The Cloze Procedure as a Test of ESL Proficiency

Susan Unber

INTRODUCTION

Although out of the mainstream of recent research, the cloze procedure still appears to be a viable method of proficiency testing. The author hopes that the reader, after having thoroughly digested the following, will see the applicability of the cloze procedure.

The cloze test is an attractive test because it offers simplicity of construction, administration and scoring, and at the same time, it has been shown to be a reliable and valid measure of language proficiency for the learner of English as a Second Language (ESL).¹

The original cloze procedure, developed by Wilson Taylor in 1953, was designed to measure the level of readability of a passage of prose. By deleting every nth word (a positive, whole number), an average score of a sizable population could be used as an estimate of the difficulty of the text. In 1956, he further suggested using the cloze to determine second language (L2) proficiency. Since then, the written cloze procedure has been utilized and highly regarded for tests of reading comprehension, global language proficiency and other aspects of language competence for both native and non-native speakers of English. The orally presented cloze has been used as a measure of comprehensibility of materials, as well as for listening comprehension of examinees (Potter, 1968). Klare, Sinaiko and Stolourow (1971) found it valuable for use with languages other than English. There appears to be a universality of cloze test performance across languages (Oller, Bowen, Ton That Dien, and Mason, 1972). Darnell (1968) found it useful with ESL, as have many others since then. Darnell also designed the procedure, “clozentropy,” a test of responses of native speakers used to define and weigh acceptable answers based on the information theory.

As Upshur (1971) declared, a critical issue in the cloze or any test is its use. The cloze has provided valid and reliable results through an economical procedure and is versatile in its use. In this paper, I will define the cloze procedure and discuss its relevance as a tool for measuring ESL global proficiency.

I. BY DEFINITION

Taylor (1953: 416) defines the “cloze unit” as

“any single attempt to reproduce accurately a part deleted from a message (any language product) by deciding from the context that remains, what the missing part should be.”

¹. This applies to English as a Foreign Language (EFL) also.
The “cloze procedure,” he continues (p. 417) is a method of intercepting a message from a transmitter (writer or speaker), mutilating its language patterns by deleting parts and so administering it to receivers (readers or listeners) that their attempts to make the patterns whole again potentially yield a considerable number of cloze units.”

A simple look at the cloze procedure in light of the communication theory can be represented schematically as shown below (adopted from Evans and Hastrup, 1976, p. 35):

```
<table>
<thead>
<tr>
<th>SENDER</th>
<th>CHANNEL</th>
<th>RECEIVER</th>
</tr>
</thead>
<tbody>
<tr>
<td>(writer or speaker)</td>
<td>(reader or listener)</td>
<td></td>
</tr>
<tr>
<td>encoder</td>
<td>mode</td>
<td>decoder</td>
</tr>
<tr>
<td></td>
<td>filters</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MESSAGE</td>
<td></td>
</tr>
</tbody>
</table>
```

Generally speaking, messages undergo distortion through channel filters, such as external noise, incomplete listening, speaker disfluencies, distractions, etc., yet the native speaker can still interpret such an incomplete message because he can make use of other contextual cues. These cues are redundant and can be used when needed to fill in the gaps created by the filtering process. The replacement of the deleted words by the receivers may be analogous to the process of closure in Gestalt psychology, hence the name cloze (Weaver and Kingston, 1963). This process of completing broken patterns is a natural human psychological tendency and ability. The more gaps in the message, the fewer redundant cues available, to the native and non-native speaker alike.

The cloze procedure may be paralleled to this communication and psychological model. The message or cloze passage has gaps systematically designated by the process of every nth word deletion, artificially creating an incomplete message. The receiver must correctly complete the text by filling in the gaps, calling upon his knowledge of the language and the remaining contextual cues. In real communication situations, these cues are not confined to just verbal messages but also include a variety of extralinguistic clues such as gestures, facial expressions, pointing and pictures, etc.

