

*Cloze Procedure as an Instrument
in Testing ESL*

CHAPTER = I

CLOZE PROCEDURE AS AN INSTRUMENT IN TESTING ESL

1.1 Discrete point testing

There have been a number of new trends in the overall system of teaching and learning English as a second language for several years. These various types of tests were based on different theories of language. Almost all the tests -- traditional type to the recent communicative competence type have had to rest satisfied with testing parts of language ability, even though the parts tested may be claimed to be representative of the subjects' overall abilities. The parts tested were sometimes basic language skills i.e. listening, reading, speaking and writing: sometimes they were components of these skills such as phonology, graphology, points of grammar and items of vocabulary; sometimes a combination of both composite and subskills was attempted. But there was dissatisfaction with the traditional type of tests, although many experienced teachers have long suspected that no replacement has yet been found for the traditional essay or oral interview despite the attendant marking problems (Wainman 1979).

Researchers introduced the objective type of test items which were discrete-point tests as against the traditional tests. Their attempt to be very specific and accurate in measuring language skills led to the pendulum swinging to the other extreme in the test types proposed by some of them- the structuralists

who believed that language skills could be broken into discrete points and tested. Thus this new type or the objective type of language test tests the use of articles, prepositions, sequence of tenses etc., separately in individual sentences. This could overcome the defects of essay-type questions as there was no possibility of avoiding any specific grammatical point in such tests.

The content of such foreign language tests is typically specified in terms of a two-dimensional frame-work--operations and skills on one axis and the linguistic knowledge on the other.

Language Aspect

Skill	Phonology or Orthography	Morphology	Syntax	Lexicon
Auditory Comprehension				
Oral Production				
Reading				
Writing				

(Reproduced from Carroll, 1961: 30-40)

Thus the operations and the linguistic knowledge are tested separately. But it is defective because discrete-point tests are tests of receptive or passive skills. Many times as

their options are limited and the clues are so overtly provided in the given sentence, the students respond correctly, almost mechanically. Again, language skills are not, by any means passive skills only. They cannot be acquired in isolation. They are used synthetically. They are associative and active. What is to be tested is not dead bits of language cut into parts, but chunks of language at work. This can be done not by testing every small bit separately in isolation from the rest of the language structure, but by viewing the same and presenting parcel of the entire language. Second testing the language requires not only the knowledge of language but the knowledg of context also. Therefore, in this testing model Robert L. Cooper added the fifth column 'Context'. (Cooper, 1967)

Knowledge

Skill	Language Aspect				Context
	Phonology	Morphology	Syntax	Lexicon	
Auditory Comprehension					
Oral Production					
Reading					
Writing					

(Reproduced from Cooper, 1967: 58-60)

Items based on this model test the ability to comprehend and produce spoken and written utterances in connected discourse. Communicative competence is tested in connection with contextual competence and linguistic competence. The two kinds of inference made from such language test scores concerns knowledge of language i.e. competence and the ability to communicate in that language in specific situations i.e. performance.

According to Cooper, this language testing model is adequate for testing competence, but inadequate for testing performance. It is inadequate for two reasons. First, only the knowledge of the target language is not a sufficient condition for effective communication. Hymes makes the distinction between linguistic and communicative competence (Hymes, 1967). By linguistic competence is meant the ability in principle to produce and comprehend any and all grammatical utterances of a given language (Chomsky, 1965). As Hymes points out that effective communication requires more than linguistic competence. To communicate effectively a speaker must know not only how to produce any and all grammatical utterances of a language but also how to use them appropriately. The speaker must know, What to say, with 'whom' and 'when' and 'where' (Cooper 1967). Communicative competence consists of two sets of competencies--linguistic and contextual. So both are useful to test a person's ability to communicate effectively. Neither is a sufficient condition.

Second, the socio-linguistic aspects of the language must be considered while testing the communicative competence "If

codes are clusters of co-variants which shift in response to changes in the social context, then speaker must possess more than a single code. Therefore it is inadequate to test a person's communicative competence by single-code (Neutral language)". (Cooper, 1967)

So Cooper improved the validity of inferences by a socio-linguistic elaboration of his test model.

Table

KNOWLEDGE

	Language Aspect				Context		
	phon. or orth.	morph.	syntax	lex.	A	B	C
SKILL							
Auditory Comprehension							
Oral Production							
Reading							
Writing							

(Reproduced from Cooper, 1967: 62-65)

It is not enough to test the knowledge of the language code (grammar or how to use language) but it is necessary to test whether a person can use the appropriate variety of language in a given social situation.

