would not be penalized for spelling errors and every intelligible restored word would be considered correct because, as Oller (1979, p.281) reports, "nil correlations exist between spelling and the ability to process discourse." The time allotted for the completion of the test was 25 minutes. Finally, the test papers were randomly distributed among the subjects. Some subjects received SV test and others AV. The reason why test papers were randomly distributed was that, as already noted, the subjects were found to represent the same population. So it was not necessary to view the two classes as separate samples any more and to administer one version of the test to one class and the other version to the other class. The random distribution of the test papers also eliminated the possible effects of other factors such as cheating; as a result, any possible observed difference between the performances of the subjects could be attributed to chance only.

### 5. Results and Discussion

After the tests were administered to the subjects, the completed C-tests were scored. The mean and standard deviation of the scores of the subjects who had taken the standard version of the test (SV) were 90.12 and 6.463, respectively, and those of the subjects taking the adapted version (AV) were 69.84 and 17.03, respectively. To investigate the statistical significance of the data, an independent t-test was conducted. Table 3 illustrates the results obtained from this analysis.

<Table 3 about here>

The results obtained show that although the subjects were initially found to be comparable and representing the same population, they performed significantly differently on the two different versions of the test. Subjects taking the standard version of the test outstandingly performed better on the test than those taking the adapted version. This indicated that the order of the deletion of the letters in a C-test affects the completion or, in other words, the comprehension of it greatly. Such being the case, the answer to the research question raised above is a positive one. In fact, based on the results of the above t-test, it can be stated that the position of the deleted letters in a C-test can affect the performance of the subjects taking the test and those taking the test with right-hand deletions outperform those taking the test with left-hand deletions.

The findings of the present study are in line with those obtained by Cleary (1988), Jafarpur (1995), Heidari (1999), Farhadi and Jamali (2006), and Rahmani Anaraki (2007) which show that a change in the position of deleted letters in a C-test leads to a change in the performance of language learners on such a test, and that such a test usually has a higher discrimination power.

As mentioned before, this study is an example of a mixed study in which use was made of both quantitative and qualitative methods of data collection and data analysis and so far we have been just dealing with the quantitative aspects of the study. However, as far as the qualitative aspects are concerned, the following points can be of great importance to the interested readers.

After scoring the completed tests, it was observed that, most interestingly, the highest grade (99) belonged to a subject who had taken the AV although the subjects taking the SV had generally performed better on the test (AV subjects' scores ranged from 34 to 99 and those of SV subjects ranged from 76 to 98.). Among subjects taking the AV, there were

four other subjects who had scores ranging from 90 to 97 on the test. These five subjects had performed significantly better on the test than other subjects taking the AV. This interesting observation made the researcher conduct a follow-up study to find the reason behind this hidden aspect.

## 6. The Follow-up Study and Secondary Findings

In the follow-up study, the five subjects already mentioned were identified by the information they had provided on their test papers. Two of these subjects belonged to class I and the other three to class II. It was decided to interview these five subjects. Eventually, they were interviewed individually. First, they were asked some introductory questions and the main questions were posed afterwards. Answering the question "How come you got such a high score on the test?", three of the subjects said that they had a lot of outside reading practices, reading short stories in English. The other subjects said they regularly read English newspapers such as *Tehran Times* and *Iran News*. All the five subjects claimed that they usually had good performance on the cloze tests given to them as a sub-test in the mid-term and final exams of the institute. The data obtained from the interviews revealed that having a lot of outside reading practice and using authentic reading materials such as short stories, newspapers, and magazines will increase the ability to process language holistically and at a higher level. This will, in turn, lead to further development of learners' expectancy grammar in the sense that their ability to complete an item will depend on their ability to predicate the language that appears next (Chastain, 1988). In fact, some language experts (e.g., Jorstad, 1974) maintain that this "learners' internalized grammar of expectancy" is the central core of language competence. In addition, Richards and Schmidt (2002, p. 191) contend that "knowledge of the expectancies of occurrence of language items is made use of in the comprehension of language." This suggests that if language learners are exposed to authentic reading materials and are encouraged to use such materials, it will help them to improve their reading comprehension ability in particular and their language proficiency in general.