In the case of the cloze, language cues will be the only source of redundancy, but as language itself is unusually redundant, it is this property of language that allows the receiver to predict missing symbols from the context. Aitkens (1977: 65) points out “Highly redundant texts tend to be repetitive and contain relatively little information per symbol.” Therefore, the greater the redundancy of the message, the easier it is to comprehend. Aitkens (1977: 65) cites the example below:

The (A)_____ is going (B)_____ sink in the quicksand.

(A) is a high information symbol. Because it could be almost anything (although it is limited grammatically to the category, noun), it is highly dependent on the context and requires
grammatical and contextual cues. (B) is a low information symbol and can be easily predicted if the structure of the language is known. Redundancy reduces the possibility of errors and aids when interference occurs, but it is only helpful to the extent that it is recognized. Pertaining to the ESL speaker, George (1972) keenly summarized that the language is only redundant to the extent that his concept of the English language, and his own communication strategies, match those of a native speaker. The less experienced speaker requires more redundancy (Spolsky, 1971).

The cloze procedure requires that the test taker predict missing words. It tests the ability to continuously draw from contextual clues to the nature of the content immediately to follow. This is a true measure of reading or listening fluency, the creative aspect of receptive language (Porter, 1976). The ability for the listener to predict his speaker’s next words and for the reader to guess the upcoming words or lines in a written passage is dependent upon the ability to fully comprehend the language being processed at any given moment. This is what Oller (1973: 114) terms “analysis by synthesis” and the reverse, “synthesis by analysis.” These two processes are probably manifestations of the same underlying competence, which Oller, in many of his articles on cloze testing, calls an “expectancy grammar.” This predictive ability is a cognitive function related to expressive functions as well, and it is interesting to note at this point, that tests of reading and aural comprehension, dictation, and reduced redundancy (like cloze) seem to highly correlate with each other, around 80% or better (Oller and Conrad, 1971; Oller, 1972; Irvine, 1974). Many researchers now feel that the productive and receptive processes of language are not distinct, but really representative of the unique underlying skill, the active process of hypothesis sampling and expanding, involving constant revision of one’s expectations, i.e. the expectancy grammar (Oller, 1972, 1973; Irvine, 1974; Aitken, 1977). Oller (1971) has suggested this to be the foundation of all language skills, the ability to anticipate elements in a sequence, i.e. predictability and expectancy. If the language user is able to anticipate forthcoming information, processing will be facilitated. There will be less stress on the short-term memory. Short-term memory constraints are thought to be limitations on underlying language competence, rather than on performance, again a hint at the expectancy grammar (Oller, 1973). The cloze procedure tests these constraints.

More recently, language researchers, through qualitative empirical procedures, have come to better understand the processes involved in the strategies in taking a language test. Cohen (1984), one of the first to report on the self-perceived strategies by the test taker in cloze tests, found the top two to be guessing and using immediate context, otherwise referred to as short term memory constraints. This, again, substantiates earlier findings.

The reverse process, the prediction of the sender’s words by the receiver, is the idea of which Darnell

---

2. Other aspects of learning theory seem applicable to language learning and the cloze procedure. One such proposal by Gagne (Ingram, 1978) suggests four steps in the process:
(a) Perceptual identification—the basic step underlying the following three,
(b) Concept learning—semantic and syntactic categorization,
(c) Chaining—the ability to produce as a smooth sequence, an activity that has several steps.
(d) Problem-Solving—the structuring of a task so that the learner can make decisions about relevant concepts and procedures. All are reminiscent of the idea of predictability, expectancy grammar, and cloze.
(1970) sets forth pertaining to his definition of proficiency, which he calls “decision compatibility”:

“A speaker’s communicative ability is best stated by his choice of words to get his point across to another listener” and “to choose the ones a listener would make if they were to exchange roles.” (p. 38)

Cloze, too, is a measure of this skill.

