

# 琉球大学学術リポジトリ

## クローズ法の識別力についての実験

メタデータ	言語: 出版者: 琉球大学法文学部 公開日: 2010-01-26 キーワード (Ja): キーワード (En): 作成者: Yamauchi, Susumu, 山内, 進 メールアドレス: 所属:
URL	<a href="http://hdl.handle.net/20.500.12000/15227">http://hdl.handle.net/20.500.12000/15227</a>

An Experiment with Cloze Procedure on Japanese EFL Learners:  
On the Diagnostic Power of Cloze Procedure\*

Susumu Yamauchi

INTRODUCTION

Cloze procedure, in which every  $n$ th word is deleted from a reading passage and a testee is asked to restore such words, was originally introduced by Taylor (1953) as a method for measuring English native-speakers' ability to read a passage. Later some researchers found that the procedure could also be a useful technique for assessing non-native speakers' reading comprehension and/or overall language proficiency in various ESL/EFL settings (e.g., Aitken, 1977; Anderson, 1972; Bachman, 1982; Briere, Clausing, Senko, & Purcell, 1978; Brown, 1983; Darnell, 1968; Hanania & Shikhani, 1986; Irvine, Atai, & Oller, 1974; Oller, 1972, 1973; Stubbs & Tucker, 1974; Wainman, 1979). Other researchers, however, criticized the technique for not being fully effective for that purpose, claiming that in various experimental studies (e.g., Alderson, 1979, 1980; Klein-Braley, 1983; Porter, 1978; Shanahan, Kamil, & Tobin, 1982) the reliability and validity of cloze tests were neither satisfactory nor consistent.

In order to investigate the reliability and validity of the cloze procedure, researchers have controlled numerous variables such as scoring methods, deletion frequency, textual difficulty, and number of cloze items and have been able to reach a certain degree of consensus on some of these variables (Brown, 1989). Although there is still controversy concerning the use of the cloze procedure, an area for minimum agreement seems to be that cloze tests are cloze tests, "just like other testing techniques"

(Alderson, 1979; Brown, 1989). Whether one approves or disapproves of the technique, it remains widely used in many ESL/EFL institutions both as a testing and a teaching device. This is also indicated by the fact that the cloze procedure is discussed in many ESL/EFL testing and reputable textbooks, especially those published in the 1980s (Baker, 1989; Briere & Hinofotis, 1979; Cohen, 1980; Croft, 1980; Heaton, 1988, 1990; Hubbard, Jones, Thornton, & Wheeler, 1983; Madsen, 1983; Oller, 1979a, 1980, 1983).

Although cloze procedure is certainly a much-debated testing device in the field of ESL/EFL testing abroad, the situation in Japan is not necessarily the same. It is true that the procedure is mentioned in some English language education methodology textbooks and monographs recently published in Japan (Aoki, Tanaka, Yamaoka, & Yorozuya, 1989; Oozeki, Takanashi, & Takahashi, 1983; Sato, 1988; Takahashi, 1990), but there seems to be insufficient empirical data both quantitatively and qualitatively on the use of the procedure with Japanese students as subjects. As Brown (1989) pointed out, we cannot simply extrapolate from the results of experimental studies carried out in ESL settings in Britain and the United States to the situation in Japan; there are differences between ESL and EFL students in terms of nationality, language background, educational level, age, social environment, motivation and so on.

The necessity of gathering data on cloze tests with Japanese students as subjects is also pointed out by Sato (1988). In his book *The Role of Cloze Testing in English Teaching*, he devotes his entire discussion to reviewing the background and history of cloze procedure, summarizing past research studies undertaken mainly in ESL settings, and introducing various uses of the procedure as testing and teaching devices applicable to Japanese situations. He speculates that cloze tests are innovative

devices for measuring overall language proficiency in that they display various excellent characteristics which traditional language tests do not possess. He concludes that it is imperative to accumulate empirical data on cloze tests in Japanese settings in order to understand the possibilities and the limitations of the tests when used in the Japanese educational system.

The aim of this study, then, is to present some experimental data concerning the use of cloze procedure with Japanese university students as subjects. The specific research question investigated in this study is the diagnostic power of the cloze procedure. Two types of diagnostic power of the cloze procedure should be distinguished here. The first type concerns passage difficulty. Originally, Taylor (1953) claimed that cloze test scores could discriminate the difficulty level of passages. His experiment showed that "cloze scores ranked the three selected passages in the same order of readability as do the Flesch and Dale-Chall formulas" (p. 422). The same result was obtained by Anderson (1972), though instead of using the readability formulas, he asked four experienced teachers to rank the passages by level of difficulty. Takanashi (1983), in a study with Japanese students, found that scores in *modified* cloze tests discriminated between four passages in terms of readability, and thus partially confirmed the diagnostic power of cloze procedure. This type of research was referred to as "readability estimate" in a study by Greene (1965).

