

## The C-Test and the MC-Test

Patchara Wonghiramsombat

*Language Institute  
Thammasat University*

patcharaw@gmail.com

### Abstract

The C-Test, which was introduced in 1981, has been used to measure reading comprehension. The test has been adapted and modified into many forms with a variety of deletion methods, starting points, and scoring procedures, resulting in a number of strengths and weaknesses. Due to the weaknesses of the C-Test, an adaptation of the C-Test, the MC-Test, was introduced in 1987. Like the C-Test, the MC-Test has been adapted, modified and used in many research papers. The results show that the MC-Test has more strengths than weaknesses.

[Thammasat Review, Special Issue, 2013]

**Keyword:** The C-Test, Adapted, Weakness, Strength

## C-Tests

The C-Test, an adaptation of the cloze test, was introduced by Klein-Braley&Raatz in 1981 (Klein-Braley, 198k; Klein-Braley&Raatz, 1984; Baatz, 1985; Raatz& Klein-Braley, 1985; cited in Dornyei&Katoan, 1992). It is based on the following criteria:

1. The C-Test should be much shorter and should have at least 100 items.
2. The deletion rates and starting points of deletion should be fixed.
3. Only exact word scoring method should be employed.
4. There should be a number of different texts.
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 test (Raatz, 1985, cited in Klein-Braley, 1997, p. 63-64)
7. The test should have high reliability and validity (Klein-Braley&Raatz, 1984, p. 136).

In the C-Test, the second half of every second word is deleted and the deletion begins in the second sentence. A complete sentence at the beginning of the test and another one sentence at the end of the test are left intact.

Raatz & Klein-Braley (1981) conducted a study to investigate the using of English versions and German versions of the C-Test. The subjects were divided into two groups. The first group was pupils in a comprehensive school in Leeds, the freshmen in the English Faculty of the University of Duisburg, and English schoolchildren in the third year of their junior school at Bransley. The pupils at Leeds and the freshmen of the University of Duisburg were administered the English senior version test while the other group took the English junior version test. The results showed that the test was surprisingly highly reliable and valid. For the senior version test taken by the pupils in Leeds, no one made perfect scores. Besides, only 7 out of 125 items could have viable alternatives.

In the second group, the subjects were native speakers of German undergraduates, learners of German, and German schoolchildren in their third grade, who took the German version test. It appeared that 98 out of 100 items had no alternative solution. Again, the test proved to have high reliability and validity. Unlike the results of the first group, some subjects made perfect scores, which showed that the educated native speakers could virtually obtain perfect scores.

The reliability of the C-Test was also proved by Segal (1983; cited in Cohen, Segal and Bar-Simon-Tov, 1984). The study was to investigate how reliable a Hebrew C-Test would be, how well it would discriminate better students from poorer ones, and how well it would correlate with other tests of language ability. The study was conducted with 42 immigrant teachers in training. The tests used in this study were a C-Test, a grammar test, and a reading comprehension test. The results indicated that the Hebrew C-Test was highly reliable and discriminated well. The Hebrew C-Test highly correlated with the test of grammar and moderately correlated with the reading comprehension test. Moreover, Segal found that the C-Test provided a more thorough assessment of connectives, idiomatic expressions and knowledge of different language registers than the cloze test. However, the semantic clues encouraged guessing on the Hebrew C-Test more than on the English C-Test.

Like Segal (1983), Weiss (1983; cited in Cohen, Segal & Bar-Simon-Tov, 1984) investigated how reliable a Hebrew C-Test would be, how well it would discriminate better students from poorer ones, and how well it would correlate with other tests of language ability. The subjects were 20 intermediate and advanced non-native students of Hebrew. All subjects took a C-Test and a cloze test. Both the C-Test and the cloze test were adapted from the same passages. The results showed that the C-Test was highly reliable and discriminated well. The C-Test correlated with the cloze test.

Later in 1984, Klein-Braley&Raatz conducted a study to validate the C-Test. They constructed C-Tests in German, Turkish, Hebrew, English, French, and Spanish, which were administered to three different groups. The subjects in the first group were German and Turkish children who were learning their native languages. The German and Turkish test versions were administered to these subjects. The second group was people learning a language spoken in the community around them but that language was not the language spoken at home. They were Turks, Greeks and others, and were administered the German and Hebrew test versions. The last group was foreign language learners at school or university, who took the German, English, French, and Spanish test version. The results showed that the tests were highly reliable and valid in general. Some cases indicated unsatisfactory validity coefficients due to the lack of variance in the criterion, not the C-Test.