1.2 Integrative testing: the cloze technique

One of the most popular instruments discovered and used widely in integrative language testing is the cloze procedure. Let us briefly look at the origin and development of the cloze procedure.

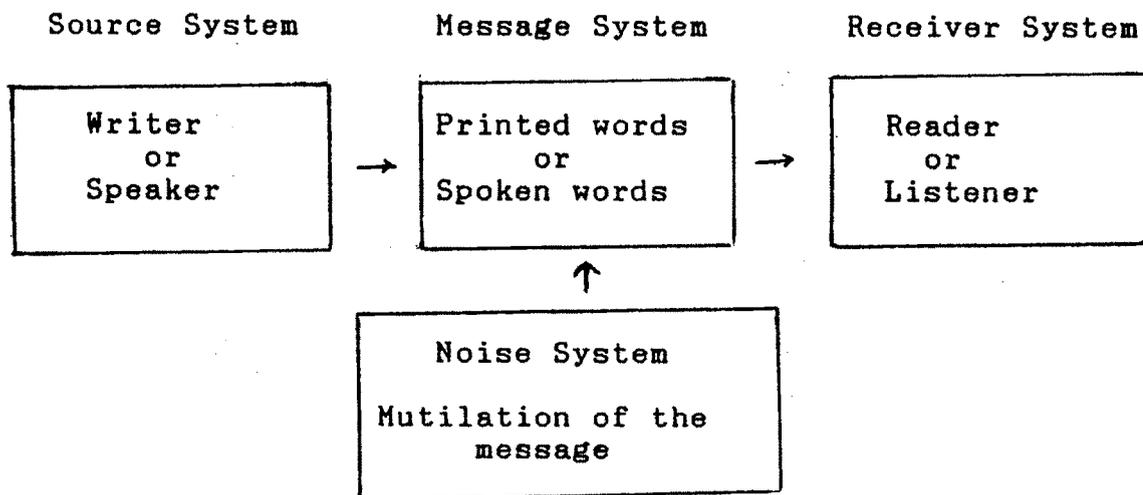
The term 'cloze' was first used by W.L. Taylor (1953) to refer to a type of test originally designed to measure the readability of passages of prose. The procedure was to select a reading passage and delete every nth word in it and replace it with a blank space. The subjects were required to complete the passage by filling in the blanks with suitable words. The average scores over a large group of subjects obtained on this test were used to measure the readability or suitability of the passage for reading at a given level.

The term 'cloze' was used with the notion of Gestalt 'closure' in mind referring to the natural human psychological tendency to fill in the gaps in patterns. In the same way that most people see an incomplete figure without any breaks or discontinuities as shown below. (Anderson, 1971)

Like this, "the restoration of words deleted from a selection of prose in order for the passage to make sense is a special use of this ability to complete the broken patterns" (Oller and Conrad, 1978: 183)

As a result of new interpretations of the rationale behind cloze procedure, it is used to measure not only comprehension but also overall language ability. In support of this, the following model was offered instead of the Gestalt closure in mind theory -- a model for the language correspondence of a Source System to the Receiver System (Anderson, 1971). Anderson considers the ability to reconstruct mutilated text as a reflection of the knowledge of the reader or decoder which should be according to him, identical with that of the writer or encoder. He explains this phenomenon with the help of the following model:

A Model for the Language Correspondence
of a Source System to a Receiver System



In this model, as shown in the figure, "the transmission and receiving of the message are seen essentially as coding operations. The source or encoder, the writer or author produces a message, a passage of printed English. Noise, which in the present context consists of mutilation of the language patterns of the message, interrupts the coded message before it is received by the decoder or reader. To decode the message, cloze procedure requires the reader to construct the mutilated language pattern by making the most likely replacement in the light of his language system and the grammatical and semantic cues that are available". (Anderson, 1971: 179-80)

1.2.1 Reduced redundancy

Another important aspect of cloze procedure is that it exploits the phenomenon of "redundancy" which is a property of all natural languages.

Bernard Spolsky (1969) emphasises the fact that language is redundant and that it is creative. Redundancy is a concept developed as part of the statistical theory of communication (Shannon and Weaver, 1949) "In this theory, a message carries information to the extent that it effects a reduction in uncertainty by eliminating certain possibilities. The greater the reduction, the greater the information". (Spolsky, 1971: 385)

Then Spolsky points out that all natural languages are redundant because in all natural languages more units are used than are theoretically necessary. Redundancy "reduces the possibility of error and permits communication where there is

interference in the communicating channel ... Messages in normal language can be understood even though a good portion of them is omitted or masked. (Spolsky, 1971: 385). Spolsky has given emphasis on the principle of redundancy for testing communicative competence.

