CHAPTER III

THE ELUSIVE N IN CLOZE PROCEDURE

Chapter III

The Elusive N in Cloze Procedure

The object of this chapter is to review the literature on the factors that determine the difficulty level of cloze tests. It had been suggested that deletion rates are a possible factor that determines the difficulty level. This led to a lot of research on the effect of varying N on performance (Oller 1972; Nihalani 1979; Alderson 1980). We have also seen that empirical findings on this problem discount such a possibility. However, owing to some intriguing and counterintuitive results produced in certain experiments (Nihalani 1979 and ours), it was suggested in the previous chapter that it may be useful to take a different line of enquiry.

It was also suggested that factors other than deletion rates may be responsible for making different cloze tests based on the same passage more or less difficult. The other factors suspected are one, the exact items on which blanks happen to fall rather than the rate of deletion and two, the amount of context required for a blank to be filled in. In what follows, we shall summarize the findings of investigations on these two points.

I

The Relation between Blank-Position and Difficulty Level

1) Porter (1978) investigated the effect of shifting the blank-position in preparing cloze tests, based on the same passage, keeping the N constant. The object of his investigation was to find out whether two different cloze tests with the same rate of deletion but different blank positions are equivalent. The following is a summary of his report.

According to him, so far it was implicitly accepted that a number of different cloze tests can be constructed on the basis of a single passage, and all will be equivalent. One aim of the experiment was therefore to determine whether two cloze tests constructed from a single passage, but beginning at two different and arbitrarily selected points, could be regarded as equivalent. A second aim of this experiment was to investigate whether student's performance was consistent - that is, whether they ranked similarly - on cloze tests constructed from passages in different styles. Finally, he refers to the claim that, (Oller 1972) whereas with native speakers the simplest and most reliable system of scoring is to count as correct only the exact words which have been deleted, significantly superior differentiation of ESL speakers

is obtained when any contextually acceptable word is counted as correct (Oller 1972). The third aim of his experiment was to check this claim with students for whom English was a foreign, not a second language.

Two cloze tests (1 and 2) were constructed; the texts were identical except that in 2 each deletion occurred one word earlier than in 1. Each test consisted of two subtests of approximately equal length, one literary ('A' - about 440 words) and one non-literary ('B' - about 460 words). Fifty blanks were derived in each subtest by deleting every eighth word, beginning with the 27th word in subtest A, the 28th in subtest B in test 1, and thus the 26th and 27th word respectively in Test 2. In this way, the first few lines were left intact.

The test was administered to 39 students studying English at the Jagiellonian University of Cracow, Poland, and these were drawn from the 3rd, 4th and 5th years. A full class-period of 45 minutes was allowed for the test. All tests were scored twice, first by exact word method, then by contextually accepted word method. Thus, a total of six distinct scores was obtained for each student-text; an exact word and an acceptable word score for subtest A, subtest B, and for the test as a whole.

To determine whether cloze tests constructed from the same original passage could be considered equivalent, the results obtained from test 1 and 2 were compared by calculating the Pearson product moment coefficient of correlation.

Pearson product moment coefficient
of correlation between tests 1 and 2

		Coefficient
Exact-word scores :	Subtest A	.39
	Subtest B	.47
	Total Test	.57
Acceptable-word scores :	Subtest A	.57
	Subtest B	.58
	Total Test	.65

The first and the most important issue considered was whether cloze tests as usually constructed, that is, beginning at an arbitrary point in a text, can be considered reliable indicators of students' general proficiency in a language as revealed in scores. The data from the experiment undertaken here suggest that they can not; tests constructed from a single passage, but beginning at different points, may yield markedly different sets of scores.

The second question considered was whether cloze-test results are independent of style. No strong correlation was obtained between rankings on the literary and non-literary subtests, again indicating that the two subtests could not be considered equivalent. However, whether this was because they differed stylistically, or because any two cloze tests - even if derived from a single passage - cannot be considered automatically equivalent, cannot be stated with certainty on the basis of this experiment.

Finally, the assertion that acceptable-word scoring gives better differentiation with ESL students was checked with students for whom English was a foreign language. The assertion was found to be generally true of discrimination between individual subjects, but not true of discrimination between year-groups in the population studied.

In conclusion, this study presents evidence which seems to weaken the attractiveness of cloze procedure as a weapon in the teacher's armoury, and which suggests that construction of cloze tests should be left to the professional tester.