### 6.1. Discussion of secondary findings

In fact, it is generally believed that extracting real information from a real text in a new or different language can be very motivating and it can increase students' motivation for learning by exposing them to real language (Guariento and Morely 2001). As for the effect of extensive reading on reading comprehension and language proficiency, Brown (2001) contends that "extensive reading is a key to student gains in reading ability, linguistic competence, vocabulary, spelling, and writing" (p.301). Moreover, researchers such as Gorsuch and Taguchi (2010), Iwahori (2008), Kusanagi (2004), Taguchi et al. (2004), Sheu (2003), Bell (2001), conducting research in EFL contexts such as Japan, Taiwan and Yemen have all reached the conclusion that extensive reading can have a great impact on reading rate and/or general language proficiency of EFL learners.

## 7. Conclusion

The findings of the present research proved that the order of deletion of letters in a C-test- whether the right part of the words are deleted or the left part- affects the comprehension of the C-Test. It was found that the comprehension of a C-test with right-hand deletion was comparatively much easier than the comprehension of a C-test with left-hand deletion. This supports the role of prior knowledge in comprehension of mutilated passages such as cloze tests and C-tests. In case of C-tests with right-hand deletion, the initial letters of the words activate the related schemata and function as prior knowledge about the words and help the reader to retrieve the second part of the words from his/her memory and to comprehend the C-test more easily.

The most interesting finding which might be unique to this study was the conspicuously high performance of some subjects on the C-test with left-hand deletion. The data obtained from the interviews with these subjects revealed that the type of reading materials and the amount of reading practice can affect one's general reading comprehension ability very much. Of course, this finding, due to the small number of students having a performance different from most of the other students should be approached quite cautiously. In fact, other factors might be involved and further study is invited to detect these possible factors.