Language tests based on reduced redundancy techniques provide good measure of overall proficiency. However, redundancy is only helpful to the extent that the receiver of the message is capable of taking advantage of it. If the receiver does not know the structure or rules of the language, or has no previous experiences to give him a feeling of what is likely to be communicated, redundancy is not very helpful. English to an ESL speaker is only redundant to the extent that his concept of English language and communication strategies matches that of a native speaker. (George, 1972)

There are three main techniques used for the principle of reduced redundancy test. They are cloze, clozetrophy and the noise test. In the cloze test, portions of a written or oral text are blanked out and the subject is asked to provide the missing word or words. The masking is statistically controlled. A cloze test is just like a mutilated message. When a student responds to the items on a cloze test he is required to project a word to fill in a blank and complete a sequence on the basis of an incomplete message. He formulates hypotheses or expectations about the information that is to follow by sampling subsequent sequences, the student confirms or disconfirms these hypotheses. If they are disconfirmed, he rejects or revises his expectations

forming a new hypothesis. (Oller, 1973)

The clozetrophy procedure measures a subject's performance in terms of a group norm. Thus, foreign students are scored according to the extent to which their responses agree with the normal responses of native speakers. The possibilities of the procedure are obvious. The tester can choose a sample population (say native speakers of English in the first year of a university engineering course) and claim to show how well a subject (say a foreign student applying for admission) would fit into this group. This more precise functional definition is likely to be of great value.

In its original form, the 'noise-test' consists of fifty discrete English sentences recorded on tape with accompanying noise in the background. Except for the presence of the noise, the instrument is in essence a dictation test. It is simple to administer, relatively easy to score, and statistically reliable.

Thus the cloze procedure which is based on the principle of reduced redundancy is a very good test of language proficiency.

1.2.2 Redundancy and creativity

Creativity is the basic distinction between language like behaviour and knowing a language. While precise specification may not be possible, for there is a continuum, the interpretation of each is relatively clear. Thus language like behaviour refers to the parrot trained to speak, and equally well to the student who is able to recite a number of sentences in a second language, but

not to modify them and use them in a free conversational situation. The example of the student learning a second language makes the continuum clear, for there is a stage at which he may be able to use his stock of sentences to answer a limited set of questions. This is still not the same as knowing a language which involves the ability to produce an indefinite number of sentences in response to an indefinite number of stimuli.

As we have seen earlier, redundancy is a concept developed as a part of the statistical theory of communication (Shannon and Weaver, 1949). Redundancy may seem wasteful effort, but it is in fact of great use, for it reduces the possibility of error and permits communication where there is interference in the communicating channel. When one considers all the interference that occurs when natural language is used for communication, it is clear that only a redundant system would work. Messages in normal language can be understood even though a good portion of them is omitted or masked; or, in other words, every message contains many elements (defined statistically rather than linguistically) that can be omitted without leading to a breakdown in communication.

The relationship between creativity and redundancy has been clearly established. Knowledge of rules (the key factor in creativity) is also the principal factor in the understanding of messages with reduced redundancy. Miller and Isard (1963) have shown that intelligibility of a sentence depends on its syntactic and semantic rules.



According to Oller (1975); cloze can most appropriately be described as a learner-centred teaching and testing device in second language situations since it is thought to challenge the efficiency of the developing L2 grammar of the student in a way that reflects natural language processes rather than contrived pedagogy. As a learner-centred language test, cloze is superbly suited to counterbalance the discrete-item test.

1.2.3 Construction and scoring of a cloze test

The construction of a cloze test involves two important operations: One, the selection of a suitable passage and two, deciding on the figure 'n' in deleting every nth word.

First, select a self-contained passage of approximately 375 words. The type of passage selected will depend on the purpose of the test. A test of specific reading comprehension would require a passage previously read. On the other hand, a test of general comprehension, a readability test or a proficiency test would necessitate choosing a passage as close to the criterion language style as possible for the difficulty level of the passage. There are test has ready made procedures such as by Dale and Chall (1948) or the grade placement formulas such as by Bormuth (1969). In the circumstances the teacher has to depend either on his own judgement based on empirical knowledge or devise some ad hoc procedures. It may be useful to state that such procedures will involve determining the linguistic difficulty in terms of vocabulary level (with refernce to appropriate word lists) and sentence level in terms of length and complexity of sentence; content difficulty in terms of the

teacher's judgement (more than one teacher's judgement) and registeral and stylistic features in some way.

