

THE CLOZE PROCEDURE AND ITS APPLICABILITY FOR TESTING HEBREW AS A FOREIGN LANGUAGE

by

ELANA SHOHAMY

Stanford University / University of California, Berkeley

ABSTRACT: The Cloze is a procedure widely used for testing and teaching second languages. This paper defines and describes the Cloze procedure, discusses its uses, and reviews research findings about the Cloze as a testing device for language proficiency. It then proceeds to review research findings about the Cloze specifically in Hebrew: reliability, validity, deletion rules and difficulty level of texts used for Cloze tests. These research findings indicate that the Cloze in Hebrew tends to follow similar patterns to those obtained in other languages. The last section of the paper discusses the applications and uses of the Cloze procedure for testing and teaching Hebrew.

Introduction

The Cloze is a procedure which in recent years has been widely used for testing and teaching second languages. It has been extensively researched and found to be efficient, reliable and valid. In spite of its popularity as a language proficiency testing tool in various language programs, the Cloze is not very well known in the field of teaching Hebrew. Recently, however, several research studies have been conducted on the use of the Cloze procedure for testing Hebrew language proficiency. The findings from these studies indicate that the Cloze procedure in Hebrew tends to follow patterns similar to those obtained in other languages.

The purpose of this study is to expose the reader to useful information about the Cloze procedure so that it may be used as a device for teaching and testing Hebrew.

The study consists of three main sections: the first provides definitions and background information, and reviews research findings about the use of the Cloze procedure as a testing device for language proficiency; the second section reports on research findings about the Cloze procedure in

Hebrew; the third section discusses the applicability and uses of the Cloze procedure for testing and teaching Hebrew.

1. Definitions, Background and Findings About the Cloze as a Testing Procedure

Definitions

The Cloze, first developed by Taylor in 1953, is based on the Gestalt psychological theory where one can see an overall pattern based on partial information. The term Cloze is derived from "closure" which occurs when a subject supplies a missing part based on clues available from its surroundings. For example, it is obvious that the reader knows how to complete the imperfect patterns in Figure 1.



Figure 1. Examples of Visual Closure

The Procedure

In language study, the "Cloze procedure" refers to the systematic deletion of letters, words or phrases from a continuous text. A student is then required to resupply the missing portions on the basis of contextual clues available in the remainder of the text. Filling in the blanks by guessing the missing words is a type of closure which is similar to the perceiver's completion of the imperfect visual patterns in Figure 1.

The Cloze procedure is based on the assumption that continuous texts in a language can be shown to exhibit many kinds of statistical regularity, i.e., the probability that a noun will follow an adjective is very high. Taylor (1953) stated that language structure gives rise to transitional probabilities so that the expression "Merry Christmas" is more probable in a language unit than "Merry Birthday" (p. 419). Some words are more likely than others to appear in patterns of word sequences. For example, in the sentence, "The computer's memory is _____ into sections, and each section _____ a certain number of parts in _____," it is obvious to the reader that the words most likely to appear in the sentence are: "divided," "has," and "it," respectively.

Types of Cloze

There are many types of Cloze procedures. The most commonly used, and the best researched, is the Cloze constructed by deleting every 'n'th word of a passage. According to this method, every sixth word, for example, can be deleted, resulting in one-sixth of the words deleted from a complete sentence. Cloze texts can also be constructed by randomly omitting words with the aid of a random table of numbers. Another type of Cloze is that in which words are not randomly deleted, but rather are selected on some other basis: the deleted words can be richly laden with meaning, such as verbs, adjectives, adverbs, or they can be only function words, such as prepositions and conjunctions. Valette (1977) mentions the "reading input" format, in which every fifth word presents the students with a choice between the right response and a distractor. Thus, the student does not have to interrupt his reading to write a needed word, but simply circles the preferred option.

The Cloze procedure was originally developed to measure readability of texts, but in recent years has gained popularity as a second language testing device, capable of assessing students' language proficiency in a practical and efficient way.

What Does the Cloze Measure?