## References

- Alderson, J. C. (2002). Testing proficiency and achievement: principles and practice. In J. Coleman, R. Grotjahn, & U. Raatz, (eds) *University language testing and the C-test*. Bochum: AKS-Verlag.
- Baghaei, p., Monshi Tousi, M., & Boori, A. A. (2009). An investigation into the validity of conversational C-test as a measure of oral abilities. *Iranian EFL Journal*, 4, 94-109.
- Bell, T. (2001). Extensive reading: Speed and comprehension. *The Reading Matrix*. 1, 1.
- Brown, H. D. (2001). *Teaching by principles: An interactive approach to language pedagogy*. Second Edition. White Plains, NY: Pearson Education.
- Chastain, K. (1988). *Developing second-language skills: theory and practice* (Third edition). New York: Harcourt Brace Jovanovich, Inc.
- Cleary, C. (1988). The C-test in English: left-hand deletions. *RELC Journal* 19, 26-35.  
<http://dx.doi.org/10.1177/003368828801900203>
- Eckes, T. & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, 23 (3), 290-325.  
<http://dx.doi.org/10.1191/0265532206lt330oa>
- Farhady, H., Jafarpur, A., & Birjandy, P. (1994). *Testing language skills: from theory to practice*. Tehran: The Center for Studying and Compiling University Books in Humanities (SAMT).
- Farhady, H. & Jamali, F. (2006). Varieties of C-test as measures of general language proficiency. In H. Farhadi (Ed.), *Twenty-five years of living with applied linguistics: collection of articles* (pp. 287-302).
- Gorsuch, G. & Taguchi, E. (2010). Developing reading fluency and comprehension using repeated reading: Evidence from longitudinal student reports. *Language Teaching Research*, 14, 27-59.  
<http://dx.doi.org/10.1177/1362168809346494>
- Grotjahn, R. (2002). 'Scrambled' C-Tests: eine Folgeuntersuchung. In Rudiger Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Vol. 4, pp. 83-115). Bochum: AKS-Verlag.
- Guariento, W. & Morely, J. (2001). Text and task authenticity in the EFL classroom. *ELT Journal*, 55, 347-353.  
<http://dx.doi.org/10.1093/elt/55.4.347>
- Heidari, A. (1999). *C-test: Solving strategies, changing the position of deleted letters and discriminatory power*. Unpublished master thesis. Shiraz University, Iran.
- Ikeguchi, C. B. (1998). Do different C-tests discriminate proficiency levels of EL2 learners? *JALT Testing & Evaluation*, 2, 1, 2-10.
- Iwahori, Y. (2008). Developing reading fluency: A study of extensive reading in EFL. *Reading in a Foreign Language*, 20, 1, 70-91.
- Jafarpur, A. (1995). Is C-testing superior to cloze? *Language Testing* 12, 194-216.  
<http://dx.doi.org/10.1177/026553229501200204>
- Jafarpur, A. (1999b). What's magical about the rule-of-two for constructing C-tests? *RELC Journal*, 30, 86-100.  
<http://dx.doi.org/10.1177/003368829903000205>
- Jorstad, H. L. (1974). Testing as communication. In G. A. Jarvis (Ed.) *The challenge of communication*. Skokie, III.: National Textbook, pp. 223-73.
- Klein-Braley, C. & Raatz, U. (1984). A survey of research on the C-test. *Language Testing*, 1, 134-146.  
<http://dx.doi.org/10.1177/026553228400100202>
- Kusanagi, Y. (2004). The class report 2: Course evaluation of Pleasure Reading Course. *The Journal of Rikkyo University Language Center*, 11, 29-42.
- Lee-Ellis, S. (2009). The development and validation of a Korean C-test using Rasch analysis. *Language Testing*, 26, 245-274. <http://dx.doi.org/10.1177/0265532208101007>
- Oller, J. W. jr. (1979). *The Tests at School: A Pragmatic Approach*. London: Longman.
- Rahmani Anaraki, R. (2007). *Effects of changing the deletion direction and deletion ratio on the validity and reliability of the C-test*. Unpublished master thesis. Shiraz university, Iran.

- Richards, J. C., and Schmidt, R. (2002). *Longman dictionary of language teaching and applied linguistics*. Third Edition. Essex: Longman.
- Seal, B. (1988). *Vocabulary Builder 2*. London: Longman.
- Sheu, S. P. H. (2003). Extensive reading with EFL learners at beginning level. *TESL Reporter*, 36, 8-26.
- Spolsky, B. (1973). What does it mean to know a language? Or how do you get someone to perform his competence? In: J.W. Oller, Jr., & J.C.Richards, (Eds.), *Focus on the learner* (PP. 164-176). Rowley, MA: NewberyHouse.
- Taguchi, E., Takayasu-Maass, M., & Gorsuch, G. J. (2004). Developing reading fluency in EFL: How assisted repeated reading and extensive reading affect fluency development. *Reading in a Foreign Language*, 16, 1-23.
- Taylor, W. L. (1953). Cloze procedure: a new tool for measuring readability. *Journalism Quarterly*, 30, 415-433.

Table 1. The results of the t-test on students' previous term scores

Groups	No of S's	Mean	S.D.	d.f.	t.obs	t.value	2-tail Sig.
Cont.	25	77.00	1.948	48	1.65	2.011	0.83
Exp.	25	76.14	1.739				

P&lt;.05

Table 2. The results of the t-test on students' proficiency test scores

Groups	No of S's	Mean	S.D.	d.f.	t.obs	t.value	2-tail Sig.
Cont.	25	51.5	1.653	48	1.47	2.020	0.81
Exp.	25	53	1.719				

P&lt;.05

Table 3. The results of t-test on the students' performance on the two C-tests

Groups	No of S's	Mean	S.D.	d.f.	t-obs.	t-value	2-tail sig
SV S's	25	90.12	6.46	48	2.568*	2.011	0.00
AV S's	25	69.84	17.03				

P&lt;.05