II. LANGUAGE PROFICIENCY AND LANGUAGE TESTING

Aitken (1977) defined language proficiency as the facility with which an individual can handle the communication needs of a given task in a given language, perhaps a more functional definition than Darnell’s and more useful in practical application as in the testing situation. Proficiency is dependent on the individual’s mastery of meaning at various levels and modes of presentation (oral vs. written, formal style vs. informal, etc.) and his language learning experience. Clark (1978) states that L2 proficiency must encompass all communicative aspects of real life situations such as reading magazines and novels, talking with friends on various topics, in addition to viewing television, speaking on the telephone and communicating in emergency situations. This may be best stated in terms of communicative competence. Moreover, researchers realized that “the whole of the communicative event was considerably greater than the sum of its linguistic elements” (Clark 1983: 432). As language testing entered into the “integrative-sociolinguistic” period, with its emphasis on communication, authenticity, and context, so came the idea of integration. In terms of testing, communicative competence is regarded as so global that it cannot be captured in additive tests of isolated items such as grammar, reading, vocabulary or other discrete points of language. The integrative test was an attempt, made to assess the learner’s ability to use many kinds of information at the same time (Oller 1979: 37). Researchers like Cziko (1982) and Savignon (1982) are of the same opinion, as well, that communicative competence can only be tested through integrative tests. Two tests have been held up as prime examples: the cloze and the dictation (Brown 1994: 262).

The form and content of the proficiency test, Clark continues, must be completely independent of the student’s language learning history. On the other hand, achievement tests are intended to determine which linguistic elements the student has mastered in a class or formal learning situation. Achievement tests, in this sense, are experience-dependent. If an ESL learner has primarily a reading translation background, an oral achievement test would be inappropriate to use. As an oral proficiency test, Clark’s sentiments would deem the cloze a possible and acceptable technique to use. Therefore, the written cloze may be alternated with an oral cloze to measure oral proficiency. Some controversy exists in respect to this aspect of testing. Although oral and written cloze test scores correlate highly with each other and with other tests, as previously mentioned, it seems that such a justification can only be used for people who have both oral and written training. The high correlations between all these tests may be indicative of their integrative nature, but they may still fail to measure true proficiency in certain cases. Perhaps tests should be classified as oral proficiency and written proficiency.
The Cloze Procedure as a Test of ESL Proficiency

It is commonly agreed that the cloze test does measure integrative or pragmatic competence, because it is contextually based and it measures the ability of the expectancy grammar (Oller and Conrad, 1971; Evans and Haastrup, 1976; Aitken, 1977). The problem for the written cloze proficiency test is that it presupposes that a L2 speaker is literate in the L2. Yet he may not only be illiterate in the L2 but in the L1 as well, although his speaking skills may be quite good. In cases like this rather extreme example, variations on the original cloze procedure have been proposed and somewhat verified. These shall be discussed in Section V.

III. RATIONALE FOR THE CLOZE

The cloze procedure is an integrative and pragmatic approach that appears to measure the same skills used by native speakers in sending and receiving messages. It provides semantic, contextual clues and grammatical structure, suitable to use as a global proficiency measure (Irvine, 1974; Jonz, 1976; Clark, 1978; Oller, 1978).

A good proficiency test should be integrative, pragmatic, reliable, valid, practical and efficient.

A valid test measures what it purports to measure, language proficiency in the case of the cloze. Comparatively speaking, to the extent that it correlates with other tests of this nature, the cloze appears to be valid. It does, however, lack the face/content validity of a direct test of proficiency. Direct measurements are, of course, the preferred method of testing, but they are not always practical. Obviously, the best direct method would be to follow a person around for a period of time to observe how effectively he handles himself in various communicative situations, but this is highly unfeasible (Clark, 1978). The cloze is thus better suited for use, but the results of its validity are not conclusive. Weaver and Kingston (1963) claim that cloze results were more related to each other than to factors of other language tests, and that their validity was not justified through correlation. Porter (1978) also found low intra-cloze test correlation contrary to Oller (1973) shadowing doubt on cloze reliability. This suggests to the examiner the use of very discrete interpretation of cloze test scores. Since the amount and quality of language accomplishment represented by a given performance level on a direct test is readily apparent to the examiner, but is not the case with an indirect test, Clark (1978: 28) warns that “appropriate extrapolation of test results to specified types and levels of real life situations” is a crucial matter.

