



The Cloze Test and the C-Test

Patchara Wonghiransombat

Language Institute, Thammasat University

ABSTRACT

The Cloze Test, which was introduced in 1953, was first used to measure the readability of the written materials, and then was applied to first language testing for 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 Cloze Test, an adaptation of the Cloze Test, the C-Test, was introduced in 1981. Like the Cloze Test, the C-Test has been adapted, modified and used in many research papers. The results show that the C-Test has strengths as well as weaknesses.

Keywords: *Cloze test, C-test, Reading comprehension, Language testing*

INTRODUCTION

This article reviews the related literature of the Cloze test and the C-Test.

CLOZE TEST

The cloze procedure was introduced by Taylor (1953 cited in Culhane, 1970). The principle of the cloze procedure is based on the gestalt concept of ‘closure’ which means the ability to fill in the parts in an incomplete message (Culhane, 1970; Heaton, 1975). The cloze was first invented to measure the readability of the written materials, and then was applied to first language testing for reading comprehension (Heaton, 1975), and for overall ESL proficiency (Brown, 1993; Weir, 1990). Culhane (1970) suggested that the cloze test should be used to the benefit of both teachers and pupils. The cloze test was used as the teaching device for teachers to improve the students’ reading skills. The pupils were forced to read more carefully, to use



contextual clues, and to become actively involved with what they were reading. In addition to the original cloze test, the test has been adapted and modified into many forms such as, the multiple-choice cloze test (Brown, 1980; Prapphal & Tanapongpipat, 1984), the matching cloze test (Prapphal & Tanapongpipat, 1984), the conversational cloze test (Lennon, 1989), and the summary cloze test (Coniam, 1993).

- ***Deletion Methods***

The original deletion method of the cloze test is the systematic deletion procedure, in which every n th word is deleted (where n is normally a number between 5 and 10) (Klein-Braley, 1997).

Another method invented by Bachman (1985) is the rational deletion procedure. He conducted a study to develop criteria for rationally selecting which words to delete in developing a cloze test. The deleted words were selected according to these criteria: (a) within clause; (b) across clause, within sentence; (c) across sentence, within text; and (d) extratextual.

Alderson (1978, 1979, 1980, 1983 cited in Davies, 1985) found that changing the deletion rates of the test produced a different test which seemed to measure different abilities and also affected the validity of the test.

Farhady & Keramati (1996) did research of cloze tests with a text-driven method for the deletion procedure. The deletion rates were considered on the basis of the number of linguistic and discourse structures in the passage. The results showed that the test-driven method was superior to the fixed-ratio deletion method in terms of reliability.

- ***Stating Points***

Unlike the changing of the deletion rate, Brown (1983) conducted a study to investigate whether the cloze tests that were based on deleting every n th word with different starting points would be able to tap equally well whatever cohesive devices existed in the prose. The results showed that different starting points did not change the role of the cohesive devices.



Later in 1988, Brown (1988a) found different results from his previous study. The cloze test in his 1988 study consisted of two forms which were constructed from the same passage. The only different criterion was the starting point. The original cloze test began deletion at the 7th word while the ‘tailored’ cloze test’s starting point depended on the results of item facility and item discrimination analyses of the five pilot versions. The tailored cloze was selected in terms of those items that closely approximated 50% in item facility, and (b) those items with the highest discrimination indices. The results showed that the tailored cloze test appeared to be considerably more reliable and valid than the original cloze test. Moreover, the tailored cloze test yielded higher scores than the original one. It can be concluded that the dispersion of scores, reliability, and validity were improved by the item analysis and selection processes.

- ***Scoring Procedures***

There are four major scoring procedures to score the cloze test. The first procedure introduced at the same time as the cloze test was invented is the exact word scoring method. Taylor (1953) explained that “the ability of a reader to fill in the very word used by a writer would be a suitable index of the degree of correspondence between the language system employed by the writer and that employed by the reader” (p.367 ,as cited in Oller, 1979).

The second scoring procedure is acceptable word scoring method, which accepts either the exact word or any word that is contextually acceptable.