While Klein-Braley&Raatz (1984) found low validity in some cases in the C-Test, Klein-Braley (1985) sought to establish the validity of the four hypotheses:

1. If the same C-Test is administered to subjects at different stages of language development, then the C-Test scores will become successively higher as the subjects become more proficient in the language. (p. 85).
2. Subjects learning a language ‘naturally’ will exhibit similar behavior on C-Tests in that language. (p. 86).
3. If texts have an inherent ‘C-Test processing difficulty, which is dependent of the subject groups involved then it will be possible to discover characteristics of the texts which can be used to predict the rank order of difficulty of texts, possibly even the actual empirical difficulty levels, for specific subject groups. (p. 88).
4. Learners with more efficient language processing strategies will make higher scores on C-Test. (p. 97).

Hypotheses 1, 2 and 3 were validated. However, the first hypothesis was validated by L1 speakers in childhood and adolescence only, while the second hypothesis was validated by L1 speakers in childhood in a German host environment. The third hypothesis was validated by L1 speakers in childhood and adult FL learners from highly inflectional L1 backgrounds.

Nevertheless, the fourth hypothesis has remained to be proven. This was because the correlation between the C-Test scores and the general intelligence of younger children, and the correlation between the C-Test scores and the general intelligence of older children were low and medium, respectively. Raatz (1984; cited in Klein-Braley, 1985) also found low and medium correlations between the C-Test results and non-verbal intelligence tests. Klein-Braley concluded that the results of all four hypotheses were sufficient evidence to substantiate the claim that the C-Tests were authentic tests of the construct of general language proficiency.

Satisfactory results were found in McBeath's report (1990) as well. He reported the results of the C-Test used as a final examination in Oman's Air Force. The results suggested that the C-Test was able to discriminate between students who had acquired an overall competence and those who had serious weaknesses.

Dornyei & Katona's study (1992) provided positive results. The study was carried out in a research program to replicate the very positive results reported in the literature to see to what extent they could be generalized. The Hungarian EFL learners were divided into two groups: 120 first-year English major university students and 53 secondary pupils. There were five subtests in the study which were the Department Proficiency Test, the Test of English for International Communication (TOEIC), the Oral Interview, the cloze test, and the C-Test. The first-year English major university students took all five subtests while the secondary pupils took TOEIC and the C-Test only. The results showed that the C-Test was a reliable and valid instrument among Hungarian learners. It was found to have high concurrent validity and to measure global language proficiency. In comparison to the cloze test, the C-Test appeared to be a better measure of language proficiency. Another result suggested that the truncated structure words were easier to reconstruct than the truncated content words.

Unlike Dornyei & Katona (1992), Stemmer (1992) found the opposite result. The assumption was that C-Test solving was a cognitive task. He also investigated the way learners of French tried to solve the C-Test. This study employed the verbal report techniques such as thinking aloud and intermediate retrospection as data collection procedures. The data were analyzed in three components: (a) a task analysis comprising item and text analysis, (b) an analysis of the verbal protocol comprising strategic problem-solving behavior, and (c) an analysis of the performance data. The findings from the various analyses suggested that the C-Test did not measure high level comprehension and could not be regarded as a measure of general language proficiency.

Another contradictory result was found in the C-Test as well. Jafarpur (1995) investigated the feasibility of the procedure with native and non-native speakers of English. Native speakers and non-native speakers of English took the C-Test, and the cloze test, and filled out a questionnaire. The C-Test consisted of 20 versions varying in different deletion ratios and starting points. It appeared that changing the deletion ratios and starting points produced different C-Tests, which affected the validity of the test.

Connelly (1997) used English C-Tests with postgraduate students at the Asian Institute of Technology (AIT) in Thailand. The same test was given as both a pre- and post-test to help maintain the reliability. The results indicated that the C-Test had high validity and reliability. Furthermore, Connelly claimed that the C-Test was a useful tool for measuring general language proficiency.

Like Connelly (1997), Klein-Braley (1997) discovered positive outcomes. She compared the empirical performance of the C-Tests with other representatives of the family of reduced redundancy tests, which were the classical cloze test, the cloze-elide test, and the multiple-choice cloze test. The Duisburg English Language Test for Advanced students was used as the criterion for empirical validity. The findings revealed that the C-Test emerged as the most economical and reliable procedure and had the highest empirical validity.