Despite these limitations, Clark (1978) and others still recommend the use of the cloze test as a proficiency measure. Its advantages outweigh its disadvantages. I leave this up to the reader to decide.

IV. SIGNIFICANT VARIABLES OF THE CLOZE

Oller (1973) mentions that certain factors affect a test’s level of difficulty and may affect test performance. The distance between deleted words is important (Oller 1975; Alderson 1979). Having blanks appear less frequently than every 5th word is too difficult, and more frequently than every 12th word is probably too easy. The best range is the frequency between
every 5th deletion and every 12th. Most tests commonly employ deletion of every 5th, 7th or 9th word.

Deletion of function words (rather than content words) is easier for native speakers but the effects on non-native speakers are unclear. Effects due to the type of deletion are reported by Bachman (1982, 1985). In addition, the C-Test (Gratjahn 1996), a variation on the cloze, with its letters as cloze units, may be acceptable as a technique of choice. The multiple word deletion pattern, such as deletions of phrases, clauses, sentences, and paragraphs may make possible cloze items, but is beyond the scope of this paper.

Finally, results of written and oral presentations of the cloze appear to yield similar scores for native speakers, but studies with L2 speakers still need to be done. Chihara et. al. (1977) did an interesting study to measure the effect of discourse constraints on the cloze procedure and concluded that cloze items embedded in normal prose are sensitive to discourse constraints ranging beyond the immediate limits of a single sentence. The effects of these constraints are incremented with increased proficiency. They must be recognized to be used (see George, 1977, and Aitkins, 1977, on redundancy). Oller et. al. (1972) notes an interesting fact about native and non-native speaker errors. L1 speakers, in contrast to L2 speakers, made almost no responses which failed to conform to contextual constraints. Errors of native speakers were not only fewer in number, but different in type.

Another aspect of the cloze test which should be considered is the method of scoring. The best and most convenient way to score is the exact word method. Contextually appropriate responses are likely to be less reliable because the judgment of acceptability is, out of necessity, a subjective one (Stubbs and Tucker, 1974). In response to this, Darnell’s procedure of clozentropy (1968) might be employed. The clozentropy procedure was designed to objectively approve the acceptability of the approximate word, thereby rendering the use of the acceptable word scoring method more feasible. It may even make possible standardization of these tests. Cloze, in any case, remain objective in as far as the test taker’s response is determined entirely by predetermined criteria, so that no judgment is required on the part of the scorer (Bachman 1990: 76).

Oller (1972) said that the scoring method used with native speakers is irrelevant, but this may not necessarily be the case with the non-native speaker. The exact word scoring criterion may create a cloze test that is too difficult for the L2 learner and Oller suggests the use of the acceptable word method. The exact word score may really be a measure of vocabulary rather than language fluency. Anderson (1971), and Stubbs and Tucker (1974) hold the more popular opinion that both methods yield equal results for the ESL speaker. Perhaps this proves correct statistically, but intuitively, for Oller and myself, it still sits uneasy.

