Cloze test and c test

A - Rostamzadeh [*]

Abstract

The present study was an attempt to investigate varieties of cloze test and C-test as measures of reading comprehension ability. It aimed at comparing the performances of subjects on different cloze and C-test Forms constructed from the same passage. To this end, five versions of cloze test and two versions of C-test were developed by applying different deletion techniques and deletion rates, and were labeled Form A. (Standard Cloze), Form B. (First – Half C-test), Form C. (Rational cloze), Form D. (Noun – phrase cloze), Form E. (Sentence cloze), Form F. (Independent – Dependent clause cloze), and Form G. (standard C-test). The test Forms were randomly administered to 247 EFL students along with Michigan ECPE (Examination for the Certificate of Proficiency in English) test. Nearly 35 students took each test Form. To find out which test from(s) showed a better picture of testees' reading comprehension ability, the correlation coefficients between each test Form and ECPE reading section and variance overlap were calculated. It revealed that Form C. had the highest correlation and variance overlap with ECPE reading section (.62) and enjoyed, with the exception of Form F. the highest reliability (.96). To compare the performances of subjects on test Forms and to determine their reliability and validity, two one-way ANOVAs were run and KR-21 and Pearson product moment formulas were applied. The observed F-ratio came out to be 73.15, which was

[*] Islamic azad university.

considerably higher than the critical F-ratio at D.F. (6.24), i.e., 2.14. This meant that the subjects performed differently on these Test Forms. The reliability and validity of Test Forms showed that their psychometric characteristics were also different.

Key words: redundancy in language, test of reduced redundancy, standard cloze, noun – phrase cloze, sentence cloze, independent – dependant clause cloze, standard C-test, first – half C-test, reading comprehension.

Introduction

Testing is a very complicated and intricate responsibility. The delicacy of decision-making and the intricacy of different related fields have made the language testing a challenging field (Farhady et al., 1994:3). Myriads of studies have been conducted on testing the reading comprehension, which has attracted the attention of so many researchers. Disagreement on the nature of reading has resulted in different tests with different forms, which evaluate different kinds of abilities such as pre-reading, reading speed, and reading comprehension.

Using the principle of reduced redundancy is a development within the so-called "integrative – sociolinguistic"(formerly called psychometric- sociolinguistic) era. The idea behind reduced redundancy is that native speakers of a language can restore the damaged messages by using their knowledge of the language. Following this, non-native speakers cannot reconstruct the impaired messages of the language as efficiently as native speakers can do, because they lack the knowledge that native speakers possess. In particular, "cloze test" procedures are the result of the

idea that linguistic performance can be tested by setting the task of "reconstructing" damaged texts whose redundancy has been reduced by omitting some parts of the texts. Cloze test was first introduced in 1953 by Taylor as a measure of readability, but later some researchers investigated its use as a measure of ESL/EFL general language proficiency (Hinofotis, 1987) and reading ability (Anderson, 1971; Trollope, 1995); and it came out to be a reliable and valid measure of these constructs.

Cloze test – an operationalization of the theory of reduced redundancy – was first constructed by deleting every nth word from a text, but later rational deletion was proposed by researchers such as Porter (1978), Klein-Braley (1983) and Bachman (1982;1985). C-test was first introduced into the realm of language testing by Klein-Braley and Raatz (1984) as a modification of cloze procedure. The original C-test was developed by mutilating the second half of every other word, beginning from the second word in the second sentence of a passage.

The present study was conducted to compare the performance of Iranian university students on cloze passages constructed from the same text using different procedures (fixed - ratio, text - driven, rational, and C-test) to find out which procedure can better tap testees' reading ability. Cloze test, as a measure of language proficiency and reading comprehension, has been one of the controversial issues in language testing. It has been the subject of numerous studies during the last two decades most of which have supported that the procedure is an integrative, reliable and valid measure.

Instructors have been constructing cloze tests and administering them for years to measure their students' reading ability. In this way, the need for investigating different cloze methods and finding the most appropriate one for measuring Iranian university students' reading ability was felt. In order to discover the best possible method for constructing cloze tests to assess Iranian university students reading ability, the following research questions were formulated:

1- Is there any difference between cloze procedures and C-tests in measuring Iranian university students' reading ability?

2- Is there any relationship between the deletion based on a specific rationale and the psychometric characteristics of cloze tests?

3- Are there any differences or similarities in the performances of Iranian university students on different methods of cloze tests?

Based on these questions, three null hypotheses were formulated.

