Grotjahn, Rüdiger. (Hrsg.). (1996). Der C-Test. Theoretische Grundlagen und praktische Anwendungen (Bd. 3, S. 183–195). Bochum: Brockmeyer

Tetsuro Chihara, William D. Cline and Toshiko Sakurai

If the cloze test is a question, is the C-test the answer?

While the cloze test has been criticized for varying in its validity depending upon the deletion rate (Alderson, 1979), the C-test has been proposed a viable alternative. To find whether or not the C-test is superior to the cloze test, four C-tests were prepared for the present study. Two tests followed the basic procedure of deleting the second half of every second word beginning from the second sentence. In two experimental tests, deletions started from the first word of the second sentence. Four hundred forty Japanese junior college students took part in the experiment. One hundred seventy-five of the students had also taken the TOEFL so their TOEFL scores were correlated with the C-test results. Significant differences were found in the scores of two tests. All four tests correlated satisfactorily with TOEFL total scores and all four tests had satisfactory reliability.

1. Introduction

The cloze test has achieved considerable popularity and credit as a reliable and valid instrument for measuring foreign language proficiency. It has been in use over thirty years and numerous studies have supported its validity and reliability (Oller, 1979). In recent years, however, the cloze test has been reported to have some serious defects (Alderson, 1979; Klein-Braley, 1983, 1985; Klein-Braley & Raatz, 1984; Raatz & Klein-Braley, 1982). One of the major problems seems to be in the construct validity of the cloze tests. Alderson (1979) argues that the cloze procedure is not automatically valid and reliable for producing tests. Since the deletion rate chosen affects the test performance, each cloze test must be validated each time it is administered. Raatz & Klein-Braley (1982) share his doubts about generalization of the cloze procedure. They claim that every *n*th word deletion procedure does not ensure the random sampling theory which is crucial for tests of reduced redundancy. Cloze tests are tests of reduced redundancy which require test takers to utilize what Oller calls their pragmatic expectancy grammar. A cloze passage is a sample of language and the student's performance on any one passage is generalized to represent his performance on any other sample of language. In order for the results of a cloze test to

be generalized, the sample of language must be random. However, changes in texts, deletion rate, and starting points also change the ratio of content and structural words deleted in the cloze text, and that affects not only the difficulty of the test but its reliability and validity. Some cloze tests turn out to be valid and reliable while others have low reliability and validity coefficients.

The C-test, developed by Raatz & Klein-Braley in 1981, is a modification of the cloze procedure, still retaining the underlying theory of general language proficiency. In the C-test, starting from the second sentence, the second half of every second word in the text is deleted (cf. Grotjahn, 1987). Dörnyei & Katona (1992), and Raatz & Klein-Braley (1982), have reported results with the C-test which support the superiority of the C-test to the traditional cloze procedure.

In order to reveal whether every second word deletion in C-tests gives a representative sample of the words in the whole texts, Raatz & Klein-Braley (1982) chose 40 English and 40 German texts, each with 200 words, and compared the proportion of content and structural words deleted for the C-tests to that of those in the original texts. The results show that the procedure assures the random sampling theory.

Concurrent validity of C-tests was also examined. Klein-Braley & Raatz (1982, 1984) report correlations of the C-tests in several languages: the English version, .36 – .72 and .62 – .90 with the Delta Test; the German version, .82 – .85 with the AWO Test; the French version, .87 with the Bochum French Placement Test. In Negishi's study (1987), the correlation coefficients of a C-test with the ELBA Test were .76 for the total test and .80 for the reading comprehension section. Dörnyei & Katona (1992) report positive correlations of the C-test with a Department Proficiency Test, (total) .43, with the TOEIC (total) .62, and .43 with an Oral Interview.

In the present study we asked the following questions:

1. Will the C-test have the same results with either first or second word deletion starting points in terms of the ratio of content and structure words and test difficulty? Our hypothesis was that there should be no significant difference in C-test results using either the Basic second or an Experimental first word deletion starting point (cf. Grotjahn, 1987, pp. 227f., however, for a possible significant effect of a first word deletion starting point).

2. Does the C-test produce high validity coefficients with a reliable and valid criterion in the Basic and Experimental versions?
3. Will two different C-tests developed with different passages both produce valid results as claimed? Klein-Braley & Raatz (1984, p. 145) write "It seems probable that, provided the C-Test is suitable in its level of difficulty for the target group envisaged, it will, in the majority of cases, produce valid results."