---

3. Discourse constraints may be grammatical and contextual and act as redundancy cues.
4. Darnell (1970) makes some very interesting assumptions that apply here:
   (a) The primary assumption of language is communication;
   (b) This function is best served when a group, by compliance within the group, establishes the
       norms of language use;
   (c) A measure of proficiency should index one’s ability to conform to existing group norms
       rather than to a prescriptive model. This seems relative to decision compatibility.
V. MODIFICATIONS ON THE BASIC EGG

Variations on the original cloze procedure have been proposed by many researchers. Jonz (1976) and Porter (1976) suggest the multiple-choice cloze to increase redundancy for those less proficient, decreasing stress and making the test easier to score and more reliable. Propst and Baldauf (1979) offer the alternative of the matching cloze for younger children or those L2 adults with poor reading ability. Other adaptations include the cloze test of English prepositions by Oller and Inal (1971), while still integrative, may be a step closer to the discrete point test. An interesting test by V.F. Allen (1968), not as pragmatically oriented as the cloze, has a slightly different format for deletion and context. Another variation on an old theme is the context test by Bondurak, Child and Tetrault (1972). It is very similar to the cloze in format, but the authors have developed a different rationale and criterion mode for item selection, control methods and approaches to item analysis. The deletions are not as systematically random as in the cloze.

Other possible variations can be constructed from one passage by changing the deletion pattern, the scoring method, the method of presentation (oral, written) and the response procedure. See Appendix I for guidelines on test construction.

VI. CONCLUSION

Cloze tests appear to measure communicative competence in a practical, reliable and economical way. They are an indirect measure of language proficiency with a relatively high validity correlation to other integrative proficiency tests. They also offer general internal consistency, and ease of construction, modification, administration and scoring.

Used with caution, they show

". . . considerable promise as indices of language proficiency in situations where more direct tests cannot be administered." (Clark, 1978, p. 29).

BIBLIOGRAPHY


5. Taken from Jonz (1977), “Improving the Basic Egg: The Multiple-Choice Cloze”


1. The passage should be approximately 375 words.
2. The type of passage varies with the purpose. For general comprehension, readability and proficiency, choose a text as close to criterion language style as possible.
3. Beware of content matter, so as not to be overtaxing.
4. Systematically delete every nth word (commonly 7th) until fifty blanks are made.
5. Some alternatives to deletion process do not permit deletions made for dates, proper nouns, numbers and formulas.
6. Leave the first and last sentences intact.

APPENDIX 2: Guidelines for Test Administration, based on Aitken, 1977.

The examiner should read the directions aloud, describe the test and then make the following suggestions to the students:

1. Write one word per blank.
2. Try to fill-in every blank; guess if you do not know.
3. Don’t ponder over too difficult a blank. Leave it and return to it at the end.
4. Spelling errors are not penalized.
5. Write neatly.
6. Take as much time as you need (within the limits set by the instructions). This should take 1/2 hour.
7. Give a sample for practice.

APPENDIX 3: Cloze Examples—Deletion of every nth word

(Original)

From the quiet narrow streets, the doorway curtains part to admit Koichi. A shout of welcome comes to him from somewhere within the kitchen behind another curtain, further inside. The television way up in the corner tosses out some noise. In a single downward motion he snuffs out his cigarette as he sits down, reaching across the counter to a small bookrack, and pulling out some dog-eared comic book of samurai baseball players and naked women. “Moon noodles,” he calls toward the kitchen, and flips the pages. Behind him customers move in and out of the small place accompanied by all the expected sounds, and so Koichi is free from distraction.

Deletion of every 5th word:


— 103 —
ward as he sits down across the counter to small bookrack, and pulling some dog-eared comic book samurai baseball players and women. “Moon noodles,” he toward the kitchen, and the pages. Behind him move in and out the small place accompanied all the expected sounds, so Koichi is free distraction.

**Deletion of every 7th word:**

From the quiet narrow streets, the curtains part to admit Koichi. A of welcome comes to him from within the kitchen behind another curtain, inside. The television way up in corner tosses out some noise. In single downward motion he sniffs out cigarette as he sits down, reaching the counter to a small bookrack, pulling out some dog-eared comic book samurai baseball players and naked women. “noodles,” he calls toward the kitchen, flips the pages. Behind him customers in and out of the small accompanied by all the expected sounds, so Koichi is free from distraction.