An adaptation of the C-Test was performed by Cleary (1988). The test employed in the study consisted of two versions: the standard version and the adapted version, called the variant version. In the standard version, all items (103) were deleted on the right hand. In the variant version, 63 items, which were all grammatical unmarked, were deleted on the left hand. It was discovered that the mean score of the standard version was higher than that of the variant version, which made the variant C-Test discriminate better than the original C-Test. The author also found that the C-Test became unstable when it was applied to learners at a low level of attainment, which meant that the C-Test lacked validity. Cleary claimed that the lower-level learners did not have enough ability in discourse for general proficiency which the C-Test purported to measure.

Mitchell (1991) conducted research on the C-Test to investigate whether the C-Test would be a suitable replacement for the City Polytechnics' Post-Entrance English Language Proficiency Test (Cipoldex). It was found that the C-Test discriminated adequately, but no better than the Cipoldex. The reliability of the C-Test appeared to be better than the Cipoldex. However, the correlation between the C-Test and the Cipoldex was moderate. The author concluded that the C-Test should not replace the Cipoldex.

Results from the studies reviewed can be summarized as follows:

The advantages of the C-Test

1. The C-Test is easy to construct and to score (e.g., Connelly, 1997; Klein-Braley&Raatz, 1984).
2. The C-Test is highly reliable and valid (e.g., Connelly, 1997; Dornyei&Katona, 1992; Klein-Braley, 1997; Weiss, 1983).
3. The C-Test is economical (e.g., Weir, 1990 1993).

Disadvantages of the C-Test

1. The C-Test lacks face validity (e.g., Jafarpur, 1995; Mitchell, 1991)
2. The C-Test cannot be used satisfactorily in measuring high level comprehension (e.g., Sigott&Koberl, 1993; Stemmer, 1991).

3. The C-Test processing is critical in terms of psycholinguistic completing strategies (Klein-Braly, 1997).

#### MC-Tests

Due to the disadvantages of the C-Test, a new test form called the MC-Test or the X-Test was introduced. The MC-Test was first initiated by Boonsathorn in 1987 (cited in Boonsathorn, 1988, 1990). The principle of the MC-Test construction is based on the psychological viewpoint of Goodman (1967; cited in Boonsathorn, 1988), which claims that in the processing of reading, the readers simultaneously engage in all levels of processing – graphophonic, syntactic, and semantic. Unlike the C-Test, the MC-Test deletes the first half of every second word. The deleted parts contain the semantic information of the words, while the remaining parts provide syntactic/structural information.

Boonsathorn (1987; cited in Boonsathorn, 1988) compared the C-Test with the MC-Test to establish the reliability of the C-Test and The MC-Test. Besides, he proposed to verify whether the different starting points of deletion would affect the difficulty, reliability, and validity of the test, and to explore the strategies which L2 learners used in restoring the C-Test and The MC-Test passages. The author constructed two forms of the C-Test and another two forms of the MC-test, which were administered to both L1 and L2 subjects. Only the exact word scoring procedure was employed. The results showed that the C-Test and The MC-Test were essentially different functionally and structurally. Both types of tests were highly reliable and valid for L1 and L2 subjects. It appeared that the MC-Test was more difficult and discriminated better than the C-Test for both L1 and L2 subjects. This was because the MC-Test required more of the normal reading process than did the C-Test. However, different starting points did not affect the difficulty, reliability, and validity of the tests.



Sigott and Koberl (1993) also made a comparison between the C-Test and the MC-Test. They formulated three hypotheses: (a) The results of the MC-Tests correlated highly with the results of other reduced-redundancy tests, (b) the results of the MC-Test did not correlated highly with the results of tests that claimed to measure language proficiency above sentence level, and (c) for language of English, the MC-Tests were more difficult than the C-Tests. The subjects were first-semester students of English who took a cloze test, two C-Tests, two MC-Tests, a paragraph sorting task, and a multiple-choice reading comprehension. Both exact word scoring and acceptable word scoring methods were used. The results confirmed all hypotheses. It appeared that the MC-Test seemed to measure the same competence as the cloze and the C-Test, which did not measure language competence beyond sentence level. The MC-Test was found to be more difficult than the C-Test; therefore, it might be profitably used instead of the C-Test in measuring higher levels of competence.