So the construction of a cloze test involves the selection of a suitable prose passage and systematically deleting every 'n'th word, replacing it with a blank of standard size leaving out the first and the last sentence intact.

It is important not to choose the items to be deleted, for that would vitiate the rationale of the cloze procedure. Although, generally ten spaces, there have been suggestions that it should be the same as the deleted word (Anderson, 1971). It may be argued that the length of the deleted word is a "clue" which may be treated as one more redundant symbol. (Shastri, 1980)

1.2.3.1 Constraints on the Choice of Items to fill Blanks:

There are two constraints which work upon the mind of a reader while he is attempting to produce the right word and to put it in the right grammatical form to fill in the blank. They are:

- 1) The semantic constraint
- 2) The syntactic constraint

The semantic constraint works on the principle of how well we are acquainted with the discourse and the discourse level. It tells something about the maturity and cultural development of the respondent. If it matches well with discourse level of the passage and he/she has enough control of the vocabulary, he/she produces the right word. The deviation from the correct response

in direct proportion to the difference between the discourse level of the passage and the mental and cultural level and the vocabulary of the respondent. Thus the production of the correct vocabulary item is a very objective production test of the vocabulary of the subject, the several content words competing to be fitted in the blank being in the paradigmatic relationship with each other.

The syntactic constraint works on the principle of how we have mastered the rules of grammar and the conventions of collocation of the target language. The several words competing to be fitted in the blank under this constraint (so far as their grammatical forms are concerned) are in syntagmatic relationship with each other. (ELTC, 1988)

1.2.3.2 Scoring procedure

The cloze procedure has many merits which have been demonstrated by the studies of Darnell (1968), Bowen (1969), Kaplan and Jones (1970). Oller and Conrad (1971) and Ollar and Inal (1971). But there are several questions which are yet unanswered. One of them is the problem of scoring system of cloze procedure.

Cloze is a very objective test of linguistic skills because the word to be filled in the blank is the same as the one deleted. There is no variation or subjective judgement.

There are two scoring procedures for cloze tests:

1. By accepting only the exact word in the same grammatical form



2. By accepting any contextually acceptable word in the correct grammatical form.

Anderson (1971) found that the two methods yielded the same results and recommended the former method because it was easier. Stuffs and Tucker (1974) found that the results of the two scoring procedures correlated very highly with each other ($r = 0.97$), Oller (1972), however, argues that any acceptable alternative method is better for use with ESL students. Anderson (1972: 66-67) tried any contextually acceptable answer method which was the following:

1. Verbatim Score (V): Each exact replacement of deleted word was given one mark.
2. Synonym Score (SYN): Each exact replacement of a deleted word was given one mark and each synonym half a mark. To receive credit, the synonym had to be listed under the appropriate usage of the word in "Collin's Gem Dictionary of Synonyms and Antonyms". (1964)
3. Alternative Response Score (AR): Each exact replacement of a deleted word was given one mark and each response that made sense within the context half a mark. Such responses had to be grammatically correct in terms of number agreement and had to fit the original syntactic pattern of language structure.
4. Grammatical _____ Score (GC): Each exact replacement of a deleted word was given one mark and each response of the same grammatical class as the deleted word, regardless of number or tense, was given half a mark.

Various other parameters were used in the scoring of responses of ESL students (Bowen, 1969; Oller, 1972; Oller et al 1972). But the most elaborate and objective one seems to be that of Clarke and Burdell (1977). They propose three criteria in evaluating responses, namely,

1. Syntactic Acceptability (SYNAC)
2. Semantic Acceptability (SEMAC)
3. Semantic Change (SEMCH)

1. The exact word, as it is fully acceptable on application of both the criteria, is to be awarded one full mark. A student will not be punished for spelling mistake.

2. Any other word which is syntactically. (i.e. grammatically) acceptable and semantically acceptable at the discourse level (i.e. passage level) is also to be awarded one full mark.

3. If the word used by the student is semantically acceptable only at the sentence level (i.e. if the word is meaningful in relation to both preceding and succeeding parts of the blank in the sentence, but is not acceptable at the passage as a whole level), and is in the correct grammatical form, it should be given half a mark.

4. A word not acceptable semantically either at the passage level or at the sentence level and a word in wrong grammatical form should not be given any credit.