The second type of diagnostic power of the cloze procedure concerns the effectiveness of ranking examinees according to their proficiency. Oller and Conrad (1971) confirmed the assumption in research with 5 non-native and 2 native groups as controls. These subjects were categorized into one of the following proficiency levels; (I) beginning ESL, (II) intermediate ESL, (III) advanced ESL, (IV) advanced

composition ESL, (V) non-native TEFL graduate students, (VI) native Freshman composition students, and (VII) native TEFL graduate students. A cloze test administered to these groups ranked the groups as expected, with the exception of groups IV and V, in terms of their mean scores. The same results have been reported by some researchers (e.g., Anderson, 1972; Briere, Clausing, Senko, & Purcell, 1978; Templeton, 1977). The method of comparing subjects' cloze scores, in order to ascertain their competence, was referred to as "comprehension scores" analysis by Greene (1965).

Thus, the results seem consistent in ESL/EFL settings, but whether the same assumption holds true in the Japanese EFL setting needs careful consideration.

The research questions posed in this study are as follows:

- (1) Is standard fixed-ratio cloze procedure a reliable and valid measure of overall language proficiency? (While this is not necessarily the main point of interest in this study, it is a prerequisite condition for the subsequent research questions. Without high reliability and validity, the discussion of the test results will be downgraded to some degree.)
- (2) Can standard fixed-ratio cloze procedure provide a satisfactory means for measuring the difficulty of different reading passages?
- (3) Can standard fixed-ratio cloze procedure provide a satisfactory basis for assessing subjects in terms of levels of language proficiency?

## METHOD

### Subjects

The subjects in this study consisted of 81 freshmen enrolled in the General Education Division at University of the Ryukyus. The students' major subjects were English (37 students), Business Administration (21), Sociology (8), Elementary Education (7), Japanese Language and

Literature (4), History (3), and Law and Political Science (1). The subjects were all native speakers of Japanese and included 55 females and 26 males.

### Materials

Three books from L. A. Hill's (1980 a, b, c) *Steps to Understanding* series were selected for the construction of cloze tests. The criterion for selection was based on the ease with which the difficulty level of the reading passages could be decided. The three books in the series are differentiated by the author according to the vocabulary and the grammatical structures as follows:

Elementary level	----	1000—headword level
Intermediate level	----	1500—headword level
Advanced level	----	2075—headword level

The distinction according to vocabulary and sentence structures seems reasonable since they are fundamental to various language activities especially for non-native speakers. A word list is included at the end of each book, though it is not made clear what kinds of grammatical sentence structures are controlled at each level. Another advantage of using this series is that the contents of the books comprise humorous short stories with a chronologically arranged paragraph developments. Students, thus, need no special background knowledge to understand these stories. So, it is rightly assumed that the level differences between these stories are mainly accounted for owing to vocabulary and grammatical structures.

Two independent stories were selected from each book for the construction of cloze tests. In principle, every 5th word was deleted from the passages, beginning from the second sentence, and thereby creating 25 blanks in each passage. The first sentence in each cloze test was left

uncut to give students necessary and adequate contextual information. The selected stories were taken from the latter part of the books, but the selection process was not necessarily random. Not all stories were long enough to make 25 cloze items, and therefore only stories that fulfilled this demand were chosen. But no attention was paid to the contents of the stories when making a suitable selection. As a result, lessons 21 and 27 were chosen from the elementary level, lessons 27 and 30 from the intermediate level, and lessons 23 and 27 from the advanced level respectively. The final result was six cloze tests, consisting of two tests from each different level. A sample cloze test (Test 1) is given in the Appendix.

To check the criterion-related validity of the cloze tests, the English Test in the College Entrance Examination Center Test Battery (Center Test) was used as a criterion. Which test to choose as an external criterion is a difficult question especially in Japan, since there is no consensus on it. The selection of the Center Test is of interest because it alone is used in every national and public university as well as in some private universities throughout Japan. The content areas tested in the Center Test are summarized as follows: (1) accent & sentence stress (2) vocabulary, word usage, & grammar (3) dialogue fill-in (4) controlled composition and (5) reading comprehension. Considering these content areas of the Center Test, one may doubt whether it can be a reliable and valid measure of overall language proficiency because it does not include a listening comprehension section. Although one must acknowledge that Center Test is not a perfect model of a proficiency test, it is nevertheless and widely assumed to be a well-established, standardized test measuring various aspects of language abilities of college-bound Japanese students.

