CHAPTER I
CLOZE PROCEDURE

Chapter I

CLOZE PROCEDURE

I

Discrete point Vs. Overall Language Ability

Objective assessments of proficiency in language have always attracted the attention of testing experts, 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 have been engaged in finding ways of making up for the inadequacies of the traditional testing system. 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 down into discrete-points and tested. Thus the use of articles, prepositions, sequence of tenses etc. came to be tested 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.

This discrete-point testing device did not however guarantee the acquisition of a specific skill on the part of a student. It can be seen from the public examination papers of

Secondary and Higher Secondary Certificate examinations that the students who could score good marks in discrete-point test items on grammar, do not necessarily score well in the other questions. It can also be seen that the students who answer the questions on grammar (discrete-point items) pretty correctly, are generally not able to write any free composition of their own. It is often a complaint from industries and administrative offices, that young employees, good as they appear to be in English from their scores in their examinations, are unable to write or speak well enough.

This is so because discrete-point tests, at their best, 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. Language skills are not, by any means, passive skills only. Nor can they be acquired in isolation one from the other; but they are to be put to use synthetically. They are associative and active. What is to be tested is not dead bits of language cut up into parts, but whole chunks of language at work. This can be done not by testing every small bit seperately in isolation from the rest of the language structure, but by viewing the same and presenting the same as part and parcel of the entire language. And this is what is attempted to be done, most objectively, by employing the cloze procedure (ELTC 1988).

1. Background and Uses of Cloze Procedure :

To understand the various uses of Cloze Procedure, and the rationale behind it, it is necessary to know what exactly this test is (ELTC 1988).

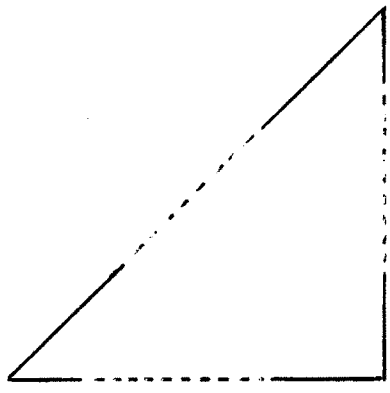
"The word 'cloze' was invented by W.L. Taylor (1953) to refer to a type of test originally designed to measure the readability of passages of prose. The test was constructed by deleting words from a selection and requiring the examinee to fill in the blanks. The average score for a large group of subjects was taken as an index of the level of difficulty of the passage. The term 'cloze' was used with the notion of Gestalt 'closure' in mind, referring to the natural human psychological tendency to fill in gaps in patterns. 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 broken patterns." (Oller et. al. 1971).

The following is an example of a cloze test.

"The idea that the ..... of a language, unlike ..... words, are probably infinite ..... numbers, so that they ..... be listed, is no ..... one, however, familiar it ..... recently have become through ..... writings of Chomsky and ..... followers."

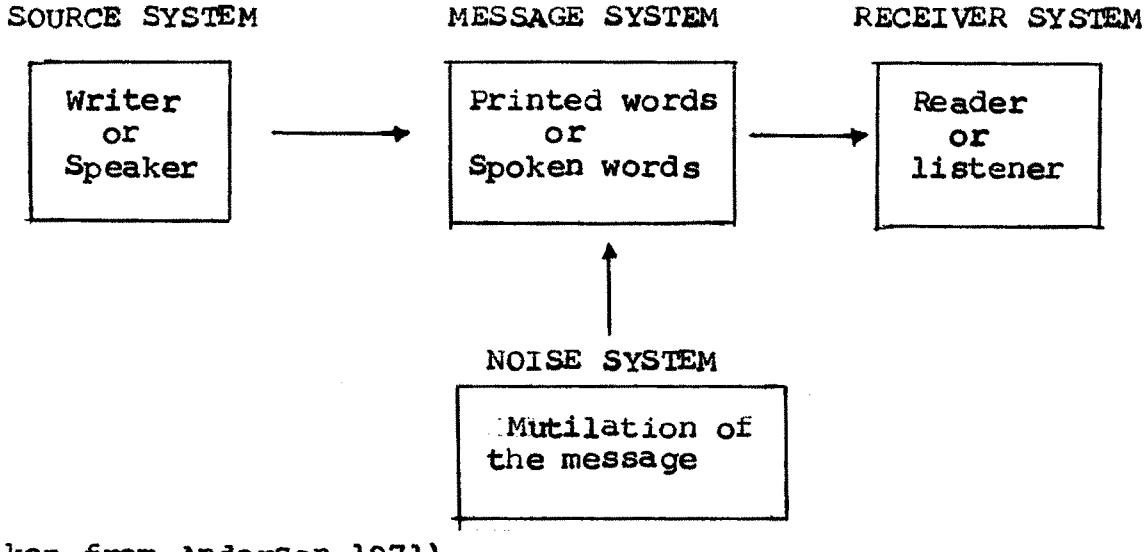
In this passage every fifth word has been deleted.