It is clear that the main thrust of the researcher is on the utility of cloze tests in general. However, his findings indicate an important lacuna in the experimental designs adopted by investigators in their work on the choice of N or deletion

rate and its effect on performance. It would appear that if equivalence on the basis of deletion rate is questionable, the basis of our research design is faulty.

2) Gorden and Haastrup (1976), in their 'Experiments with Cloze Procedure', were interested in determining the Item Descrimination Indices (I.D.Is) of cloze test items with a view to improving the instrument. In the process, although they kept the N (7th word deletion) constant, they shifted certain blanks in their later drafts. This means that they shifted the blank position keeping the N constant. This resulted in increasing (which was their objective) the I.D.Is and consequently the scores changed. The following is a summary of their investigation.

In 1971 they started experimenting with five modes for testing writing skills. One of these was the cloze procedure. In the first draft they deleted every 7th word. They tried out the test on native speakers and on a sample of their test population. They analysed item difficulty level and worked out item discrimination indices (I.D.Is) using the short method described by Harris (5, Page 106). They found that some items were very easy and some very difficult. Some had a low I.D.I. These items were discarded; that is to say, they wrote in the word. Then they deleted another word, which meant that later drafts had an irregular number of words between blanks. They

repeated this procedure until the test consisted solely of items with consistently high I.D.Is and at a level of difficulty appropriate for their test population.

Results

	Number of Items	Average I.D.I.
An early draft	47	0.34
A middle draft	20	0.43
The final draft	17	0.53

This experiment was encouraging. They had found a way of improving the performance of items in a cloze test.

Osgood and Sebeok say : 'In the process of encoding (or decoding), transitional dependencies among semantic events, among grammatical and syntactical regularities, and also.... within skill sequences are simultaneously contributing to a rather precise selection among hierarchies of alternatives at each choice point.' They continue (page 112) : 'The significant advantage of Taylor's "cloze" procedure is that it taps simultaneously all of the complex determinants affecting word choice, both at various levels of organisation and through long stretches of sequencing.' A testee appears to use all of the contextual cues available to him. He considers semantic and

grammatical clues, he does see items as interdependent over 'long stretches of sequencing'.

Since their objective was to assess cloze tests as instruments for testing comprehension, they were putting out a procedure for standardizing such test.

But our interest is slightly different - the effect of shifting blank-position while keeping the deletion rate constant. The results of this investigation shows that, shifting blank position affects performance. This is another evidence which seems to question our research design based on assumption of equivalence.

II

Sensitivity of Cloze Items to Immediate and Remote Contextual Clues

The choice of deletion rate is a mechanical determinant of the amount of context around the blank in terms of number of words. Empirical findings on the effect of varying N on performance generally agree in one respect and that is deletion rates from 5 to 12 keep results constant. One conclusion that can be drawn from this is that plus or minus five words from the blank is sufficient context for providing clues necessary to guess the cloze items. However, a closer examination of this

proposition reveals that clues may be located far away from the blank - sometimes many sentences before or after it. Also, the order of sentences which provide discourse clues is another factor. These together with several other intriguing and counterintuitive findings on this problem point towards the need for looking at the sensitivity of cloze items to immediate and remote contextual clues. Some work has already been done on this problem. In what follows we shall summarize some of these works.

According to Chihara et. al. (1977), an important practical question is how to measure discourse constraints (or alternatively, the competence to utilize them). The cloze procedure is one of the techniques that has been proposed as a possible basis for investigating discourse constraints (Carroll 1972). However, it has been claimed that "cloze scores are dependent chiefly on what may be called the 'local redundancy' of a passage, that is, the extent to which linguistic cues in the immediate environment (generally in the same sentence) of a missing word tend to supply it.....cloze scores are probably more dependent on detection of grammatical than of semantic cues." (Carroll 1972:189). Pushing this argument to the extreme they hypothesized that cloze procedure is not sensitive to discourse constraints.

They say that although it has been demonstrated Oller (1975) that greater amounts of surrounding context increases the ease with which subjects supply missing words, it has not been shown that cloze procedure is sensitive to the kinds of constraints that range across sentence boundaries.