Students were required to give "self-estimated" scores in the Center

Test, which they took in January before entering the university. Though it was impossible to obtain students' "real" scores in the Center Test, their "self-estimated" scores were assumed to be fairly close to the real scores.

#### Procedure

The six cloze tests were administered in a regular classroom session period in April, 1990. The cloze tests were administered to all of the students. Students were instructed first to read the whole of each story to grasp the outline and then to fill in each blank with an appropriate word. It was explained that contracted forms were also to be treated as one word. Some examples of the contracted forms were printed at the top of the first page of the cloze tests. The instructions and the explanation were given in Japanese.

The students were allowed 80 minutes, plus 10 extra minutes if necessary, to complete the whole set of the cloze tests, in which there were a total of 150 blanks. Cohen (1980) recommended a 60-minute time allotment for a cloze passage with 50 blanks (p. 97). Judging from his criterion, the time allotment of 80-90 minutes for 150 cloze items was apparently too short. His criterion, however, should not be rigidly applied because the time needed for completing cloze tests depends greatly on the difficulty of the given passages and the ability of each student. Because Japanese students were expected to learn many more than 2,075 words during high school years, the passages used in this study might be judged rather easy for Japanese college students in terms of vocabulary. Besides, it is a truism that teachers do not always allow ample testing time in regular test settings. The Test of English as a Foreign Language (TOEFL), for example, has been often used as an external criterion measure of cloze tests in various studies.



But it is a fact that there is a time restriction in the TOEFL: no testees are allowed sufficient time to complete the test.

In scoring the cloze tests, the acceptable word method was employed throughout in this study. According to this method, any contextually acceptable word replacement was counted as correct. While research has repeatedly found high correlations between the exact-word method and the acceptable-word method, the latter has been recommended for use in EFL settings by some researchers (Brown, 1980; Oller, 1972).

## RESULTS AND DISCUSSION

The basic descriptive data for each of the cloze tests made over the six passages and the Center Test is shown in Table 1. Included in the statistics are the number of cloze items (CI), the mean (M), the range (R), the standard deviation (SD), the reliability coefficients by the Kudar-Richardson formula 20 (KR-20) and by the Kudar-Richardson formula 21 (KR-21).

Table 1 clearly shows that the students' performance decreased as the difficulty of the chosen passages increased. The mean in test 1 and 2 were higher than those in test 3 and 4. The mean scores of the advanced cloze tests were revealed to be lower than the other two levels.

Another notable characteristic observed in Table 1 is that the mean score of test 6 was exceptionally low. The two most plausible explanations for this result were (1) that the students did not have enough time to complete this last cloze test because of the time restriction, and/or (2) that test 6 itself was extremely difficult. As for the first explanation, discussed in the procedure section, the time allotted in this experiment might not be considered satisfactory. But it is not known precisely what effect the time limitation had on the above result because some students commented that there had been enough

TABLE 1  
 Characteristics of Cloze Tests and the Center Test

Test		CI	M	R	SD	KR-20	KR-21
Elementary	1	25	15.73	1 - 22	3.37	.66	.51
	2	25	17.83	7 - 22	2.92	.45	.42
Intermediate	3	25	14.73	2 - 22	3.74	.71	.59
	4	25	15.64	4 - 24	3.58	.72	.57
Advanced	5	25	12.73	1 - 21	3.81	.72	.59
	6	25	7.67	2 - 15	3.37	.63	.55
Center Test			160.24	99 - 116	18.03		

time but that the last test had been especially difficult, while other students had reported that they needed a little more time to complete the whole test. Admittedly, the time allotment may have had an effect on some students, but that might not be the sole reason for the lowest mean score in test 6.

The other possible explanation is that test 6 might simply have been difficult for most of the students for some unidentified reason. It should be also noticed that the highest score obtained in the test was only 15 points. This was remarkably low, compared to 21 points in test 5, which was also extracted from the same advanced textbook. What made test 6 more difficult than test 5 was not clear and beyond the scope of this study. It might be added that theorists and researchers have found that there are various factors involved in reading comprehension, such as difficulty of concept, paragraph development patterns, students' background knowledge, their motivation, typography, and so on. Concerning the textual difficulty of cloze tests, Oller (1972)

found in particular that one of the influential factors involved in readability estimates of the cloze procedure was the degree of redundancy in the chosen passages (p. 155-56). Brown (1989), on the other hand, suggested that "for Japanese students, lexical factors are more highly related to performance on individual items than the other factors" (p. 61). Further research is needed to reach a consensus on the factors involved in readability estimates of cloze procedure.