The factors involved in the process of undertaking a Cloze test are not entirely known. Oller (1979), who extensively researched the Cloze procedure, writes that it would seem that the perceiver's ability to fill in gaps in imperfect patterns may be related to the ability to construct the patterns. The reader can supply the missing portions by a constructive process similar to the way the writer originally arranged the text. Oller (1973) maintains that the Cloze is a measure of the "grammar of expectancy" which he defines as "the chief mechanism underlying the skills of thinking, understanding, speaking, reading, and writing" (p. 113). He states that when a subject responds to the items on a Cloze test on the basis of incomplete information, he is required to project a word to fill in a blank and, thus, he completes a sequence. By sampling subsequent sequences, the person either confirms or denies grammatical expectations. In so doing, the "productive" and "receptive" skills of language must be used. The information provided in a Cloze test allows the students, by analysis, to synthesize a "great whole," and this synthesis may become part of the next analysis required to produce a subsequent synthesis.

Validity

The most important trait in evaluating a test is its validity, that is to say, how well a test measures what it claims to be measuring. Of all types of validity used, concurrent validity provides information on how well a performance on a specific test can predict performance on another test. Thus, high correlations between two tests indicate that similar information can be obtained from the two tests and thus, one can substitute for the other.

The strength of the Cloze procedure as a valid proficiency test has been demonstrated by correlating students' scores on Cloze tests with their scores on tests which measure other language skills, specifically reading comprehension, listening comprehension, writing, speaking and overall language proficiency. It was found that Cloze tests correlate highly with most of these tests, and especially with tests of listening comprehension and dictation.

In testing English as a second language, Darnell (1968) correlated a Cloze test with the parts and total scores of the Test of English as a Foreign Language (TOEFL) and found correlations of .73 with the Listening Comprehension section. His results were indirectly replicated several times in research with the UCLA English as a Second Language Placement Exam (ESLPE). Oller and Conrad (1971) obtained correlations of .82 between Cloze scores and dictation. Moreover, Cloze test scores have yielded correlations ranging from .69 to .89 with the total score on the UCLA ESLPE and the TOEFL tests (Oller, 1973; Oller and Streiff, 1975; Oller and Conrad, 1971).

As a reading comprehension test, Bormouth (1962) found correlations ranging from .73 to .93 between Cloze and several reading tests. The main advantage of the Cloze test as a reading comprehension test is that it does not require the reader to read the questions about the text before answering, and thus eliminates problems arising from the misinterpretation of the test questions; when undertaking a Cloze test, the test-taker responds directly to the reading passage itself.

Oller and Hinofotis (1976) correlated Cloze scores with the TOEFL, the Southern Illinois Placement Test, and the Foreign Service Institute (FSI) Oral Interview, and obtained a correlation of .71 between the Cloze and TOEFL, .80 between Cloze and the Southern Illinois Test, and .60 between Cloze and the FSI Oral Interview.

The Cloze procedure has been shown to correlate highly with tests in other languages as well: Klare, Sinaikio, and Stolorow (1971) report good success with the Cloze procedure in Vietnamese. Kerr (1972) writes that Cloze tests in French as a Second Language have the same applicability

as in English. Leong (1972) found the Cloze a suitable measure of reading comprehension in Chinese.

McLeod (1974) conducted a study in Saskatoon, Canada, designed to do a comparative assessment of reading comprehension and showed that the Cloze procedure is, in fact, "translatable." Comparable results in performance were achieved when children were tested in five languages (Czech, Polish, French, Spanish, and English).

Mason (1971) investigated the Cloze procedure as a measure of language competence for Thai students as well as native English-speaking students.

Toiemah (1978) obtained correlations ranging from .65 to .79 between two Cloze tests in Arabic and the Arab Proficiency Test which tests listening, vocabulary, grammar, reading, and writing.

Based on such high correlations, proponents of the Cloze procedure claim that the Cloze can be substituted for tests which are more complicated to construct, administer and score. This claim is made with regard to the assessment of specific skill tests such as reading, listening, writing, and speaking, as well as to overall proficiency in a language.

Reliability

The reliability of a test indicates the extent to which the test is consistent or stable and thus is an important factor in test evaluation.