Background

Different methods of testing reading comprehension are: Multiple-choice Questions (MCQs), Cloze, summary cloze, selective deletion gap filling, C-test, information transfer, short-context technique, and short answer questions.

Multiple-choice questions (MCQs) (Psychometric- Structuralist era) provide results, which are easily quantifiable and are easily administered.In Cloze method (integrative-sociolinguistic era) some words are deleted from the passage after some introductory non-deleted sentences. This method is the focus of the present

study. *In summary cloze* a reading passage is summarized by the tester and then, some blanks are left in the summary to be completed by the testees. In *selective deletion gap filling* linguistic reasoning is used to decide on deletions and so it is easier to state what each test is intended to measure(Oller,1979;Weir,1990). In *C-test*, every second word in a text is partially deleted. In *Information transfer* the information transmitted verbally is transferred to a non-verbal form. They are realistic tasks in various situations and their interest and authenticity give them high face validity in their contexts. In *short-context technique* reading ability is measured through using very brief stimulus materials followed by one or two items testing comprehension of overall meaning. In *short answer questions*, there are questions which require the candidates to write down specific answers in spaces provided on the question paper.

Cloze procedures

Cloze procedure is a technique for measuring reading comprehension ability. The word cloze comes from the idea of closure in Gestalt theory. According to Taylor, cloze is "any passage of suitable length and level of difficulty with every nth word deleted".

To determine the length of the passages for cloze tests, first, the value of "n"should be determined. Cloze tests with every 5th or 7th word deletion have become popular and cloze tests with every 7th word deleted are known as standard cloze tests.

Different varieties of cloze exist regarding the deletion procedure: Fixed-ratio Deletion, Random Deletion, Stratified Sampling, and Rational Deletion. In *fixed-ratio Deletion technique*,

cloze tests are made by fixing "deletion rate" at every nth word. This technique is the most frequently used one, also called 'pseudo-random', because it doesn't enjoy a complete randomization process. In *random Deletion procedure*, all the words in the passage will be assigned a number (for example, 1 to 250 in a passage of 250 words). The numbers will be randomly selected and corresponding words will be deleted. The deleted words shouldn't be adjacent. In *stratified Sampling technique*, the number of words from different grammatical classes is determined. Then, the number of deletions is proportionately set for each class of word. Every word class has an equal chance to be included in the deletion process. *Rational deletion* is a substitution for fixed word deletion. In this deletion technique, words are deleted according to linguistic criteria. The deletion is based on linguistic and discourse structures of language and not just on a random number such as 5, 6, 7, etc. (Farhady and Keramati, 1996).

Bachman (1985) introduces three criteria for the classification and selection of the delectable words:

1) Syntactic, which depends only on clause-level structure.

2) Cohesive, which depends on inter-clausal and inter-sentential cohesive   structures (also cited in Halliday and Hasan, 1976).

3) Strategic, which depends on "large scale" patterns of coherence.

Bachman (1985) claims that the rational deletion cloze is a better measure for the reading of connected discourse. He contends that rational deletion approach does enable the test constructor to determine what exactly is tested. However, Klein-Braley (1997) asserts that the only problem with this approach is that it is not in

agreement with the theory of reduced redundancy in testing. If the tester chooses what to test, then the random sampling model is abandoned.

The most common types of scoring cloze tests are: exact word method, acceptable word method, weighted response method and multiple choice cloze method. In exact word method (EWM), only the exact deleted word is the correct answer even if others can work. Although, EWM is easy and objective, it is a difficult task, i.e., it limits the range of possible responses. This method hinders the testee from activating the creative process of language use in producing suitable words. Acceptable word method (AWM) allows any response, which is acceptable in the context to a native speaker. It gives credit to synonyms. Therefore, the scoring procedure becomes more difficult than that of the exact word method and moves toward subjectivity. Weighted response method (WRM) considers some responses more appropriate than others, i.e., different words receive a varying degree of credit. This method seems to be impractical (Farhady et al., 1994;Oller,1979). Multiple-choice cloze technique has been introduced by Jonz (1976) for increasing objectivity and ease of implementation. In multiple-choice cloze, some other irrelevant words are added to the text and the test takers most choose the correct one from among them (Madsen, 1983).

C-Test

C-test is an alternative to cloze test, in which the second half of every other word is deleted, leaving the first and the last sentence of the passage intact. A C-test consists of a number of short

passages (maximum 100 words) on a variety of topics. In C-test, a clue (half the word) serves as a stimulus for the test takers to find the other half (Klein-Braley,1985; Jafarpur,1995). Klein-Braley (1997) describes the advantages of C-Test over cloze tests as follows:

1) Much more items are possible with much shorter texts.