The reliability coefficient of each cloze test is also shown in the table. While there exist various ways to estimate test reliability, the three most frequently used methods are test-retest reliability, equivalent forms reliability, and internal consistency reliability. Among these, internal consistency reliability estimates "have the distinct advantage of being estimable from a single form of a test administered only once - contrast to test-retest and equivalent forms reliabilities" (Brown, 1988, p. 99). A review of the past literature on cloze procedure revealed that it was a common practice for many researchers to report reliability coefficients according to the Kuder-Richardson formula 20, one of the methods of calculating internal consistency reliabilities. Likewise, this study reported the KR-20 reliability coefficients, but it also included reliability estimates calculated by the KR-21 formula, which has also been used by some researchers, to check the reliability of cloze tests further.

One of the conspicuous characteristics observed in the column of KR-20 is that the reliability coefficient of test 2 is exceptionally low (.45), while the coefficients of the other tests could be estimated to be moderately high, ranging from .63 to .74 in a rather consistent manner. The same results were also found in the KR-21 column, in which every coefficient became lower variably in every case. Henning (1987) commented that the KR-21 formula is less accurate than the KR-20 formula because it underestimates the reliability. The same tendency was also observed

by Yamauchi (1989).

How much reliability coefficient is required for a test to be considered good seems to be a somewhat difficult issue yet to be resolved by researchers. Harris (1969) suggested that reliability should be above .90 at least for a Standard Proficiency Test, but that it could be above .70 or .80 for teacher-made tests for classroom use. Carmines and Zeller (1979) recommended an above .80 reliability coefficient for a test to be regarded reliable. As observed in Table 1, no tests reached the .80 level coefficients and, furthermore, half of the results yielded coefficients lower than a .70 level. This should not be interpreted, however, as evidence of the cloze tests being unreliable. As theorists have observed, reliability certainly decreases if the number of test items is reduced (see Oller, 1983, p. 227 for discussion). Apparently, the number of cloze items investigated in this study was very small, but in reality few teachers would be convinced that students could be reliably tested with a one-shot 25-item cloze test. The subsequent question then is : what kind of results could be obtained if the number of cloze items in a test is doubled? Table 2 was compiled in order to answer this question.

Table 2 was constructed by simply compiling data in Table 1. Cloze tests 1 and 2 were combined and referred to operationally as the Elementary-level Cloze Test, tests 3 and 4 as the Intermediate-level ClozeTest, and tests 5 and 6 as the Advanced-level Cloze Test respectively. It should be noticed that the number of cloze items (CI) became 50, the number of cloze items usually recommended by some researchers as appropriate for gathering data (e.g., Cohen, 1980; Oller, 1972).

The results as observed in the table clearly indicated that reliability coefficients were quite satisfactory if 50-item cloze tests were employed, though the Elementary Test yielded a coefficient slightly below the .80 level when calculated by the KR-20 formula. It is also interesting to

TABLE 2  
 Characteristics of Combined Cloze Tests

Cloze Test	CI	M	R	SD	KR-20	KR-21
Elementary-level	50	33.56	8 - 44	6.02	.78	.71
Intermediate-level	50	30.37	8 - 44	6.34	.82	.72
Advanced-level	50	20.40	3 - 35	6.15	.80	.70
Total	150	84.33	19 - 116	16.08	.90	.86

note that very high coefficients, i.e., .90 by the KR-20 and .86 by the KR-21, were gained when these three cloze test results were combined as shown raw in the Total.

As an external criterion-related validity check, the cloze test scores were correlated with the results of the Center Test. Since 50-item cloze tests resulted in a more reliable measurement, the results of the three level-differentiated cloze tests in Table 2 were employed in the following calculation. Results are shown in Table 3.

Correlation coefficients between the cloze tests ranged from .565 to .778, showing a moderately high positive correlation between these tests. Klein-Braley (1983), in discussing the validity of cloze procedure, demanded "evidence collected from two (or more) cloze tests administered to the same examinee" (p. 222) to show that all cloze tests were essentially parallel or equivalent test in any ESL/EFL setting. While the correlations between the cloze tests in this study were not higher than the .80 level, the data seemed quite satisfactory. Though interpreting the correlation coefficient is not so easily accomplished, it seemed fair to assert, according to the general guideline indicated in Ohtomo (1969) or Kiyokawa (1990), that correlation coefficients between .565 to .778

TABLE 3  
Correlations between Cloze Tests and Center Test

	1	2	3	4	5
1 Elementary	1.000	0.778	0.565	0.852	0.509
2 Intermediate	0.778	1.000	0.632	0.898	0.646
3 Advanced	0.565	0.632	1.000	0.811	0.513
4 Cloze Total	0.852	0.898	0.811	1.000	0.618
5 Center Test	0.509	0.646	0.513	0.618	1.000

All correlations are significant at  $p < 0.01$ .

indicated a fairly high relationship between these tests.