Many studies have demonstrated that the reliability of the Cloze procedure ranges from .60 to .95 (Oller, 1972). Such reliability coefficients are considered high, especially for a test requiring little preparation and minimal changes in the original text.

Difficulty Level of Cloze Tests

Research has shown that the Cloze procedure may be appropriate to almost any text; a fairly wide latitude can be tolerated by the technique. However, the most appropriate material depends on the level of the students taking the particular test. It is suggested that a passage should be selected from material which is intended for a comparable level of native speakers of the target language. It is recommended that texts containing technical terms should be avoided and texts which are based on a topic with which only a number of students may be familiar should not be selected.

It has been demonstrated that the level of difficulty of the text does not significantly affect the spread of the scores which will be produced

(Oller, 1972). Tests of various difficulty levels seemed to perform very well, and it is to be expected that similar information will be obtained from Cloze tests of quite different difficulty levels. Still, it is suggested that texts which are too difficult for beginners, as well as texts which may be too easy for advanced students should not be used.

Deletion Rules

Deletion rules generally range from the fifth to seventh word. It has been shown that spaces closer than every fifth word make it difficult for even native speakers to comprehend. The deletion rule affects the difficulty level of the text, to the degree that the smaller the deletion rule, the higher the difficulty level. In addition, it also depends on the specific language: languages which are considered to be of a heavy semantic load may require larger deletion rules.

Number of Deletions

The most common number of deletions per text is about fifty and tests using this number of deletions seem to yield high reliability. (The length of a test affects its reliability, that is, the longer the test, the higher the reliability). However, a smaller number of deletions is also acceptable, depending on the situation in which the scores are to be used. On the average, about fifty minutes are required for completion of a 50-item Cloze test, depending on the difficulty level of the text, as well as on the level of the test-takers.

Scoring Methods

The two most common techniques for scoring Cloze tests are the "exact" and the "acceptable." "Exact" means that only the original word which was deleted from the text is accepted as the right answer. The "acceptable" method implies that words which were not originally used by the author of the text are counted as correct as well, provided that they fit contextually and grammatically into the sentence.

Most of the research has shown that there is no significant difference between the two scoring methods and also that correlation between the two scoring methods is very high. The "exact" method is more practical and very easy to score; the "acceptable" method provides the test user

with better discrimination among the various proficiency levels and is probably less frustrating for the test-takers.

2. Findings About the Cloze Procedure in Hebrew

Research in the use of the Cloze procedure in Hebrew has been conducted on a very limited scale. Still, from the available studies, it is clear that Hebrew Cloze tests follow similar patterns to those obtained in other languages.

Validity and Reliability

Nir and Cohen (1977) report an experiment they conducted for the Israeli Armed Forces, in which they had to place immigrant soldiers into Hebrew language classes. The researchers had to assess the Hebrew language skills of hundreds of these soldiers in a way that would be both efficient and valid. They administered a test battery consisting of a listening comprehension test, a grammar test, and a reading comprehension test. The three tests, similar to the tests used for placing students in the Israeli Ulpanim (intensive Hebrew language schools), generally require several hours to administer. In addition, the researchers decided to administer two Cloze tests—one of 14 deleted items and the other of 22 deleted items. The correlation between the scores on the first Cloze test and the combined score of the other three tests was significant, and the correlation between the second Cloze test and the total score of the other three sub-tests was .92. They concluded then that the Cloze procedure, which is more efficient and takes less time to administer and score, measures what the other three tests together measure and, when well constructed, can be used as a surrogate for a battery of tests in order to assess the language skills in Hebrew as a foreign language.

Shohamy (1978) conducted a study in which she administered two Cloze tests of 50 deleted items each to 106 students of Hebrew at the University of Minnesota. Each of the Cloze tests, one "easy," the other "difficult," was scored twice—once by the "exact" scoring method, whereby only the words which were originally deleted from the text were considered correct, while the second score was obtained by the "acceptable" scoring method, whereby any word which fitted contextually and grammatically was considered correct. She then correlated the scores obtained from the two Cloze tests with students' scores which were obtained from the Oral Interview speaking test. The Oral Interview is a test which as-

sesses oral proficiency globally as well as in the areas of grammar, vocabulary, pronunciation and fluency, on a scale ranging from 0 to 5. The correlations between the scores of the two tests ranged from .74 to .86, (see Table 1). Highest correlations were obtained between the difficult Cloze tests scored by the "acceptable" method.