According to Hinofotis (1987), the third scoring method, called the clozentropy scoring procedure, was the one developed by Darnell (1968). This method is a complex mathematical procedure which involves weighing the nonnative speaker’s responses with the native speaker’s responses. Since the procedure is complex, it is impractical for use in most testing situations. This method has been later modified mathematically by Reilly in 1971 (Brown, 1980).

The last method is the multiple-choice scoring method. A choice of words is provided for each blank from which the students must select the appropriate



response. The multiple-choice scoring method is probable testing of receptive skills while the exact word scoring, acceptable word scoring, and clozetrophy scoring methods measure productive skills.

The first and the second scoring methods have been examined by many researchers. Anderson (1971) found that the exact word scoring and the acceptable word scoring procedures produced the same results (as cited in Aitken, 1977). On the contrary, Alderson's studies (1979, 1983) indicated that changing the scoring procedures resulted in a different test validity. However, Stubbs & Tucker (1974, as cited in Hinofotis, 1987) indicated that there seemed to be very little difference when the exact word scoring method was substituted for the acceptable word scoring method. The acceptable word scoring procedure appeared to be the most valid procedure for the purpose of EFL language testing (Oller, 1972) and to be superior to the exact word scoring procedure (Porter, 1978). Hinofotis (1976, as cited in Hinofotis, 1987) reported that the acceptable word scoring method yielded more reliable scores and provided more accurate information about ESL proficiency levels.

In Hinofotis's study (1980), it appeared that the exact word scoring and the acceptable word scoring procedures were not equally reliable. The acceptable word scoring method was a more accurate assessment of the student's EFL ability than the exact word scoring method. Klein-Braley (1983) found the opposite results. She found that the exact word scoring procedure was more reliable than the acceptable word scoring method.

There have been few studies that examined all four scoring procedures. However, Brown (1980) did research to compare all four scoring methods. He found that there were differences among the four scoring methods in reliability, mean item facility and discrimination, as well as in usability. The results indicate that the best overall scoring method was that acceptable word scoring procedure. However, Brown suggested that the decision about which method to use would vary with the testing situation.



Oller (1972) examined the effectiveness of five different scoring methods. The first method was the exact word scoring method and the second was a combination of the exact word scoring and the acceptable word scoring methods. The other three scoring methods weighed responses in several different ways according to the extent to which they fit the context. Five categories of responses were differentiated: (a) exact words, (b) acceptable words, (c) responses that violated long-range constraints, (d) responses that violated short-range constraints, and (e) entirely incorrect fill-ins or item left blank indicating complete lack of comprehension. The results indicated that the best of the five scoring methods was the method that combined the exact word scoring and the acceptable word scoring methods. The data showed that the acceptable word scoring method was superior in terms of item discrimination and validating correlation regardless of the level-of-difficulty of the test.

- ***What Does the Cloze Test Measure?***

The cloze procedure has been used to measure English language proficiency (e.g., Aitken, 1977; Brown, 1993; Caulfield & Smith, 1981; Chavanachat, 1986; Fotos, 1991; Oller & Conrad, 1971; Stubbs & Tucker, 1974). The level of proficiency that the cloze test measures has been criticized. Alderson (1979, 1980) and Markham (1988) found that the cloze measured tested the lower-order level proficiency while some researchers indicated that the cloze test measured higher-order level proficiency (e.g., Bachman, 1982; Difabio, 1997; Jon, 1991; McKenna & Layton, 1990; Oller, 1972). Interestingly, Prapphal & Tanapongpipat (1984) and Markham (1988) found that the cloze test measured both lower-order and higher-order proficiencies.

Like other testing methods, the cloze test has both strengths and weaknesses. All the technically problematic aspects of the cloze test were summarized by Alderson and Klein-Braley (as cited in Klein-Braley, 1997) as follows:

1. The deletion rates used in classical cloze tests are too high
2. Using only one text in cloze testing is not a representative sample of the language. In addition, item bias is possible as a result of text content.