An often-quoted rationale for cloze procedure is that there is a tendency to close broken language-patterns in the same way that most people see an incomplete circle or triangle without any breaks or discontinuation as shown below (Anderson 1971).



Following the same rationale, breaks in language communication system, are created to make up a cloze test item. The following model of communication system illustrates this point.

<u>A Model for the Language Correspondence of a Source System to a Receiver System</u>



(taken from Anderson 1971)

According to the language communication model in the figure the transmission and receiving of a 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-patterns by making the most likely replacement in the light of his language system and the grammatical and semantic cues that are available (Anderson 1971).

2. Reading Comprehension and Language Proficiency :

If reading comprehension is defined as the correspondence between the semantic and grammatical habit-systems of encoders and decoders using the same language, then it can be seen that cloze procedure taps this correspondence and thus indexes reading comprehension. Reading comprehension is therefore defined operationally.

There is support from linguistics for cloze procedure as a measure of reading comprehension. Fries (1963) identified three layers of language meanings : the layer of meanings carried by the grammatical structures, the layer of meanings carried by the lexical items, and the layer of social-cultural meanings. Cloze procedure taps all these three layers of language meanings,

for, as MacLeod (1965) stated, to successfully reconstruct a message "requires a familiarity with the grammatical structure of English, an understanding of lexical meaning and if the passages selected are concerned with a variety of experiences familiar in a given culture, they reflect to some extent 'social cultural' meaning."(MacLeod, 1965, p.3). There is some theoratical basis, then, for believing that cloze scores index reading comprehension (Anderson 1971).

3. Overall Language Ability :

Spolsky et. al. (1968) and Oller (1973) regard the cloze procedure as an efficient and reliable test of overall language proficiency. It is also particularly well suited for achievement tests of linguistic competence because of the nature of the task it presents to the students and its high degree of versatility.

Oller (1973) argues that cloze measures what he calls a "grammar of expectancy", which, he further claims, directly underlies linguistic competence. According to Oller the student samples the information available to him in the passage and formulates a hypothesis or expectation of what should come next in the sequence. His expectation leads him to project a logical word to fill in the blank . Although some may prefer to wait for empirical confirmation of Oller's "grammar of expectancy", it is clear that the cloze procedure does pose a direct challenge to the the student's linguistic competence. It requires the student

to scan the information present in a sentence, analyze it for its salient features and probable meaning, and generate a word that is consistent with his analysis. After generation of the word he judges its grammatical acceptability and appropriateness within the sentence and the surrounding context. His ability to recall the word and to evaluate its syntactic and semantic pertinence is directly dependent on his linguistic competence (Fischer 1981).

4. 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 (where the examinee knows what to look for and so is placed in a mechanical and unnatural situations) 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 the 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 or 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 discourse analysis (Gefen 1979).

5. Redundancy and Language Comprehension :

One of the fundamental truths about natural language is that it is redundant. Redundancy is that property of language

that allows us to predict missing symbols from the context (Carroll 1964). Highly redundant texts tend to be repititive and contain relatively little information per symbol. By "information" is meant "the informativeness of the symbols" (Shannon and Weaver 1949).

The greater the redundancy in a message, the greater the chance that the message will be understood. That the greater the redundancy one perceives in a message, the greater the chance one will understand it, logically follows. 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.

Redundancy is very helpful as it reduces the possibility of errors and misunderstanding and permits communication where there is interference in the communication channel. Messages can be comprehended even though a portion of them are omitted or distorted. However, if we give these mutilated messages to someone who is not a native speaker or one who does not know the language well, he will experience considerable difficulty in understanding the message. He requires the full normal redundancy

(Spolsky 1971). A cloze test is just such a mutilated message. The correspondence between the ESL student's conception of English redundancy and the actual rules of English redundancy becomes an index of his overall language proficiency in English.

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 expectation, forming a new hypothesis (Oller 1973). Oller notes that this process is the "mirror image" of what takes place in productive language use (speaking or writing). When a person is trying to convey an idea in speech, he has certain expectations about what he wants to say. He samples his own output and tests and modifies it as he proceeds in order to meet his own expectations. Many of our ungrammatical utterances are the result of such midstream readjustments in our speech (Aitken 1977).

6. <u>The Relationship between Creativity and Redundancy</u> :

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.