Two passages of prose were selected from texts written for non-native speakers of English. Two cloze tests were constructed over each passage. Every 7th word was deleted from A and every 6th word from B to get two sequential cloze tests (A Seg and B Seg). No lead-in or lead-out was allowed for either test. Then the tests were divided roughly into thirds. Divisions were always at sentence boundaries. The last sentence of the last third of A Seg then became the first sentence of the scrambled version of A (A Scr). The last sentence of the second third of A seq became the second stencence of A Scr and so on until all sentences were exhausted. Then, the same procedure was applied to B seq to get B Scr. There were 56 items in A Seq and the same 56 in scrambled order in A Scr. There were 48 items in each form of B.

Forty-one native speakers of English and 201 adult students of English as a Foreign language (EFL) participated in the study. The native speakers of English were all tested at the same time, and the non-natives were tested in groups, depending on class sizes. Every subject took a test over one

of the passages in the sequential condition and a different test over the other passage in the scrambled condition (that is, each subject took two tests). The design was counter-balanced for possible effects or order and passage. There were four combinations of condition (sequential or scrambled), passage (A or B), and order (A Seq B Scr, B Scr A Seq, A Scr B Seq, B Seq A Scr). Thus, four groups of approximately equal numbers of subjects each completed one of the four combinations of tests.

The native speakers of English did almost equally well on passages A and B while non-natives consistently did better on B. There was also a greater spread between the means for sequential and scrambled conditions for natives on the two passages (9.38% for A and 13.32% for B). The significant interaction between level of proficiency and condition $F(3210) = 3.66$, $p < 0.5$ sustain the hypothesis that as language users become more proficient, they also become increasingly competent in the use of discourse constraints. Thus the contrast between sequential and scrambled conditions increases significantly as subjects become more proficient in English.

Obviously, the same conclusion is sustained if the native speakers are contrasted with the non-native speakers lumped into a single rather than a three level group. Thus, it was suggested that at least for the population sampled in

this study, the cloze procedure is sensitive to discourse constraints ranging across sentences. Furthermore, the effects of these constraints are incremented with an increase in proficiency.

2) In another study Markham (1985) investigated the intersentential sensitivity to cloze items. This study was mainly on the global comprehension in German (as a foreign language) and he used the rational deletion cloze that is a cloze procedure where only the content words are deleted. His assumption was that the ability to recognize intersentential clues is necessary for global comprehension. Although our interest is not directly in this phenomenon, his research design of using alternate forms, sequential and scrambled has something to do with a sensitivity of cloze items to remote contexts. Let us see what Markham has to say.

According to him many second language reading researchers, such as Bachman (1978), Oller (1973), Cziko (1978) and Clarke (1979), maintain that the cloze procedure is an objective, reliable measure of global comprehension. However, there are many first language reading researchers such as MacGinitie (1961), Kibby (1980), Smith-Burke, Gingrich, and Eagleye (1978) and Shanaham, Kamil and Tobin (1982) who question the sensitivity of the cloze test to intersentential processes.

Markham reports that Bachman (1982:62) and Clarke (1979:126) assert that the rational deletion cloze procedure (exclusively content-word deletions) is superior to mechanical deletion (every nth word), because content words tends to reflect cohesive processes across sentence boundaries. They assert that the function-words in mechanical deletion cloze tests, as a rule, only reflect grammatical processes on the level of individual sentences. Because the critics of cloze intersentential sensitivity have consistently used mechanical deletion, it was decided to employ rational deletion in this study.

The following research questions were addressed :

1) Is intersentential comprehension in German affected by a scrambled order of sentences within a cloze passage using rational deletion ?
2) How does subject proficiency level in German influence performance with regard to the first research question ?

A total of 84 college German students participated in the investigation.

Markham mentions that there are many techniques for constructing scrambled cloze test. For example, Shanaham, Kamil, and Tobin (1982:232) altered the sentence length of

the original, sequential cloze passage before scrambling them. This procedure might have disturbed the cohesive relationships of the passages. Oller (1975) divided sequential cloze passages into units varying from 10 to 100 segments and then scrambled the order of the segments without respect to existing sentence boundaries. As a result, both within-sentence and between-sentence boundaries were disrupted.

Chihara et. al. (1977) had created a scrambled list of individually numbered cloze sentences from a sequential text. The sequential cloze test was presented in paragraph form with an every 7th word deletion rate. The deletion rate in scrambled test was consistently fixed at every 7th word within each sentence, but the deletion rate between sentences fluctuated greatly. The scrambled cloze subjects had the disadvantage of completing a cloze test that was different from the sequential cloze test in terms of appearance and deletion rate between sentences.