For a test on Reading Comprehension some concession for a slight grammatical mistake may be given. (ELTC, 1988)

Within each criterion a range of well defined degrees of acceptability is suggested "It provides a frame work within which cloze test responses can be objectively evaluated. What is more, such a system permits researchers to determine their own criteria for acceptability depending on the specific objectives of the test". (Shastri, 1980)

1.2.4 Determining the figure 'n':

In the construction of a cloze Test, every 'n'th word is deleted. This 'n' has a special value. Because it is natural to expect that the blanks offer a fairly equal exercise on content words and structure words. The 'n' value should be so chosen as to satisfy this condition.

If it is found that the 'n'th blank falls on some proper noun or a number etc. for which there are no clues available for guessing, the 'n' value can be shifted to the next number for that blank only, keeping the rest of the pattern uniform. This is called 'modified cloze'.

If on leaving the first sentence intact and deleting every 'n'th word from the beginning of the second sentence, it is found that the majority of the blanks fall on difficult content words, we may leave the second sentence too intact, and see whether the words on which the blanks fall come off all right. This is how utmost care must be taken before selecting the passage and deleting the 'n'th word so that proper difficulty level is maintained and a fairly equal chance is given to the structure words and the content words to be deleted.

It is often believed that a 'less than every fifth word' or 'more than every tenth word' deletion pattern is either unmanageable to take or impracticable to construct (ELTC, 1988). However, Mac Ginitie (1961) was perhaps one of the first to find that deletion rates fluctuating from every fifth word to every 12th word do not substantially alter the subjects' performance in cloze test. (Paul, 1985).|

Oller (1972) as reported by Davies (1968) also found that deletion rates between every 5th and every 12th word keep results stable. Nihalani (1979) found that deletion rates between 2nd to 12th with ESL students in India did not affect the results substantially. Alderson (1980) notes that the effect of varying N on the results are themselves variant. They seem to vary with the difficulty level of the cloze passage. However, his results on 'medium level' passage show no effect of varying N on the results. It is intriguing to note that on an 'easy' passage, the results are contrary to expectation a lower (6th) rate deletion gave higher scores than a higher (12th) rate deletion. (Sawant, 1989)

1.2.5 Versatility of cloze procedure:

The cloze procedure is very flexible and allows a wide range of deletion patterns and testing formats. The test designer can follow a random order of deletion and suppress every fifth, sixth or seventh word, or he may wish to follow a rational order of deletion and suppress a specific class of words like articles, adjectives, or verbs. The test designer can adopt a variety of testing formats. He may use the basic open-ended form in which he presents the passage to the student and informs him to generate

the necessary words. He can modify the open-ended form by attaching, a list of the missing words to the end of the passage. He can go even further and list the options for each blank directly underneath. (Fischer, 1981) |

The words deleted will naturally vary in form and function and may include members of all the word classes in the language - nouns, conjunctions and sentence-connectors as well as the verb forms, articles and prepositions found in the usual completion type exercise. Appropriate tenses, interrogatives, and all the other items usually tested by separate exercises can all be covered in one cloze passage. Precisely because of the complexity of the passage chosen, both production and recognition skills are brought into play. Testing experts in U.S.A. argue that a cloze test has a very high correlation with other measurements of language skills and in fact reflects all round language proficiency. |

Thus a cloze passage is far more than a complex completion exercise; it is an aspect of controlled composition (oral and written) and demands of a learner a more creative approach to language learning and language use as well as an involvement with the passage as a whole-- since the missing word at one slot may be anything from an indefinite article to an abstract noun, while the choice of filler at any one slot influences and is influenced by choices at later or earlier slots. Hence, language has to be seen as an integrated whole, going beyond the sentence to a kind of well-knit discourse. (Gefen, 1979) |

1.2.6 Conclusion

We have seen how the cloze procedure has proved to be a versatile instrument for testing the overall language ability. We have also seen that it has been used for testing the language proficiency of ESL learners. Researchers have hinted at the utility of the cloze procedure even in teaching ESL.

The rationale on which the cloze procedure was designed has also been elucidated in the foregoing sections. The main properties of natural language, that is, creativity and redundancy are exploited in cloze procedure.

This has suggested to us a possible new utility of the cloze procedure. The object of our investigation is to assess the naturalness of language in fabricated ELT materials. We believe that the cloze procedure may reveal some of the peculiar features of fabricated texts if the text is 'mutilated' and submitted to the native speaker for completion.