Table 3 also shows correlations between the cloze tests and the Center Test. Correlations ranged from .509 to .646. The highest correlation was found between the Intermediate test and the Center Test. The correlation between Cloze Test Total and the Center Test was .617. Shimizu (1978) stated that, compared to other correlation coefficients, concurrent and predicative validity coefficients usually tend to be low because they are correlations between different kinds of tests, thus various uncontrollable factors are involved. He estimated that typical validity correlations are around .60 and that correlations above .70 were very high. Therefore, we could conclude that a correlation coefficient of .617 between the Cloze Test Total and the Center Test was estimated to be quite satisfactory.

One of the main objectives of this study was to determine whether cloze procedure could demonstrate differences between passages in terms of reading difficulty. As seen in Table 2, differences were observed in the mean scores among all three cloze tests. To test simultaneously the significance of the differences between the means, an analysis of variance

(ANOVA) was performed. Since the subjects were not independent in this study, the ANOVA by block design was employed for analysis. The results are reported in Table 4.

As the overall F-value was found to be significant at the  $p < .01$  level, the least significant difference (LSD) was computed to make the individual mean comparison possible. The LSD was computed to be 3.549 at the  $p < .01$  level and 2.510 at the  $p < .05$  level. Table 5 shows the results of the LSD test. Differences between all three pairs were statistically significant either at the  $p < .05$  or at the  $p < .01$  level. It was

TABLE 4  
ANOVA Results

Source	SS	df	MS	F
Between groups	7639.971	2	3818.486	452.534**
Within groups	7543.885	80	94.299	11.178*
Error	1350.029	160	8.438	
Total	75345.738	243		

\*\*  $p < .01$

\*  $p < .05$

TABLE 5  
LSD Test for Differences between Cloze Tests

	Intermediate	Advanced
Elementary	3.186*	13.161**
Intermediate		9.975**

\*\*  $p < .01$

\*  $p < .05$

therefore concluded that the three cloze tests were different in difficulty for the 81 subjects in this study.

To investigate whether the cloze procedure can satisfactorily measure subjects in terms of their levels of language proficiency, the 81 subjects were divided into three groups according to their scores in the Center Test. Each subdivided group was made up of 27 students (1/3 of the sample) and operationally referred to as the lower group, middle group, and upper group respectively. The ranges of the means in the Center Test were 99-150 in the lower group, 155-166 in the middle group, and 167-192 in the upper group respectively. The mean scores (M), the standard deviation (SD), and the range (R) of each cloze test for all groups are to be found in Table 6.

TABLE 6

Mean Scores, Standard Deviations, and Range of Three Groups on Cloze Tests

Cloze Test	Group	M	SD	R
(a)Elementary-level	Lower Group	30.15	5.85	8 - 38
	Middle Group	33.33	5.44	15 - 41
	Upper Group	36.74	4.11	28 - 44
(b)Intermediate-level	Lower Group	25.48	5.60	8 - 34
	Middle Group	30.74	5.19	16 - 44
	Upper Group	34.63	4.51	25 - 42
(c)Advanced-level	Lower Group	16.85	5.32	3 - 29
	Middle Group	19.78	4.87	10 - 31
	Upper Group	24.93	5.50	14 - 35



TABLE 7  
ANOVA Results

(a) Elementary-level Cloze Test

Sources	SS	df	MS	F
Between Groups	581.556	2	290.778	16.454**
Within Groups	1378.445	78	17.672	
Total	1960.001	80		

(b) Intermediate-level Cloze Test

Sources	SS	df	MS	F
Between Groups	1008.358	2	504.179	22.259**
Within Groups	1766.778	78	22.651	
Total	2775.136	80		

(c) Advanced-level Cloze Test

Sources	SS	df	MS	F
Between Groups	1655.580	2	827.79	12.663**
Within Groups	5098.773	78	17.672	
Total	6754.353	80		

\*\*  $p < .01$

An examination of Table 6 clearly shows that as the students' level of language proficiency increased, their performance on the cloze tests also increased in every case. A one-way ANOVA was conducted in each

level to ascertain if the difference between the means of the three groups was significant. The ANOVA results of the elementary-level cloze test, intermediate-level cloze test, and advanced-level cloze test are shown in Table 7(a), 7(b), and 7(c) respectively.