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

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

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

6. The high reliability and validity coefficients found in many of the research studies were partly a result of extremely heterogeneous subject groups involved in the investigations. For the highly homogeneous groups of the Duisburg placement procedures, low reliability and validity coefficients were more typical. (p. 59-60)

C-TEST

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, as 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, as cited in Klein-Braley, 1997: 63-64)



7. The test should have high reliability and validity (Klein-Braley & Raatz, 1984: 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 native speakers of English who took the English cloze test and the native speakers of German who took the German cloze test. The results of the two groups showed that the tests were surprisingly highly reliable and valid. However, 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,as 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 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,as 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 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 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. Ratz (1984, as 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 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 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. 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 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

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-Braley, 1997).

REFERENCES

- [1] Aborn, M., Rubenstein, H., & Sterling, T. D. (1959). Sources of contextual constraint upon words in sentences. *Journal of Experimental Psychology*, 57, 171-80.
- [2] Abraham, R. G., & Chappelle, C. A. (1992). The meaning of cloze test scores: An item difficulty perspective. *Modern Language Journal*, 76(4), 468-479.
- [3] Aitken, K. G. (1977). Using cloze procedure as an overall language proficiency test. *TESOL Quarterly*, 11(1), 59-67.
- [4] Alderson, J. C. (1978). *A study of the cloze procedure with native and non-native speakers of English*. Unpublished doctoral thesis, University of Edinburgh, Scotland.
- [5] Alderson, J. C. (1979). The cloze procedure and proficiency in English as a foreign language. *TESOL Quarterly*, 13(2), 219-227.



- [6] Alderson, J. C. (1980). Native and nonnative speaker performance on cloze test. *Language Learning*, 30(1), 59-76.
- [7] Alderson, J. C. (1983). The cloze procedure and proficiency in English as a foreign language. In J. W. Oller, Jr. (Ed.), *Issues in Language Testing Research* (pp. 205-217). Massachusetts: Newbury House.
- [8] Anderson, J. (1971). Selecting a suitable 'reader': Procedures for teachers to assess language difficulty. *RELC Journal*, 2, 35-42.
- [9] Bachman, L. F. (1982). The traits structure of cloze test scores. *TESOL Quarterly*, 16(1), 61-70.
- [10] Bachman, L. F. (1985). Performance on cloze tests with fixed-ratio and rational deletions. *TESOL Quarterly*, 19(3), 535-556.
- [11] Bachman, L. F., Palmer, A. S. (1996). *Language testing and practice: Designing and developing useful language tests*. Oxford, the United Kingdom: Oxford Press.
- [12] Boonsathorn, S. (1987). *C-Tests, proficiency, and reading strategies in ESL*. Unpublished doctoral dissertation, University of Alberta, Canada.
- [13] 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.
- [14] Boonsathorn, S. (1990). C-Test/MC-Test: A new instrument for measuring English proficiency. *PASAA*, 20(2), 46-50.
- [15] Brown, J. D. (1983). A closer look at cloze: Validity and reliability. In J. W. Oller, Jr. (Ed.), *Issues in Language Testing Research* (pp. 237-250). Massachusetts: Newbury House.
- [16] Brown, J. D. (1988a). Tailored cloze: Improved with classical item analysis techniques. *Language Testing*, 5(1), 19-31.
- [17] Brown, J. D. (1988b). *What makes a cloze item different?* University of Hawaii Working Papers in ESL, 7, 17-39.
- [18] Brown, J. D. (1993). What are characteristics of natural cloze tests? *Language Testing*, 10(2), 93-116.
- [19] Carlson, N. R. (1988). *Discovering Psychology*. Boston: Allyn and Bacon.
- [20] Caulfield, J., & Smith, W. C. (1981). The reduced redundancy test and the cloze procedure as measures of global language proficiency. *Modern Language Journal*, 65(1), 54-58.