Oral Interview \ Cloze	Easy Cloze Exact	Easy Cloze Acceptable	Difficult Cloze Exact	Difficult Cloze Acceptable
Overall Oral Proficiency	.81	.81	.82	.84
Vocabulary	.80	.80	.80	.82
Grammar	.81	.82	.84	.86
Pronunciation	.75	.74	.78	.79
Fluency	.78	.77	.76	.79

Table 1 Correlation coefficients between Cloze scores and Oral Interview ratings

The R^2 , which is an indicator of the amount of common variance between two tests, i.e., how well one test predicts performance on another test, can be seen in Table 2. It ranged from 65.61 to 70.56, clearly indicating that the Cloze test may be a valid measure of speaking proficiency as measured by the Oral Interview.

Oral Interview \ Cloze	Easy Cloze Exact	Easy Cloze Acceptable	Difficult Cloze Exact	Difficult Cloze Acceptable
Overall Oral Proficiency	65.61	65.61	67.24	70.56

$P \leq .001$

Table 2: R^2 figures for the Oral Interview

Shohamy also obtained reliability figures of up to .90 for the two Cloze tests used in the study, showing that the Cloze was a highly reliable test.

Difficulty Levels

In the Shohamy (1978) study, high correlations were obtained between the scores of the easy and the difficult tests; they were .81 when the tests were scored exactly and .83 when scored acceptably. Although the texts were selected from two different sources, one from a Hebrew textbook and the other from an Israeli newspaper, they still correlated very highly. Furthermore, although the mean score obtained from the easy Cloze was higher than the mean score obtained from the difficult Cloze, similar information was obtained, i.e., a student who scored low on one test obtained a low score on the other test—the order of the scores remained the same.

Deletion Rules

Nir (1974) and Nir and Cohen (1977) claim that because Hebrew is a semantically heavily loaded language, deletions smaller than $n=7$ may result in a Cloze test which is too difficult. However, Shohamy found that a deletion of $n=5$ in two Hebrew texts provided appropriate information so that the smaller deletion did not greatly affect either the test discrimination or the difficulty level; students of all levels were able to perform satisfactorily on both tests.

Scoring Methods

In the Shohamy study, the correlations obtained between the Hebrew Cloze test scored “exactly” and the Cloze test scored “acceptably” were .97 for the easy Cloze and .99 for the difficult Cloze. These high correlations indicate that the same results are obtained by the two scoring methods.

Summary

The findings obtained from the results on the Cloze procedures in Hebrew provide evidence that the Hebrew Cloze is a valid and reliable test. In terms of difficulty levels, deletion rules and scoring methods, Cloze procedures in Hebrew are similar to those in other languages.

3. The Application and Uses of the Cloze Procedure for Testing and Teaching Hebrew

Based on the findings reviewed and reported in this paper, several recommendations can be made regarding specific procedures and situations in which the Cloze procedure can be used for testing Hebrew language proficiency.

The Cloze as a Substitute for Existing Proficiency and Placement Tests

The high correlations between the Cloze and other external measures of grammar, reading comprehension and speaking in Hebrew indicate that the Cloze test is a valid test that can be used as a surrogate measure for less practical and more complex tests of other language skills, as well as for assessing overall proficiency. From a statistical point of view, when high correlations exist between two tests, it means that a large amount of variance in one can be accounted for by the other test, thus implying that it may be possible to substitute Cloze tests for the more complex tests with no loss of significant information.