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. Redundancy may seem wasteful of 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 proportion 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 the intelligibility of a sentence depends on its following syntactic and semantic rules (Spolsky 1971).

Cloze can most appropriately be described as a learner-centered teaching and testing device in second-language situations since it is thought to challenge the efficiency of the developing L_2 grammar of the student in a way that reflects natural language processes (Oller 1974), rather than contrived pedagogy. As a learner-centered language test, cloze is superbly suited to counterbalance the discrete item test. The latter is generally the distilled product of an evolutionary process which clearly emphasizes administrative, pedagogical and procedural considerations. The former is a contextualized challenge to learner grammar efficiency, displaying an enviable simplicity which Oller (1973) calls nothing short of a stroke of new genius (Jonz 1976).

II

Construction 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 N th word.

To take up the first point, the selection of a passage depends on the object of the test and on the level of the objects to be tested. "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." As for the difficulty level of the passage, the West has readymade procedures such as the application of readability formula devised 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 a procedure will involve determining the linguistic difficulty in terms of vocabulary level (with reference to appropriate word lists) and sentence level in terms of length and complexity of sentences, content difficulty in terms of the teacher's judgement (more than one teacher's judgement) and registeral and stylistic features in some way.

So then the construction of a cloze test involves the selection of a suitable prose passage of about 400 words long 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, now the size of the blank is kept to a standard length, 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. Constraints on the Choice of Items to fill Blanks :

There are two constraints which work upon the mind of a reader and drive him to produce the right word and to put it in the right grammatical form to fill in the blank. They are : (i) the Semantic Constraint, and (ii) 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 us something about the mental maturity and cultural development of the respondent. If it matches well with the 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 varies 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 well 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).

2. Scoring Procedure :

The possibility of utilizing the cloze procedure as a measure of ESL proficiency has recently aroused considerable interest. Studies by Darnell (1968), Bowen (1969), Kaplan and Jones (1970), Oller and Conrad (1971), and Oral and Inal (1971) have demonstrated that the cloze method has merit, but several important questions are yet unanswered (Oller 1972). Among them, one is the problem of scoring system for the cloze procedure.

'Cloze' is a very objective test of linguistic skill because the word to be filled in the blank is the same as the one deleted. There is no scope for variation or subjective judgement. There are two scoring procedures for cloze tests :

a) To accept only the exact word in the same grammatical form,

b) To accept any contextually acceptable word in the correct grammatical form.

For deciding the acceptability of the contextual alternative of the exact word, a standard dictionary of synonyms can be made use of (e.g. Collin's Gem Dictionary of Synonyms and Antonyms). Yet in practice, the scorer's own judgement to decide the degree of acceptability can be relied upon. The question of acceptable alternative remains more pertinent to the content words, there being very little scope for variation in the case of structure words (ELTC 1988).

Anderson (1971) found that the two methods yielded the same results and recommended the former method because it was easier. Stubbs and Tucker (1974) suggest that the exact word replacement method is as valid as the other method. They found that with a group of 206 University students in ESL Cloze tests scored both ways correlated very highly with each other (r = .97). Oller (1972) however, argues that ~~the~~ any acceptable answer method is better for use with ESL students (Aitken 1977).

Anderson's any-contextually-acceptable answer used the following scale for scoring :

a) Verbatim Score (V) : Each exact replacement of a deleted word was given one mark.

b) Synonym Score (SYN) : Each replacement of a blank with a synonym of the deleted word was given half a mark.

c) Alternative Response Score (AR) : 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.

d) Grammatical class Score (GC) : Each response of the same grammatical class as the deleted word, regardless of number or tense was given half a mark (Anderson 1972).

Various other parameters have been used in the scoring of the responses of ESL students (Bower 1969, Oller 1972, Oller et.al. 1972). But the most sophisticated one proposed so far is that by Clarke and Bordell (1977). They adopted three criteria in evaluating a response viz. syntactic acceptability (SYNAC), Semantic acceptability (SEMAC) and semantic change (SEMCH). 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)

III

Determining the 'N' :

In the construction of a Cloze Test, every 'n'th word is deleted. What should be the value of 'n' ? 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, MacGinitie (1961) was perhaps one of the first to find that deletion rates fluctuating from every 5th word to every 12th word do not substantially alter the subject's performance in cloze research (Paul 1985).

Oller (1972) as reported by Davies (1978) also found that deletion rates between every 5th and every 12th word keep results stable. Nihalani (1979) found that deletion rates between 2nd and 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 that a lower (6th) rate deletion gave higher scores than a higher (12th) rate deletion.

With these somewhat contradictory empirical findings on the effect of varying N on student's performance , it was thought worthwhile to make one more attempt by way of replicating such an experiment.