ปีที่ 31 ฉบับที่ 2 (2555)

- [21] Chavanachat: (1986). The use of a cloze test as a measure of English proficiency for Thai university graduates. *PASAA*, 16(1), 26-35.
- [22] Cleary, C. (1988). The C-Test in English: Left-hand deletions. *RELC Journal*, 19(2), 26-38.
- [23] Cohen, A. D., Segal, M., & Bar-Simon-Tov, R. (1984). The C-Test in Hebrew. *Language Testing*, 1(2), 221-225.
- [24] Coniam, D. (1993). *Co-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.
- [25] Connelly, M. (1997). Using C-Test in English with post-graduate students. *English for Specific Purposes*. 16(2), 139-150.
- [26] Culhane, J. W. (1970). Cloze procedures and comprehension. *Reading Research*, 23(5), 410-413, 464.
- [27] Darnell, D. K. (1968). *The Development of an English Language Proficiency Test of Foreign Students Using a Clozentropy Procedure*. Boulder: Department of Speech and Drams, University of Colorado.
- [28] 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.
- [29] Difabio, H. (1997). Psychometric characteristics of the cloze test in the testing of reading comprehension [Abstract]. *Language Testing*, 31(2), 97.
- [30] Dornyei, Z., & Katona, L. (1992). Validation of the C-Test amongst Hungarian EFL learners. *Language Testing*, 9(2), 187-206.
- [31] Farhady, H., & Keramati, M. N. (1996). A test-driven method for the deletion procedure in cloze passages. *Language Testing*, 13(2), 191-207.
- [32] Fotos, S. S. (1991). The cloze test as an integrative measure of EFL proficiency: A substitute for essays on college entrance examinations? *Language Learning*, 41(3), 313-336.
- [33] 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.
- [34] 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 Pual.



- [35] Heaton, J. W. (1975). *Writing English Language Tests*. London: Longman.
- [36] Hinofotis, F. B. (1976). *An Investigation of the Concurrent Validity of Cloze Testing as a Measure of Overall Proficiency in English as Second Language*. Unpublished dissertation, Southern Illinois University.
- [37] Hinofotis, F. B. (1980). Cloze as an alternative method of ESL placement and proficiency testing. In J. W. Oller, Jr. (Ed.), *Research in Language Testing* (pp. 121-128). Massachusetts: Newbury House.
- [38] Hinofotis, F. B. (1987). Cloze testing: An overview. In M. H. Long & J. C. Richards (Eds.), *Methodology in TESOL: A Book of Readings* (pp. 412-416). Massachusetts: Newbury House.
- [39] Jafarpur, A. (1995). Is C-Testing superior to cloze? *Language Testing*, 10(1), 195-216.
- [40] Jongsmas, E. (1971). *The Cloze Procedure: A Survey of the Research* (Occasion Papers in Reading). Bloomington: Indiana University, school of Education.
- [41] Jon, J. (1991). What cloze tests measure: A factor analysis with replications. *Language Testing Update Abstract*, 9, 44.
- [42] Klein-Braley, C. (1981). *Empirical Investigations of Cloze Tests*. Unpublished dissertation, University of Edinburgh, Scotland.
- [43] Klein-Braley, C. (1982). On the suitability of cloze tests as measures of reading comprehension. In A. J. M. van Geest, O. J. Koster, & J. F. Matter (Eds.), *Toegepaste Taalwetenschap in Artikelen: Vol. 3.* (pp. 49-61). Amsterdam, the Netherlands: VU Bockhandel.
- [44] Klein-Braley, C. (1983). A cloze is a cloze is a question. In J. W. Oller, Jr. (Ed.), *Issues in Language Testing Research* (pp. 218-228). Massachusetts: Newbury House.
- [45] Klein-Braley, C. (1985). A cloze up on the C-Test: A study in the construct validation of authentic tests. *Language Testing*, 2(1), 76-104.
- [46] Klein-Braley, C. (1997). C-Test in the context of reduced redundancy testing: An appraisal. *Language Testing*, 14(1), 47-84.
- [47] Kober, J., & Sigott, G. (in press). Deletion patterns and C-Test difficulty in German. *Moderne Sprachen*.
- [48] Koberl, J., & Sigott, G. (1996). Deletion patterns in C-Test: Native speakers vs. foreign learners. *Language Testing Update*, 20, 45-48.