Koberl&Sigott (1996) conducted a study to examine whether native speakers and foreign learners, who might differ from each other in the structure of their language competence, actually used the MC-Test solving strategies. As a result, two hypotheses were formulated: (a) The difference between the C-Test and the MC-Test scores was not influenced by whether the subjects were native speakers or foreign learners, and (b) the ranking of item facilities was not influenced by whether the subjects were native speakers or foreign learners. The C-Test and the MC-Test were administered to the sixth-form native speaker pupils and German learners of English. The results showed that the areas of language competence assessed by the C-Test and the MC-Test were similarly structured in the mother tongue and the foreign language. It also showed that the C-Test and the MC-Test were equally suitable for establishing the text difficulty for native speakers and for judging foreign language learner's competence.

A new aspect of the MC-Test investigation was done by Prapphal (1994). She investigated the order of presenting the C-Test and The MC-Test, as well as the reliability and validity of the tests. The two forms of the tests were constructed from one single passage. They were given to Thai science-oriented first-year students in the Faculty of Engineering and the Faculty of Science at Chulalongkorn University. The subjects were randomly divided into two groups. The first group took the C-Test first and the MC-Test one week later. The reverse order was employed with the other group. The results were that both tests were reliable and concurrently valid. The MC-Test seemed to be more closely related to the cognitive and academic skills than the C-Test. In terms of the order of presenting the tests, students scored the same when the C-Test was presented before the MC-Test. On the other hand, the subjects got low scores on the C-Test when the MC-Test was presented before the C-Test. The author reasoned that students might differ in the retention and reading processing skills.

Using the same method, Prapphal (1996) investigated the role of the MC-Test in the following aspects: (a) whether the MC-Test could be used as a means to investigate the ability to transfer language competence from General English to Academic English, (b) if the MC-Test could measure transferable ability, which grammatical competence it could measure better, syntactic competence or lexical competence, and (c) whether there was any difference in the retention of grammatical competence as measured by the General MC-Test and the Academic MC-Test. There were two MC-Tests in the study. The first MC-Test was considered as General English while the second as Academic English. The subjects who got the General English MC-Test as the pre-test got the Academic English MC-Test as the post-test, and vice-versa. The results were that the MC-Tests were reliable and concurrently valid with the English for Academic Achievement Tests. The MC-Test could be used as a means in examining the ability to transfer language competence from General English to Academic English. Besides, the MC-Test could assess lexical competence better than syntactic competence. With respect to retention, they could be used in measuring the subjects' ability to retain grammatical competence both in General English and Academic English.

## Conclusion

The MC-Test has been proved to be highly reliable and valid. Due to the more appropriate level of difficulty, the MC-Test appears to discriminate better than the C-Test and to be suitable for measuring higher levels of competence.

## References

- Aborn, M., Rubenstein, H., & Sterling, T. D. (1959). *Sources of contextual constraint upon words in sentences*. Journal of Experimental Psychology, 57, 171-80. (pp. 205-217). Massachusetts: Newbury House.
- Anderson, J. (1971). *Selecting a suitable 'reader': Procedures for teachers to assess language difficulty*. RELC Journal, 2, 35-42.
- Bachman, L. F., Palmer, A. S. (1996). *Language testing and practice: Designing and developing useful language tests*. Oxford, the United Kingdom: Oxford Press.
- Boonsathorn, S. (1987). *C-Tests, proficiency, and reading strategies in ESL*. Unpublished doctoral dissertation, University of Alberta, Canada.
- Boonsathorn, S. (1988). *Cloze test revisited: A validation and implications of the C-test/MC-test*. In *Thai/TESOL (Ed.)*, The eighth Thai/TESOL convention proceedings (pp. 303-324). Bangkok: Thai/TESOL.
- Boonsathorn, S. (1990). *C-Test/MC-Test: A new instrument for measuring English proficiency*. PASAA, 20(2), 46-50.
- Carlson, N. R. (1988). *Discovering psychology*. Boston: Allyn and Bacon.
- Cleary, C. (1988). *The C-Test in English: Left-hand deletions*. RELC Journal, 19(2), 26-38.
- Cohen, A. D., Segal, M., & Bar-Simon-Tov, R. (1984). *The C-Test in Hebrew*. *Language Testing*, 1(2), 221-225.
- Coniam, D. (1993). *C-Test or no text: A study of an adapted cloze technique for testing comprehension/summary skills*. Hong Kong papers in linguistics and language teaching, 16, 1-10.