2. Method

2.1 Pilot test

A standard C-test suggested by Raatz & Klein-Braley (1982) consists of four independent paragraphs with about 80 words and 25 deletions in each paragraph giving a total of 100 deletions. The purpose of the pilot test was to prepare two C-test forms for the present experiment. Twelve passages of about 80 words with varying degrees of readability were collected and divided into three forms. The four passages in each form were arranged from easiest to most difficult so that each form became fairly comparable to one another. Deletion of the second half of words began from the second word of the second sentence in each passage. The subjects were 180 Japanese junior college students. They were randomly divided into three groups and each group took one of the three forms of the C-test. We calculated the mean and standard deviation of each passage and selected eight out of the twelve original ones, so that we had two forms of almost identical difficulty, each form consisting of four passages. These were A Basic and B Basic forms. (However, in the main experiment, the difficulty of the two tests turned out to be different.)

2.2 Main experiment

2.2.1 Materials

Two additional forms of the C-test were prepared, A Experimental and B Experimental, using the same tests as A Basic and B Basic. In Experimental forms, deletions started from the first word of the second sentence. Basic and Experimental C-tests are shown in the Appendix.

2.2.2 Subjects

We used a completely different population from the one that took the pilot test. The subjects tested were 440 Japanese junior college students, all female, majoring in English. Of 440 students tested, 175 had taken the TOEFL Test.

2.2.3 Procedure

The students were randomly divided into two main groups. In order to minimize the order effect, one half of the first group took a set of tests consisting of A Basic and B Experimental, and the other half took B Experimental and A Basic. The order was reversed for the second group with one half taking B Basic and A Experimental, while the other half took A Experimental and B Basic. A written instruction was given on how to perform the C-test with an example. The students were allowed to spend 60 minutes writing answers.

The exact-word scoring method was employed. First, reliability coefficients were calculated on each form and condition, followed by analysis of variance with Conditions (Basic or Experimental) and Forms (A or B) as independent variables. Meanwhile, the ratio of content words and structural words in each C-test form was calculated and compared with the ratio of content and structural words in the whole texts. We used Fries' (1952) definition of content and structural words.

The C-test's validity was studied through correlation with the TOEFL scores. Pearson Product-Moment Correlation Coefficients were calculated between the C-test scores and TOEFL total scores as well as the listening, structure, and vocabulary and reading subtests.

3. Results and discussion

The four C-tests forms were scored by the exact-word scoring method. Table 1 shows the mean scores, standard deviations, and KR-21 formula reliability coefficients for the tests. All four tests were found to be satisfactorily reliable. Overall, subjects did better on the second word deletion starting point than the first regardless of form (66.850% versus 64.996%), and did better on form A than form B regardless of condition (68.539% versus 63.307%). However, the interaction between condition and form should be taken into account; there is no effect of condition for form A but probably an effect for form B.

Table 1
Mean scores, standard deviations and reliability coefficients on C-tests by form and condition

Form	Condition	Mean Score (%)	*SD*	Reliability (KR-21)	*N*
A	Basic	68.582	9.29	.758	220
A	Experimental	68.496	10.439	.810	220
B	Basic	65.118	10.515	.803	220
B	Experimental	61.496	9.559	.748	220
Overall Means					
A		68.539			440
B		63.307			440
Basic		66.850			440
Experimental		64.996			440

A chi-square test of the proportion of content words and structural words in the four C-tests compared with the original texts revealed that there was no significant difference in the ratios of such words with the original texts (see Tables 2 and 3). Therefore, the four C-tests used in the present study were found to provide a random sampling of language for testing.

Table 2
Ratio of content words and structural words in C-tests by form and condition

	Form A			Form B		
	Basic (deletions only)	Experimental (deletions only)	Whole Passage	Basic (deletions only)	Experimental (deletions only)	Whole Passage
Content Words	49	43	152	49	53	165
(Ratio)	(49%)	(43%)	(48.7%)	(49%)	(53%)	(49.8%)
Structural Words	51	57	160	51	47	166
(Ratio)	(51%)	(57%)	(51.3%)	(51%)	(47%)	(50.2%)
Total	100	100	312	100	100	331

Table 3
Chi-square analyses of the proportion of content words and structural words in the four C-tests compared with the original texts

Form and Condition	*df*	Chi-square	*p*
Basic Form A	1	.002	.964
Experimental Form A	1	.993	.319
Basic Form B	1	.022	.882
Experimental Form B	1	.305	.581

A two factor analysis of variance was run over the C-test scores with condition (Basic or Experimental) and with form (A or B). The results from this analysis are given in Table 4. The F-ratio for condition was 7.618 and that for form 60.676. The effects of condition and form exceeded the critical value at $p < .01$, so the hypotheses of no difference between the deletion starting points and no difference for form can be rejected. The interaction of condition and form was also significant ($p < .01$). This means that significant differences in the mean scores by different deletion starting points may be due to the second factor, form.