ปีที่ 31 ฉบับที่ 2 (2555)

- [49] Lennon: (1989). Conversational cloze tests for advanced learners [On-line].
ELT Journal, 43(1), 38-44. Abstract from: <http://www.ericir.syr.edu>:
Item: EJ 384095.
- [50] Low, L., & Toh, H. M. (1995). *Principles of economics*. Singapore: Addison
Wesley.
- [51] Markham: L. (1988). The cloze procedure and intersentential
comprehension in college-level German. *IRAL*, 26(1), 44-51.
- [52] McBeath, N. (1990). C-Tests – some words of caution. *English Teaching
Forum*, 28(2), 45-46.
- [53] McKenna, M. C., & Layton, K. (1990). Concurrent validity of cloze as a measure
of intersentential comprehension [On-line]. *Journal of Educational
Psychology*, 82(2), 372-377. Abstract from: <http://www.ericir.syr.edu>:
Item: EJ 442331.
- [54] *Microsoft Word for Windows 7.0* [Computer software]. (1995). California:
Microsoft Corporation.
- [55] Mitchell, R. (1991). *Validating language tests: A Hong Kong case study (Rep.
No.10)*. Hong Kong: City Polytechnic of Hong Kong, Department of
English.
- [56] Mondy, R. W., & Noe III, R. M. (1993). *Human resource management* (5th ed.).
Boston: Allyn and Bacon.
- [57] Oller, J. W., Jr. (1972). Scoring methods and difficulty levels for cloze tests of
proficiency in English as a second language. *Modern Language Journal*,
56(3), 151-158.
- [58] Oller, J. W., Jr. (1979). *Language Tests at School*. London: Longman.
- [59] Oller, J. W., Jr., & Conrad, C. A. (1971). The cloze technique and ESL
proficiency. *Language Learning*, 21(2), 183-195.
- [60] Perkins, K., & German: (1985). *The Effect of Information Gain on Different
Structural Category Deletions in a Cloze Test*. (Paper presented at
Midwest TESOL Annual Meeting). Milwaukee, Wisconsin.
- [61] *Perfect Copy Classic 1.18* [Computer software]. (1995). New York: Logicus
Incorporated.
- [62] Porter, D. (1978). Cloze procedure and equivalence. *Language Learning*,
28(2), 333-341.
- [63] 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.



- [64] 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.
- [66] Prapphal, K., Pas. K., & Tanapongpipat, S. (1984). Modified cloze tests as predictive measures of language acquirer and academic achievers. *PASSA*, 14(2), 75-89.
- [67] 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: University of Essex.
- [68] 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: Arbeitskreis Sprachenzentrum [AKS].
- [69] 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.
- [70] 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: Arbeitskreis Sprachenzentrum [AKS].
- [71] Reilly, R. R (1971). A note on ‘clozentropy’: A procedure for testing the English language proficiency of foreign students. *Speech Monographs*, 38,350-353.
- [72] Segal, M. (1983). *The C-Test*. (Seminar Paper in Hebrew). Jerusalem, Israel: Hebrew University, School of Education.
- [73] Sigott, G., & Koberl, J. (1993). Validating the X-Test. *Language Testing Update*, 14,53-58.
- [74] 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
- [75] Stillman II, R. J. (1996). *Public Administration*. (6th ed.). Boston: Houghton Mifflin.
- [76] Stubbs, J. B., & Tucker, G. R. (1974). The cloze test as a measure of English proficiency. *Modern Language Journal*, 58(5-6), 239-241.



ปีที่ 31 ฉบับที่ 2 (2555)

- [77] Tannembaun: H., Williams, F., & Clark, R. A. (1969). Effects of grammatical information on word predictability. *Journal of Communication*, 19, 41-48.
- [78] Taylor, W., L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 30, 415-453.
- [79] Taylor, W., L. (1957). Cloze' readability scores as indices of individual differences in comprehension and aptitude. *Journal of Applied Psychology*, 41, 19-26.
- [80] Weir, C. J. (1990). *Language Testing*. New York: Prentice Hall.
- [81] Weir, C. (1993). *Understanding and Developing Language Tests*. London: Prentice-Hall.
- [82] Weiss, R. (1983). *The C-Test*. (Seminar Paper in Hebrew). Jerusalem, Israel: Hebrew University, School of Education.