Thus, in situations where there is a need for a highly efficient test, the Cloze procedure offers major advantages as a test which is easy to construct, administer and score. Furthermore, it is easy to create parallel forms—an important factor since tests are so rapidly “consumed.” A parallel form can be created by simply moving the blank one word to the right or to the left, thereby creating indefinite numbers of texts which can be used for the Cloze tests. Substituting the Cloze test for less practical tests is a recommendation that is made for specific situations where a large number of applicants apply to a certain program and there is a need to obtain information on the proficiency of the candidates for screening or placement purposes. When a Cloze test is administered for screening and placement, results can be obtained immediately, inexpensively and with a limited amount of manpower. Preparing, administering and scoring a 50-item Cloze test can speed the process significantly, compared to more lengthy and costly administration of other tests.

Using Cloze procedures can be helpful in situations such as (1) placement of students to Israeli Ulpanim, (2) placement of foreigners in the Israeli Armed Forces, and (3) placement of students who have had previous exposure to Hebrew in Hebrew schools or on visits to Israel. However, it should be noted that, in spite of the high common variance between Cloze and other tests, there is some variance in the tests that is not accounted

for by the Cloze test, and some information is lost by the substitution. Thus, decisions which are based on information obtained from the Cloze tests should be used with caution, especially since information obtained from a Cloze test cannot be used to diagnose candidates' weaknesses and strengths, and information obtained from these tests does not lend itself to qualitative interpretation.

A Classroom Test and Teaching Device

The Cloze procedure in Hebrew can also be used as a classroom test, (1) as a vocabulary test where specific words are deleted and the students have to reinsert them, and (2) as a reading comprehension test to examine how well students comprehend the text. It can also be used as a listening test where the teacher reads a text to students but leaves every 'n'th word blank and the students are required to supply the words. Finally, it can be used as a speaking test where students listen to a text from which certain words have been deleted and are expected to say the missing words.

In terms of specific suggestions, it is recommended that a text be selected from a textbook or other material which is on the same level as the one that students are studying at the time of the test. A deletion of 50 items is sufficient to obtain reliable information for an overall proficiency test.

In terms of the "best" deletion rule to be selected for the Hebrew Cloze, less than $n=5$ may result in a Cloze test which is too difficult, and 5 and above should provide the necessary information. It is recommended, however, that when a Cloze test is used as a proficiency test, a small pilot study should be conducted to obtain information as to the difficulty of the text. A pilot study is helpful in deciding if the text should be revised and how it should be done.

In terms of scoring methods, the "exact" method yields similar information to the "acceptable" one in Hebrew, and is much easier to score. However, from the students' point of view, it is probably more ethical to use the "acceptable" procedure, in spite of the additional work involved. There is something inherently improper in counting an answer incorrect when it is actually correct, just because the author of the original text did not choose that particular word.

The type of Cloze procedure chosen will depend on how the test developer intends to use the test. If it is used to test specific vocabulary or grammar, then such items should be deleted; if it is used as an overall proficiency test, then randomly selected items are preferable.

Conclusion

It has been shown in this study that results obtained from the research on the use of the Cloze procedure in Hebrew seem to follow similar research findings obtained in other languages. The Cloze is a highly efficient, reliable, and valid test which may be used as a substitute for less economical Hebrew tests, without substantial loss of information. It offers considerable advantages in testing Hebrew as a second language and as such, should be more widely utilized and more widely researched.

Further research is needed to validate the Cloze with other tests in Hebrew by using other validation procedures, by experimenting with other scoring procedures, and by examining differences in results by using a variety of texts obtained from difference sources.

It is most important to examine carefully what the Cloze test measures, how it samples the subject matter, and how this relates to the surrounding items, so that we will be able to discover *how* a Cloze test taps other language skills and what type of linguistic and contextual elements it tests.