2) C-Test scoring is exact and objective.

3) Since each second word is damaged the probability of obtaining a representative sample of all the "word classes" in the text is very high.

However, C-Test is irritating for students and has low face validity (Weir, 1990; Jafarpour, 1995).

Method

Subjects

The subjects were 247 EFL students majoring in English. They were selected randomly from Azad University's Teacher Training College, Ministry of Education's Teacher Training College, Ministry of Education's in-service training college, Tehran University's Faculty of foreign languages and Qom Azad University. The subjects were about 21 to 31 years of age. Both males and females participated in classes; so, the classes were heterogeneous.

Instrumentation

The following research instruments were used to measure the variables in this project:

1- A Michigan's ECPE test, which was intended to act as a criterion to correlate against the other test types, i.e., cloze tests and the C-tests.

2- Five versions of cloze tests, which were provided out of the same passage.

3- Two versions of C-test, which were developed out of the same passage used for making cloze tests.

The Design of the Study

The present study was a descriptive one including a number of test administrations. Farhady (1995) has classified descriptive methods into three major groups of *survey, interrelational,* and *developmental* methods. The interrelational method has four components one of which is *correlational* method. This very method is the basis of this study. As Seliger and Shohamy (1989) put it, data collection in descriptive research is achieved through tests, surveys and questioners, self-reports and interviews, and observations. This study used only the first data collection instrument, i.e., tests.

Procedures

All the participating subjects were first administered the Michigan test. Through the following steps, five versions of cloze test based on the same text were developed:

1- The text books of BA students majoring in English were examined carefully to ensure the readability of the passage.

2- Using the Fog's index formula, the readability of every ten pages in the same text books was calculated.

3- The average readability of each book was determined, and finally the average readability of all books was calculated , i.e., 26.9. Based on this readability level, a passage with readability of 26.4 was selected.

4- Five fill-in cloze tests called Forms A, C, E and F were made.

In form A, which was a standard cloze, the deletion rate was fixed at 7. In forms D, E and F, the deletion rate was based on the number of noun phrases, independent plus dependent clauses, and sentences, respectively. In these three Forms (D, E and F), first the number of words in the first sentence was subtracted from the total number of words in the passage (270 words). Then, the result was divided by the number of noun phrases, independent and dependent clauses, and sentences in the passage. The resulting number came to be the rate of deletion in Forms D, E, and F. Form C was constructed by deleting specific words. In other words, the deletion rate was not fixed. In form C, which was a rational cloze, cohesive devices and discourse markers that play a key role in establishing relationship between text propositions were deleted. Thirteen items in this test Form fell into four categories: *discourse markers* (1, 7, 8, 9, 10 and 12), *anaphoric pronouns* (2), *lexical substitutes* (3, 4, 5, 11, and 13), and *lexical items* (6). Each fill-in version was administered to twenty university students majoring in English. Using the answers provided by these testees, a multiple – choice version of these 5 test forms was made. The C-tests were constructed out of the same passage used for developing the cloze tests and labeled

Form B and Form G. these two tests were constructed in the following way:

Form B was FH4 (First half every fourth word) C-test in which as its name indicates the first half of every fourth word was omitted. This version was developed by Jamali (1998) and was shown to have the highest psychometric characteristics among the other versions of C-test. Form G was a standard C-test in which the second half of every second word was dropped. The time allocation varied form 15 to 30 minutes depending on the test Form. Seven groups (each one including 35 students) for seven test forms participated in test administration process. The test Forms as well as the criterion measure were administered concurrently to the subjects. The Test forms were distributed among the subjects randomly. Table 1 presents the number of items and deletion rates in test Forms A- G.

Table 1 : Number of Items and Deletion Rates

Form	Description	No. of Items	Deletion Rate
A	Standard cloze	34	7
B	First-Half C-test	60	4
C	Rational cloze	13	Not Fixed
D	Noun-phrase cloze	60	4
E	Sentence cloze	12	20
F	Independent-dependent clause cloze	22	11
G	Standard C-test	121	2

Results and Discussion

The Michigan test consisted of 100 multiple-choice items completed in 60 minutes. Due to practical limitations, only the grammar, vocabulary and, reading comprehension sections were administered.

The descriptive statistics of the tests used in this study were computed(table 2).