An obtained F-value was found significant at the  $p < .01$  level of significance in every case, meaning that there was an overall significant difference with the probability of 99%. Multiple comparison analysis, therefore, was performed to see which pair had a significant difference in each level-differentiated cloze test. Analysis was made by computing the LSD. The results are shown in Table 8(a), 8(b), and 8(c) respectively.

As shown in the table, the results were quite satisfactory. The differences between each pair in the elementary-level and the intermediate-level cloze tests were all found to be significant at the  $p < .01$  level. In the advanced-level cloze test, however, the difference between the lower group and the middle group was significant neither at the  $p < .01$  level nor at the  $p < .05$  level. So, taken as a whole, the three cloze tests

TABLE 8  
The LSD Test

(a)Elementary Cloze			(b)Intermediate Cloze			(c) Advanced Cloze		
	M.G.	U.G.		M.G.	U.G.		M.G.	U.G.
L.G.	3.556**	6.556**	L.G.	5.223**	9.482**	L.G.	2.555	8.074**
M.G.		3.000**	M.G.		4.259**	M.G.		5.519**

L.G.=Lower Group

M.G.=Middle Group

U.G.=Upper Group

\*\*  $p < .01$

provided a satisfactory means for measuring the differences between these students in the three language proficiency levels. This means, assuming that the Center Test provide satisfactory assessment of students in terms of levels of language proficiency, the cloze tests in this experiment achieved exactly the same result.

## CONCLUSION

(1) Is standard fixed-ratio cloze procedure a reliable valid measure of overall language proficiency?

The results of this study demonstrate that, for the three 50-item fixed-ratio cloze tests analyzed in this experiment, reliability coefficients were found to be moderately high, ranging from .78 to .82 with an average of .80 by the KR-20 and from .71 to .72 with an average of .71 by the KR-21. Furthermore, the total reliability coefficient, i.e., combined test scores of the above three tests, proved to be surprisingly high: it was estimated to be .90 by the KR-20 formula and .86 by the KR-21 formula. It was also found, however, that, when the number of test items became 25, the reliability coefficients decreased to a certain degree.

In terms of validity, the correlation coefficient between the cloze tests and the Center Test was calculated to arrive at an external criterion-validity estimate. The use of the Center Test as a criterion here was based on the assumption that it was a well-established, standardized test of measuring overall English proficiency for Japanese English learners. So the correlation coefficients between the Center Test and the cloze tests might be considered valid indices of the cloze tests as a measure of overall English proficiency. The correlation coefficients between the Center Test and the three cloze tests ranged from .509 to .646 with an average of .556, and the correlation between the Center Test and the total cloze test were found to be .618, showing a moderately high relationship.

So, we were able to conclude that the cloze tests analyzed here had a fairly high validity as a measure of overall English proficiency for Japanese students. But some researchers demanded .80 above the validity coefficient for a test to be seen as valid. So further research seems necessary for the validity estimate of cloze tests.

(2) Can standard fixed-ratio cloze procedure provide a satisfactory means for measuring the difficulty of different reading passages?

The results of this study found positive answer to the above question. The three cloze tests analyzed in this study were level-differentiated as elementary, intermediate, and advanced texts beforehand by the author. The distinction, according to the author, was made in terms of level of vocabulary and grammatical sentence structure. The scores of the 81 subjects could well reflect the difference between these passages in terms of difficulty, just as the author had designated. Differences between mean scores in the three cloze tests were all found statistically significant at either the  $p < .01$  or at the  $p < .05$  level.

(3) Can standard fixed-ratio cloze procedure provide a satisfactory basis for assessing subjects in terms of language proficiency?

To check the potential value of cloze procedure as a measure of the subjects' language proficiency, 81 subjects were divided into three proficiency groups (lower, middle, and upper) according to the scores of the Center Test. Each group consisted of 27 students, i.e., of  $1/3$  of the sample. What we were interested in here was whether three cloze tests (elementary, intermediate, and advanced) satisfactorily measure the difference these groups, as the Center Test had. This analysis was based on the assumption that the Center Test was able to measure the difference between the students in terms of the level of language

proficiency. The results were reasonably satisfactory. In the elementary and intermediate cloze tests, the subjects had been well sorted into three groups. In the advanced cloze tests, the mean difference between the upper group and the other two groups was significant at the  $p < .01$  level, but the difference between the middle group and the lower group was not significant at the  $p < .05$  level. So, taken as a whole, the data obtained here showed the provable value of the cloze procedure as a means of assessing language proficiency.

#### \* Acknowledgement

I would like to thank Professors Katsunobu Sunagawa, B.M. Kresovich, and A.P. Jenkins for reading and commenting on the first draft of this paper. Responsibility resides with the writer for any remaining errors.