Table 4
Two factor analysis of variance for form and condition

Source	*df*	*SS*	*MS*	*F*-ratio	*p*
Condition	1	756.655	756.655	7.618	.006
Form	1	6021.823	6021.823	60.626	.000
Condition × Text	1	687.823	687.823	6.925	.009
Explained	3	7466.300	2488.767	25.056	.000
Residual	876	87010.445	99.327		
Total	879	94476.745	107.482		

The C-test scores of the four forms were correlated with the TOEFL total scores as well as TOEFL subtests of 175 subjects (Section 1: listening; Section 2: structure; and section 3: vocabulary and reading comprehension). Table 5 shows the matrix of Pearson product-moment correlation coefficients between those tests. Correlations between TOEFL total scores and C-tests were higher than those between TOEFL subtests and the C-tests, which suggests that the C-tests measure students' overall ability in English rather than any one sub-skill. While Negishi (1987) reported the highest correlation between the C-test and the ELBA reading comprehension subtest (.80) and Chapelle & Abraham (1990) reported that the C-test correlated most strongly with the EPT vocabulary test (.836), rather high correlation with the TOEFL Listening section (.641) and Structure section (.615) were observed in the present study. With the results obtained in the present study, it can be said that the C-tests seem to measure something similar to what the TOEFL measures in students' English proficiency.

Table 5
Matrix of product-moment correlations

(a) Between the C-tests and TOEFL

	A Basic ($N = 82$)	A Experimental ($N = 93$)	B Basic ($N = 93$)	B Experimental ($N = 82$)
TOEFL Sec 1	.417	.611	.569	.357
TOEFL Sec 2	.474	.582	.615	.430
TOEFL Sec 3	.381	.469	.498	.482
TOEFL Total	.565	.646	.652	.552

(b) Between TOEFL and subtests

	TOEFL Sec 1 ($N = 175$)	TOEFL Sec 2 ($N = 175$)	TOEFL Sec 3 ($N = 175$)	TOEFL Total ($N = 175$)
TOEFL Sec 1	–			
TOEFL Sec 2	.630	–		
TOEFL Sec 3	.562	.637	–	
TOEFL Total	.857	.873	.850	–

4. Conclusion

Our present study revealed that the four C-tests were representative samples of language as far as the ratio of content and structural words in the text were concerned. It also showed that these C-tests were validated with TOEFL. In addition, reliability for all four tests was quite satisfactory. The highest correlations were with the four C-tests and the TOEFL Total scores. This seems to indicate that the C-test is a measure of general language pro-

ficiency. In these respects, the four C-tests used in this study did not show the variance reported with cloze tests (Alderson, 1979).

Two C-tests in this study had significantly different scores with differing deletion starting points. This may have been due to the differing deletion starting points or it may have been due to differing forms. By starting deletions in the C-test from the first word of the second sentence, we are not proposing a new type of test. Nevertheless, significantly differing scores obtained from different C-tests indicate that we should continue to heed the advice given more than ten years ago that, "... every C-test developed at this point of time should be very carefully examined" (Raatz & Klein-Braley, 1982, p. 134).

Acknowledgments

The authors gratefully acknowledge the assistance of Tomoko Yashima, Yasuyo Edasawa, and Yoko Hookabe.

References

Alderson, Charles J. (1979). The cloze procedure and proficiency in English as a foreign language. *TESOL Quarterly, 13*, 219-226.

Chapelle, Carol A. & Abraham, Roberta. (1990). Cloze method: what difference does it make? *Language Testing, 7*(2), 121-146.

Dörnyei, Zoltán & Katona, Lucy. (1992). Validation of the C-test amongst Hungarian EFL learners. *Language Testing, 9*(2), 187-206.

Fries, Charles C. (1952). *The structure of English*. New York: Harcourt, Brace & Co.

Grotjahn, Rüdiger. (1987). How to construct and evaluate a C-Test: A discussion of some problems and some statistical analyses. In Rüdiger Grotjahn, Christine Klein-Braley & Douglas K. Stevenson (Eds.), *Taking their measure: The validity and validation of language tests* (pp. 219-253). Bochum: Brockmeyer.