BIBLIOGRAPHY

- Bormouth, J. R. 1962. *Cloze Tests as Measures of Readability*. Unpublished Ph.D. dissertation, Indiana University.
- Brown, J. D. 1980. "Relative Merits of Four Methods for Scoring Cloze Tests." *Modern Language Journal* 64(3): 311-317.
- Darnell, D. K. 1968. *The Development of an English Language Proficiency Test for Foreign Students Using a Clozentropy Procedure*. (EDRS: ED 024039) Boulder, Colorado.
- Gregory-Panopoulos, J. R. 1966. *An Experimental Application of Cloze Procedure as a Diagnostic Test of Listening Comprehension Among Foreign Students*. Ph.D. dissertation, Southern Illinois.
- Jones, R. L. 1977. "Testing: A Vital Connection." *The Language Connection: From the Classroom to the World*, pp. 237-265. Ed. June A. Phillips. The ACTFL Review of Foreign Language Education Series, Volume 9. Skokie, Illinois.
- Kerr, A. H. 1972. "Psycholinguistics and a New Empirical Approach to the Measurement of Comprehension and Readability of French as a Second Language." *Changing Education*, pp. 278-293. Eds. Simpkins and Miller. Sydney, Australia.
- Klare, G. R., M. W. Sinaiko, and L. M. Stolorow 1971. *The Cloze Procedure: a Convenient Readability Test for Training Materials and Translations*. Arlington, Virginia.

- Leong, S. N. 1972. "Cloze Procedure as a Measuring Device for Reading Comprehension in the Chinese Language." *NRC No. 4*. Ministry of Education, Singapore.
- Mason, V. 1971. *Language Competence as Measured by the Cloze Test*. Unpublished M.A. thesis. University of California, Los Angeles. Workpaper in TESL: UCLA, 6.
- McLeod, J. 1974. *Comparative Assessment of Reading Comprehension: A Five-County Study: Saskatoon, Canada*. Institute of Child Guidance Development.
- Nir, R. 1974. "Haššimus bəšitat ha'Cloze' libədikat šī'ur haqqəri'ut." *Iyyunim bəhinnuk* 4:71-84.
- Nir, R. and A. Cohen 1977. *Pituah mibhaneh miyyun ləlomdey 'ibrit*. Paper published at the Hebrew University Center for Applied Linguistics, Jerusalem. (Paper presented at the International Congress of Jewish Studies, August 1977.)
- 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): 151-157.
- _____. 1973. "Cloze Tests of Second Language Proficiency and What They Measure." *Language Learning* 23(1):105-118.
- _____. 1976. "Evidence for a General Language Proficiency Factor: An Expectancy Grammar." *Die Neueren Sprachen* 25:165-174.
- _____. 1979. *Language Tests at Schools*. London.
- _____. (Forthcoming). "The Factorial Structure of Language Proficiency: Divisible or Not?" To appear in *Language Tests at Schools: A Pragmatic Approach*.
- Oller, J. W., Jr. and C. A. Conrad 1971. "The Cloze Technique and ESL Proficiency." *Language Learning* 21(2):183-195.
- Oller, J. W., Jr. and F. B. Hinofotis 1976. "Two Mutually Exclusive Hypotheses About Second Language Ability: Factor Analytical Studies of a Variety of Language Tests." (Paper presented at the Annual Meeting of the Linguistic Society of America, Philadelphia.)
- Oller, J. W., Jr. and V. Streiff 1975. "Dictation: A Test of Grammar Based Expectations." *Testing Language Proficiency* pp. 71-82. Eds. R. L. Jones and B. Spolsky. Arlington, Virginia.
- Shohamy, E. 1978. *Investigation of the Concurrent Validity of the Oral Interview with Cloze Procedure for Measuring Proficiency in Hebrew as a Second Language*. Unpublished Ph.D. dissertation, University of Minnesota.
- Spolsky, B. 1973. "What Does it Mean to Know a Language? How Do You Get Someone to Perform His Competence?" *Focus on the Learner*, pp. 164-176. Eds. J. W. Oller, Jr. and J. C. Richards. Rowley, Massachusetts.

- Taylor, W. L. 1953. "Cloze Procedure: A New Tool for Measuring Readability." *Journalism Quarterly* 30:414-438.
- Toiemah, R. A. 1978. *The Use of Cloze to Measure the Proficiency of Students of Arabic as a Second Language in Some Universities in the United States*. Unpublished Ph.D. dissertation, University of Minnesota.
- Valette, R. M. 1977. *Modern Language Testing*. Second Edition. New York.