#### References

- Aitken, K. G. (1977). Using cloze procedure as an overall language proficiency test. *TESOL Quarterly*, 11, 59-67.
- Alderson, J. C. (1979). The cloze procedure and proficiency in English as a foreign language. *TESOL Quarterly*, 13, 219-227.
- Alderson, J. C. (1980). Native and connative speaker performance on cloze tests. *Language Learning*, 30, 59-76.
- Anderson, J. (1972). The application of cloze procedure to English learned as a foreign language in Papua and New Guinea. *English Language Teaching*, 26, 66-72.
- Aoki, S., Tanaka, M., Yamaoka, T., & Yorozyua, R. (1989). *Eigo no Testing* [Testing in English]. Tokyo: Kairyudo.
- Bachman, L. F. (1982) The trait structure of cloze test scores. *TESOL*

*Quarterly*, 16, 61-70.

- Baker, D. (1989). *Language testing: A critical survey and practical guide*. London: Edward Arnold.
- Briere, E. J., Clausing, G., Senko, D., & Purcell, E. (1978). A look at cloze testing across languages and levels. *Modern Language Journal* 62, 23-26.
- Briere, E. J., & Hinofotis, F. B. (Eds.). (1979). *Concepts in language testing: Some recent studies*. Washington DC: TESOL.
- Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. *Modern Language Journal*, 64, 311-317.
- Brown, J. D. (1983). A cloze is a cloze is a cloze. In J. handscombe, R. A. Orem, & B. P. Taylor (Eds.), *On TESOL '83; The question of control* (pp. 109-119). Washington DC: TESOLL.
- Brown, J. D. (1988). *Understanding research in second language learning: A teacher's guide to statistics and research design*. London: Cambridge University Press.
- Brown, J. D. (1989). Cloze item difficulty. *JALT Journal*, 11, 46-67.
- Carmines, G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.
- Cohen, A. D. (1980). *Testing language ability in the classroom*. Rowley, MA: Newbury house.
- Croft, K. (Ed.). (1980). *Readings on English as a second language: For teachers and teacher trainees* (2nd ed.). Boston: Little, brown and Company.
- Darnell, D. K. (1968). *The development of an English language proficiency test of foreign students, using a clozentropy procedures. Final report to DHEW Bureau No. BP-7-H-010*. Boulder, Colorado: University of Colorado. (ERIC ED 029 039)
- Greene, F. P. (1965). Modification of the cloze procedure and changes

- in reading test performances. *Journal of Educational Measurement*, 2, 213-217.
- Harris, D. (1969). *Testing English as a second language*. New York: McGraw-Hill.
- Hanania, E., & Shikhani, M. (1986). Interrelationships among three tests of language proficiency: Standardized ESL, cloze, and writing. *TESOL Quarterly*, 20, 97-109.
- Heaton, J. B. (1988). *Writing English language tests*. London: Longman.
- Heaton, K. B. (1990). *Classroom testing*. London: Longman.
- Henning, G. (1987). *A guide to language testing*. Rowley, MA: Newbury House.
- Hill, L. A. (1980a). *Elementary steps to understanding*. Tokyo: Oxford University Press.
- Hill, L. A. (1980b). *Intermediate steps to understanding*. Tokyo: Oxford University Press.
- Hill, L. A. (1980c). *Advanced steps to understanding*. Tokyo: Oxford University Press.
- Hubbard, P., Jones, H., Thornton, B., & Wheeler R., (1983). *A training course for TEFL*. Oxford: Oxford University Press.
- Irvine, O., Atai, P., & Oller, J. W. (1974). Cloze, dictation, and the test of English as a foreign language. *Language Learning*, 24, 245-252.
- Kiyokawa, H. (1990). *Eigo kyoiku kenkyu nyumon* [An introduction to English education research]. Tokyo: Taishukan.
- Klein-Braley, C. (1983). A cloze is a cloze is a question. In J. W. Oller (Ed.), *Issues in language testing research* (pp. 218-228). Rowley, MA: Newbury House.
- Madsen, H. (1983). *Techniques in testing*. New York: Oxford University Press.
- Oller, J. W. (1972). Scoring methods and difficulty levels for cloze tests