Klein-Braley, Christine. (1983). A cloze is a cloze is a question. In John W. Oller, Jr. (Ed.), *Issues in language testing research* (pp. 218-228). Rowley, Mass.: Newbury House.

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

Klein-Braley, Christine & Raatz, Ulrich. (1984). A survey of research on the C-test. *Language Testing, 1*, 134-146.

Oller, John W. Jr. (1979). *Language tests at school: a pragmatic approach*. London: Longman.

Negishi, Masashi. (1987). The C-Test: an integrative measure? *IRLT Bulletin, 1*, 3-26.

Raatz, Ulrich & Klein-Braley, Christine. (1982). The C-test – a modification of the cloze procedure. In Terry Culhane, Christine Klein-Braley & Douglas K. Stevenson (Eds.), *Practice and problems in language testing 7* (pp. 113-138). Colchester: University of Essex.

Appendix: C-test passages

Basic Form A:

1. The evening news on television is very popular with many Americans. They li_____ to fi_____ out wh_____ is happ_____ in t_____ world. O_____ television th_____ can s_____ real peo_____ and pla_____. They bel_____ it i_____ easier th_____ reading t_____ newspaper. Ma_____ people th_____ television ma_____ the ne_____ seem mo_____ real. Th_____ also th_____ the ne_____ on telev_____ is mo_____ interesting. T_____ television news reporters sometimes tell funny stories and even jokes. This makes the news about wars and crime seem less terrible.

2. The journalist must write only what is true. He mu_____ never cha_____ facts t_____ please a_____ person o_____ any gr_____. You sho_____ know n_____ that so_____ of yo_____ readers m_____ not li_____ and m_____ even b_____ angered b_____ your sto_____. If yo_____ facts a_____ true, th_____ is a_____ that i_____ important. I_____ is n_____ necessary f_____ you t_____ please everyone. It is much better for your readers to get news which is complete and true.

3. Luckily he was flying at a good height when this misfortune happened, and he had time to look for a place to land. Below h_____ he co_____ see a num_____ of fie_____ which loo_____ flat eno_____ to la_____ on, a_____ he succ_____ in brin_____ his mac_____ down o_____ one o_____ them. I_____ was n_____ as fl_____ as i_____ had loo_____ from t_____ air, b_____ he lan_____ safely a_____ jumped o_____ to lo_____ round, wond_____ where he was.

4. What has occurred since the last Ice Age, when the Alps were buried under an ice cap so deep that only the highest peaks extended above the glacier? Several mil_____ years a_____, as t_____ ice mel_____, it scr_____ the moun_____ until th_____ became st_____ rocky wa_____ with U-sh_____ valleys bet_____ them. So_____ of t_____ valleys we_____ also dam_____ up b_____ rock a_____ soil, s_____ that wh_____ the gla_____ melted compl_____ it le_____ many cl_____ mountain la_____ behind i_____.

Basic Form B:

1. Fog is really a low cloud near the ground. Fog a_____ clouds a_____ made o_____ many lit_____ drops o_____ water. Th_____ drops st_____ in t_____ air bec_____ they a_____ so sm_____. You can_____ see ea_____ drop. B_____ fog c_____ make i_____ hard t_____ see ot_____ things. I_____ can b_____ dangerous i_____ you a_____ driving, f_____ example. Some_____ where th_____ is a lot of fog you cannot see the road.

2. It was once more common than it is now to see men open doors and let women go ahead of them. On bu_____ and tra_____ no gent_____ would rem_____ in h_____ seat wh_____ a wo_____ had t_____ stand. M_____ not on_____ stood i_____ the tic_____ line t_____ buy mo_____ tickets, b_____ they pa_____ for t_____ women's tic_____ as we_____. Men cal_____ women o_____ the tele_____, but wo_____ did n_____ call m_____. A man always walked on the outside of the sidewalk to keep his woman companion away from the danger of the street.

3. There is a saying in the United States that "You are what you eat." While exagg_____, this stat_____ reflects t_____ strong li_____ between di_____ and hea_____. Most peo_____, of cou_____, pay sc_____ attention t_____ their di_____, eating what_____ the trad_____ of fam_____ and cul_____ dictate. Even med_____ doctors sel_____ study t_____ relationship bet_____ diet a_____ health. They pre_____, instead, t_____ treat a pat_____ after h_____ becomes i_____, through use of medicines and surgery.