Table 2: Descriptive statistics of the test Forms and Michigan ECPE

		A	B		C	D	E	F	G
		35	37		34	35	37	35	34
Grammar	$\overline{X}$	21.54	21.16		21.94	23.23	23.00	21.82	23.29
	S	9.54	9.08		12.23	7.54	10.87	9.76	6.40
Vocabulary	$\overline{X}$	16.22	15.29		17.50	17.11	19.13	15.60	13.14
	S	10.25	7.22		13.50	7.43	10.78	9.40	4.48
Reading	$\overline{X}$	4.88	5.27		3.91	5.57	5.97	5.97	3.82
	S	2.98	2.43		3.38	3.31	4.31	2.95	2.20
Test Form A-G	$\overline{X}$	20	24.78		7.64	30.14	7.86	15.45	44.35
	S	5.97	9.44		3.82	9.98	2.77	4.61	11.34
No. of Items		34	60		13	60	12	22	121
Maximum		34	40		13	58	12	21	70
Minimum		7	5		1	15	2	7	30
Adjusted Means		11.76	8.26		11.75	10.04	13.1	14.04	7.33

Highest SD   = 11.34
Lowest SD   = 2.77
Highest Mean = 14.04
Lowest Mean = 7.33

Since the main objective of this study was to find out which cloze or C-test technique can better tap the reading ability of the university students, the correlation coefficients between scores on each test Form and the scores on Michigan ECPE reading section were computed (table 3).

Table 3: Correlation Coefficients between Each Test Form and Michigan ECPE Reading section

	A	B	C	D	E	F	G
ECPE Reading section	.66	.62	.79	.76	.64	.59	.58

The best way to interpret the correlation coefficient is to convert it into variance overlap between the two measures (Hatch & Farhady, 1982).The overlap showed that the two measures provided similar information. Therefore, to find out the common variance between scores on each test Forms and ECPE reading section, the correlation coefficients were squared (Table 4).

Table 4: Variance Overlap between Each Test Form and ECPE Reading Section

	A	B	C	D	E	F	G
ECPE Reading section	.43	.38	.62	.57	.40	.34	.33

The best indicator of testees' reading ability

So, the first null hypothesis - there is no difference between cloze procedures and C-tests in measuring university students' reading ability - was rejected.

To determine the psychometric characteristics of the tests, the reliability coefficient of the tests was determined. Then, the Pearson product moment formula was employed to compute the concurrent validity of the test Forms with the subtests of Michigan ECPE(Table 5).

		A	B	C	D	E	F	G
Grammar	Rel.	.91	.90	.95	.84	.94	.91	.50
	Val.	.67	.76	.63	.68	.73	.60	.53
Vocabulary	Rel.	.93	.83	.97	.84	.93	.91	.89
	Val.	.48	.52	.43	.45	.52	.36	.33
Reading	Rel.	.90	.80	.92	.87	.86	.90	.84
	Val.	.66	.62	.79	.76	.64	.59	.58
Test Form	Rel.	.79	.85	.84	.86	.70	.82	.78

Table 5: Reliability and validity of the study measures

Therefore, the second null hypothesis – there is no significant relationship between rationale behind deletions and the psychometric characteristics of cloze tests – was rejected. Moreover, the findings of the study indicated that Forms B, C and D showed better psychometric qualities than the other Forms. These findings support the findings by Farhady and Kerakati (1996), and Jamali (1998). That is, Form B, which was a First-Half C-test, has yielded better reliability and validity than Form G, which was a standard C-test. Also, Form D in which deletion rate was based on the number of noun phrases had higher reliability and validity compared with Forms E and F in which the deletions were based on the number of sentences and the total number of independent and dependent clauses, respectively.

To determine whether there were significant differences among the performances of groups on test Forms, two one-way ANOVAs were run on.

The first ANOVA was carried out to see whether the differences between the standard deviations and means of the scores on the subtests of the criterion measure are significant or not (table 6).

Table6: ANOVA on Michigan ECPE

	sum of squares	df	mean square	F	sig
Between Groups Within Groups Total	1230.722 93908.493 95139.215	6 240 246	205.120 391.285	.254	.790

F critical > F observed

2.12 > 0.53

P < 0.05

Therefore, the difference between the groups regarding their language proficiency level was not significant. The second one-way ANOVA was carried out to compare the performances of students on test Forms (Table 7).

Table 7: ANOVA on Test Forms A-G

	sum of squares	df	mean square	F	sig
Between Groups Within Groups Total	35524.65 19425.09 54949.74	6 240 246	5920.77 80.93	73.15	.000

$P < 0.05$

The results indicated that the performances of subjects on test Forms were significantly different (f = 73.15). To see where the difference occurred a Scheffe test was conducted (Table 8).