- of proficiency in English as a second language. *Modern Language Journal*, 56, 151-158.
- Oller, J. W. (1973). Cloze tests of second language proficiency and what they measure. *Language Learning*, 23, 105-118.
- Oller, J. W. (1979). *Language tests at school*. London: Longman.
- Oller, J. W. (Ed.). (1983). *Issues in language testing research*. Rowley, MA: Newbury House.
- Oller, J. W., & Conrad. (1971). The cloze technique and ESL proficiency. *Language Learning*, 21, 183-195.
- Oozeki, T., Takanashi, M. (1983). *Eigoka kyoiku-hou*. [Methodology in English education]. Tokyo: Kinseido.
- Porter, D. (1978). Cloze procedure and equivalence. *Language Learning*, 28, 333-341.
- Sato, S. (1988). *The role of cloze testing in English teaching*. Tokyo: Nan'un-do.
- Shanahan, T., Kamil, M. L., & Tobin, A. W. (1982). Cloze as a measure of intersentential comprehension. *Reading Research Quarterly*, 17, 229-255.
- Shimizu, T. (1978). *Kyositsu-no tokei-hou* [Statistics in classroom] Tokyo: Taken Shuppan.
- Stubbs, J. B., & Tucker, G. R. (1974). The cloze test as a measure of English proficiency. *Modern Language Journal*, 58, 239-241.
- Takanashi, Y. (1983). An experimental study of matching cloze procedure. *KASELE Bulletin*, 11, 30-39.
- Takanashi, Y., & Takahashi, M. (1990). *Eigo kyoikugaku gairon* [An introduction to English education]. Tokyo: Kinseido.
- Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism quarterly*, 30, 415-433.
- Templeton, H. (1977). A new technique for measuring listening com-



prehension. *English Language Teaching*, 31, 292-299.

Wainman, H. (1979). Cloze testing of second language learners. *English Language Teaching*, 33, 126-132.

Yamauchi, S. (1989). Cloze tests for Japanese English-learners: Their reliability and validity. *Southern Review*, 4, 126-139.

### Appendix

Bruce was an Australian and worked for a newspaper in Sydney. Then he thought, "I (1) to see Europe, so (2) go to England and (3) for a newspaper there (4) a few years."

He (5) to London and soon (6) work, because he was (7) at his job.

He (8) in a small but (9) house near London, and (10) had a small garden. (11) enjoyed working in it (12) Saturdays and Sundays. He (13) nice neighbors on both (14), and they often worked (15) their gardens on Saturdays (16) Sundays, too, and then (17) talked and joked together.

(18) day he was digging (19) hole in his garden (20) plant a bush when (21) of these neighbors came (22) the fence between the (23) gardens and looked at (24) work. He laughed and (25), "Are you making a swimming-pool?"

"Oh, no," answered Bruce, "I'm going home."

- (Answers) 1. want 2. I'll 3. work 4. for 5. flew 6. got  
7. good 8. lived 9. comfortable 10. he 11. He  
12. on 13. had 14. sides 15. in 16. and 17. they  
18. One 19. a 20. to 21. one 22. to 23. two  
24. Bruce's 25. said

## クローズ法の識別力についての実験

山内 進

クローズ法は読解力あるいは総合的語学能力を容易に、しかも客観的に測定できる可能性を持つテスト方式として最近日本でも論議されるようになったが、日本人英語学習者を対象とした実験報告はまだ十分とはいえない。本論文では特にクローズ法の文章難易度識別力及び受験者の英語能力識別力という二種類の識別力について、琉大生一年次の学生81名を対象として行った実験結果を報告し、クローズ法の持つ特性について論じた。クローズテストは初級、中級、上級のテキストから5語毎に単語を消去するスタンダード型でそれぞれ50のクローズ項目を作成、適語法により採点を行った。

これらのクローズテストの信頼性係数は、KR-20の公式では平均 .80でありKR-21では平均 .71であった。上記の三つのテストを総合して信頼性係数を算出したところ、KR-20では .90で、KR-21では .86という非常に高い数値が得られた。更に基準関連の妥当性として大学入試センターテスト(英語)との相関を調べたら、三つのクローズテスト間とは平均で .56であり、全体との相関は .62というかなり高い相関関係が見られた。

文章難易度識別力については、三つのテストの平均はそれぞれ、33.56、30.37、20.40でありこれらの平均値の差は、初級と中級で $p < .05$ レベルで有意、初級と上級、中級と上級間では $p < .01$ で有意であり、クローズ法が高い識別力を持つことが判った。

又、受験者の英語能力識別力についてもきわめて満足すべき結果が得られた。即ち、受験者をセンターテストの得点で上、中、下位のグループに分け、それぞれのグループが上記の三つのテストで得た点数を分析したら、初級、中級、のテストでは $p < .01$ で各グループ間に有意差が見られ、上級のテストでは3グループのうち2グループ間で $p < .01$ レベルで有意差があった。従って、クローズ法は、受験者の英語能力を極めて高い確率で識別できるという結果が得られた。