- Connelly, M. (1997). *Using C-Test in English with post-graduate students. English for Specific Purposes*. 16(2), 139-150.
- Davies, A. (1985). *Simple, simplified, and simplification: What is authentic* In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 181-198). London: Longman.
- Dornyei, Z., & Katona, L. (1992). *Validation of the C-Test amongst Hungarian EFL learners. Language Testing*, 9(2), 187-206.
- Fry, E. (1977). *Fry's readability graph: Clarifications, validity, and extension to level 17*. In L. E. Burmeister (Ed.), *Reading strategies for middle and secondary school teachers* (pp. 34-37). Massachusetts: Addison-Wesley.
- Goodman, K. S. (1967). *Reading in Psycholinguistic guessing game*. In F. V. G. Goodman (Vol. Eds.), *Language and Literacy: Vol. 1. The selected writings of Kenneth Goodman*. (pp. 33-43). London: Routledge & Kegan Paul.
- Heaton, J. W. (1975). *Writing English language tests*. London: Longman.
- Jafarpur, A. (1995). *Is C-Testing superior to cloze?* *Language Testing*, 10(1), 195-216.
- Klein-Braley, C. (1997). *C-Test in the context of reduced redundancy testing: An appraisal*. *Language Testing*, 14(1), 47-84.
- Kober, J., & Sigott, G. (in press). *Deletion patterns and C-Test difficulty in German*. *Moderne Sprachen*.
- Kober, J., & Sigott, G. (1996). *Deletion patterns in C-Test: Native speakers vs. foreign learners*. *Language Testing Update*, 20, 45-48.
- Mitchell, R. (1991). *Validating language tests: A Hong Kong case study* (Rep. No. 10). Hong Kong: City Polytechnic of Hong Kong, Department of English.
- Mondy, R. W., & Noe III, R. M. (1993). *Human resource management* (5<sup>th</sup> ed.). Boston: Allyn and Bacon.

- Oller, J. W., Jr. (1979). *Language tests at school*. Longman. Perfect Copy Classic 1.18 [Computer software]. (1995). New York: Logicus Incorporated. 341.
- Prapphal, K. (1994). *A study of the C-Test and the X-Test performed by first-year science-oriented university students*. PASAA, 24, 16-23.
- Prapphal, K. (1996). *An investigation of the general and academic English X-Tests in measuring grammatical competence of Thai science students*. PASAA, 26, 73-79.
- Prapphal, K., Pas. K., & Tanapongpipat, S. (1984). *Modified cloze tests as predictive measures of language acquirer and academic achievers*. PASSA, 14,(2), 75-89.
- Raatz, U. (1984). *The factorial validity of C-Tests*. In T. Culhane, C. Klein-Braley, & D.K. Stevenson (Eds.), *Practice and problems in language testing* (Department of Language and Linguistics Occasional Papers No. 26). Colchester, the United Kingdom: University of Essex.
- Raatz, U. (1985). *Test of reduced redundancy – the C-Test, a practical example*. In C. Klein-Braley, & U. Raatz (Eds.), *C-Test in der Praxis. Fremdsprachen und Hochschule, AKS-Rundbrief, 13/14* (pp. 14-19). Bochum, West Germany: ArbeitskreisSprachenzentrum [AKS].
- 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 problems in language testing* (pp. 113-138). Colchester: University of Essex.
- Raatz, U., & Klein-Braley, C. (1985). *How to develop a C-Test*. In C. Klein-Braley, & U. Raatz (Eds.), *C-Test in der Praxis. Fremdsprachen und Hochschule, AKs Rundbrief, 13/14* (pp. 20-22). Bochum, West Germany: ArbeitskreisSprachenzentrum [AKS].

- Segal, M. (1983). *The C-Test (Seminar Paper in Hebrew)*. Jerusalem, Israel: Hebrew University, School of Education.
- Sigott, G., & Koberl, J. (1993). Validating the X-Test Language Testing Update, 14, 53-58.
- Stemmer, B. (1992). *What's on a C-Test taker's mind? Mental processes in C-Test taking*. Language Testing Update Abstract, 11, 42-43.
- Stillman II, R. J. (1996). *Public administration (6<sup>th</sup> ed.)*. Boston: Houghton Mifflin.
- Tannembaun, P. H., Williams, F., & Clark, R. A. (1969). *Effects of grammatical information on word predictability*. Journal of Communication, 19, 41-48. 19-26.
- Weir, C. J. (1990). *Language testing*. New York: Prince Hall.
- Weir, C. (1993). *Understanding and developing language tests*, Hertfordshire, the United Kingdom: Brince.
- Weiss, R. (1983). *Hebrew The C-Test (Seminar Paper in)*. Jerusalem, Israel: Hebrew University, School of Education.