4. Giraffes, who live in the African grasslands south of the Sahara Desert, eat the highest leaves, twigs, and fruit from trees. After th_____ swallow th_____ food, th_____ bring i_____ up ag_____ and ch_____ it t_____ way co_____ chew th_____ cud. I_____ order t_____ drink, t_____ giraffe spr_____ its fr_____ legs wi_____ apart a_____ lowers i_____ head bet_____ them t_____ lap u_____ the wa_____ in a str_____ or ri_____. Giraffes usu_____ sleep stan_____ up, but the few times they lie down they keep their necks upright, often resting their heads on the low branch of a tree.

Experimental Form A:

1. The evening news on television is very popular with many Americans. Th_____ like t_____ find o_____ what i_____ happening i_____ the wo_____. On telev_____ they c_____ see re_____ people a_____ places. Th_____ believe i_____ is eas_____ than rea_____ the news_____. Many peo_____ think telev_____ makes t_____ news se_____ more re_____. They al_____ think t_____ news o_____ television i_____ more inter_____. The television news reporters sometimes tell funny stories and even jokes. This makes the news about wars and crime seem less terrible.

2. The journalist must write only what is true. H_____ must ne_____ change fa_____ to ple_____ any per_____ or a_____ group. Y_____ should kn_____ now th_____ some o_____ your rea_____ may n_____ like a_____ may ev_____ be ang_____ by yo_____ stories. I_____ your fa_____ are tr_____, that i_____ all th_____ is impo_____. It i_____ not nece_____ for y_____ to please everyone. It is much better for your readers to get news which is complete and true.

3. Luckily he was flying at a good height when this misfortune happened, and he had time to look for a place to land. Be_____ him h_____ could s_____ a number o_____ fields wh_____ looked fl_____ enough t_____ land o_____, and h_____ succeeded i_____ bringing h_____ machine do_____ on o_____ of th_____ It w_____ not a_____ flat a_____ it h_____ looked fr_____ the a_____ but h_____ landed saf_____ and jum_____ out t_____ look ro_____, wondering where he was.

4. What has occurred since the last Ice Age, when the Alps were burried under an ice cap so deep that only the highest peaks extended above the glacier? Sev_____ million ye_____ ago, a_____ the i_____ melted, i_____ scraped t_____ mountains un_____ they bec_____ steep ro_____ walls wi_____ U-shaped val_____ between th_____. Some o_____ the val_____ were al_____ dammed u_____ by ro_____ and so_____, so th_____ when t_____ glacier mel_____ completely i_____ left ma_____ clear moun_____ lakes beh_____ it.

Experimental Form B:

1. Fog is really a low cloud near the ground. F_____ and clo_____ are ma_____ of ma_____ little dr_____ of wa_____.These dr_____ stay i_____ the a_____ because th_____ are s_____ small. Y_____ cannot s_____ each dr_____. But f_____ can ma_____ it ha_____ to s_____ other thi_____. It c_____ be dang_____ if y_____ are dri_____, for exa_____. Sometimes wh_____ there is a lot of fog you cannot see the road.

2. It was once more common than it is now to see men open doors and let women go ahead of them. O_____ buses a_____ trains n_____ gentleman wo_____ remain i_____ his se_____ while a woman h_____ to st_____. Men n_____ only st_____ in t_____ ticket li_____ to b_____ movie tic_____, but th_____ paid f_____ the wom_____ tickets a_____ well. M_____ called wo_____ on t_____ telephone, b_____ women d_____ not ca_____ men. A m_____ always walked on the outside of the sidewalk to keep his woman companion away from the danger of the street.

3. There is a saying in the United States that "You are what you eat." Wh_____ exaggerated, th_____ statement refl_____ the str_____ link bet_____ diet a_____ health. Mo_____ people, o_____ course, p_____ scant atte_____ to th_____ diet, eat_____ whatever the tradi_____ of fam_____ and cul_____ dictate. Ev_____ medical doc_____ seldom st_____ the relati_____ between di_____ and hea_____ They pre_____, instead, t_____ treat a pat_____ after h_____ becomes ill, through use of medicines and surgery.

4. Giraffes, who live in the African grasslands south of the Sahara Desert, eat the highest leaves, twigs, and fruit from trees. After th_____ swallow th_____ food, th_____ bring i_____ up ag_____ and chew i_____ the w_____ cows ch_____ their c_____. In or_____ to dr_____, the gir_____ spreads i_____ front le_____ wide ap_____ and low_____ its he_____ between th_____ to l_____ up t_____ water i_____ a stream o_____ river. Gira_____ usually sl_____ standing up, but the few times they lie down they keep their necks upright, often resting their heads on the low branch of a tree.