Given the various alternatives for scrambling the order of sentences within a cloze passage, he decided to construct a scrambled cloze task that deleted the same words in each passage. The deletion rate in the scrambled passage was allowed to vary from 7 to 9 while the deletion rate in sequantial passage was fixed at every 8th word. MacGinitie had (1961:121-30) found that deletion rates fluctuating from every

5th word to 12th word do not substantially alter the subject's performance in cloze research.

Three factors were examined in the study. A 2x3x2 fixed factor design was used. The first consisted of two levels of cloze context passages. One cloze passage was a normal, sequential test with every 8th word deleted. This provides a measure of the student's performance with an unaltered, cloze format. A scrambled passage was constructed by randomizing the order of sentences. A second factor, subject proficiency in German, included three ability groupings: beginning, intermediate and advanced students of German. A third factor consisted of the two different passages. Passage A and Passage B. The use of two passages enhanced the generalizability of the findings.

All subjects were asked to complete one cloze test. Time was limited to 30 minutes to minimize the possibility that the subjects might have time to reconstruct the original sequence from the scrambled items.

Only content-word deletions (substantives, verbs and modifiers) were scored. Deletions that involved function words (determiners, conjunctions, prepositions and interjections) were not included. Passage A consisted of 58 blanks of which 39 were content-word deletions. In Passage B there were 69

deletions of which 48 were content words. Three scorers participated in the research effort and interrater readiability indices of 0.93 (exact score) and 0.81 (acceptable score) was determined.

There were no significant differences in performance on sequential versus scrambled cloze tasks in the exact or the acceptable scoring conditions. None of the F values was significant for any level. There was, however, a significant passage effect. The scores associated with Passage A were consistently higher. Nevertheless, this did not render the data difficult to interpret because there was no significant interaction between passage and any of the other variables. The results were consistent across all levels and all orders. The results concerning the second research question were also easy to interpret. The significant results with regard to subject proficiency in German were logical and consistent with the student's actual progress in learning the language. These results serve to establish the credibility of the cloze test instruments that were used in this experiment.

In earlier studies Bachman (1982:61-69) and Clarke (1979:121-45) assumed that the cloze procedure with rational deletion would have greater intersentential sensitivity than mechanical deletion cloze tasks. The results of this study, however, provide no evidence to support that assertion. Despite the intuitive appeal that content words should be more heavily

involved in cohesive processes across sentence boundaries than function words, the results of this investigation did not support rational deletion.

When the conclusions of this study are combined with Markham's (1984) mechanical deletion cloze investigation, a more complete picture of global cloze sensitivity is revealed. A comparison of the mechanical deletion and rational deletion results, using the same cloze passages, yielded the following generalizations;

1) There are no significant performance differences between sequential and scrambled tests with mechanical or rational deletion.
2) The means associated with rational deletion (sequential or scrambled) were generally lower than the means associated with mechanical deletion. The second generalization is not surprising because one would expect that content words would be considerably more difficult to fill in than function word.

The findings of these investigations point to the conclusion that the cloze procedure, with mechanical or rational deletion, does not provide an adequate assessment of intersentential comprehension. Rather, the nature of the cloze task itself seems to reinforce a sentence-level or even

subsentence-level processing strategy. Apparently, students attend to the immediate environment of the blank, not the larger context. Students are by and large, not rewarded for paying attention to the global cohesive ties in the cloze test. This brings into question the assumption that cloze tests are a valid means of assessing global comprehension proficiency.

Obviously, these findings are contradictory to Chihara et.al. (1977). This may be due to two possible factors - one, the cloze procedure used here was to test the global comprehension in German as a second language, whereas Chihara used English both as first and second language; and two, intersentential or discourse clues may be text specific. Therefore, Markham's findings does not necessarily warrant a generalization that cloze items are not sensitive to remote context.

Conclusions :

From the limited review of literature presented in this chapter, we have seen that empirical findings on the sensitivity of cloze items to the surrounding context can not be specified in generalized terms. So it would appear that cloze test items are uniquely sensitive to the text in which they occur. It may be hypothesized that the difficulty level of both cloze tests and cloze items are text-dependent. Also,

perhaps performance is dependent on the passage, the cloze procedure which produces a uniquely determined set of cloze items, and the subject's profieicncy in that type of text. Some useful findings in this regard are found in Chavez-Oller et. al. (1985). There is a need for further research in this regard.