Table8: The Results of the Scheffe' Test
Group

Mean	Group	3	5	6	1	2	4	7
7.64	3							

7.86	5						
15.45	6	7.81*					
20.00	1	12.35*	12.13*				
24.78	2	17.13*	16.91*	9.32*			
30.14	4	22.49*	22.27*	14.68*	10.14*		
44.35	7	36.70*	36.48*	28.89*	24.35*	19.56*	4.21*

Comparing the means of the groups through a scheffe test revealed that cloze test Forms were different from one another. These results supported the findings of Alderson (1983), Farhady, etal (1994), Farhady and Keramati (1996), and Jamali (1998) that different cloze and C-test are not parallel. As the results showed, the greatest difference is between group 6 and 3. These results indicate that the groups performed differently on test Forms.

Conclusion

This study aimed at finding a justified method for constructing cloze tests and C-tests to measure reading comprehension ability by comparing performances of students on seven different test Forms. This study also sought to determine the reliability and validity of the test Forms constructed from the same text and that different cloze procedures had not the same reliability and validity coefficients.

Cloze test and C-test are not parallel or equivalent, and that text difficulty and type of deletion are important factors. Deleting words must be done in a careful procedure and also random and fixed – ratio deletion of words with no regards to what specific ability it is going to test is an oversimplification.

It does not seem advisable to pick up a random number as the deletion rate and to apply it to any text to develop cloze tests for testing students reading comprehension. It is recommended that the test constructors delete specific words such as discourse markers and cohesive ties, which connect propositions of a text together or at least base the deletion rate on the number of linguistic and discourse structures, below the clause preferably, of the text. Therefore, the test constructors in university level should base the deletion of words in constructing cloze tests on a sounder basis.

References

Alderson, J. C. (1983). The Cloze Procedure and Proficiency in English as a Foreign Language, In Oller, J. W. Jr, (Editor). Issues in Language Testing Reaserch. Rowley, MA: Newbury House: 205-17.

Anderson, J (1971). A technique for measuring reading comprehension and readability. English language Teaching Journal, 25(2): 178-82.

Bachman, L. F. (1985). Performance on cloze tests with fixed-ratio and rational deletions. TESOL Quarterly, 19(3): 535-66.

Bachman, L. F. (1982). The Trait Structure of Cloze Test Scores. TESOL Quarterly, 16(1): 61-8.

Farhady, H. (1995). Research Methods in Applied Liguistics. Tehran: Payam-e-Nour Publications.

Farhady, H., Jafarpur, a. & Birjandi, P. (1994). Testing Language Skills. Tehran: SAMT Publication.

Farhady, H., & Keramati, M. N. (1996). A Text-driven method for the deletion procedure in cloze passage. Language Testing. 13(1): 191-207.

Halliday, M. A. K. and Hasan R. (1976). Cohesion in English. London: Longman.

Hatch, E. and Farhady, H. (1982). Research Design and Statistics for Applied Linguistics. California: University of California Publication.

Jafarpur, A. (1995). IS C-testing superior to cloze? Language Testing, 12(2): 194-216.

Jamali, F. (1998). Varietiest of C-test as Measures of General Language Proficiency. Unpublished master's thesis, Tehran: Tarbiat Modarres University.

Jonz, J. (1976). Improving the basic egg: the multiple-choice cloze. Language Learning, 26: 255-65.

Klein-Braley, C. (1997). C-Test in the context of reduced redundancy testing:An appraisal.Language Testing,14(1):47-84.

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.

Klein-Braley, C. (1983). A Cloze Is a Question. In Oller, J. W. Jr, Editor. Issues in Language Testing Research. Rowly, MA: Newbury House.

Klein-Braley, C. & Rattz, U. (1984). A survey of research on the c-test. Language Testing, 1(2):134-46.

Madsen, H. S. (1983). Techniques in Testing. Oxford: Oxford University Press.

Oller, J. W. (1979). Languge Tests at School: A Pragmatic Approach. London: Longmamn.

Seliger, H. W. , and Shohany L. (1989). Second Language Research Methods. Oxford: Oxford University Press.

Trollope, J. (1995). On-and Off-Line Reading Strategies: The cloze Technique (Electronic Version). TESOL Journal.

Weir, C. J. (1990). Communicative Language Testing. NJ: Prentice hall Regents.