CHAPTER II

REPLICATION OF THE EXPERIMENT

Chapter II

Replication of the Experiment

The Objective :

The need for replicating an experiment on determining the effect of varying N has been elucidated in Chapter-I. The object of the present experiment was to replicate one such experiment and see if the earlier claim is sustainable through one more such experiment.

On going through the available literature on experiments on this particular topic, it was felt that Nihalani (1979) came nearest to our idea of a viable experiment for the purpose. Before describing our experiment it may be useful to summarize Nihalani's work.

The aim of this experiment was to varify some of Oller's findings regarding the effect of varying deletion rates from every fifth to every twelfth word. Oller (1972) reports that there is no significant effect on the performance of varying N between the frequencies every fifth and every twelfth word. Nihalani also wanted to varify if a very low frequency deletion, that is, deletion of every second word would affect the performance of a second language learner significantly.

In this experiment, a taught prose passage entitled 'Oil' was chosen for cloze test construction. It was about 900 words in length and was taken from Bhaskar and Prabhu, English Through Reading, MacMillan, 1975. The four deletion rates taken up for experimentation were : every second, fifth, eighth and twelfth word. Using the same prose passage, four test formats, each comprising all the four cloze procedures differently ordered, were constructed. They were as follows :

PROCEDURES

FORMATS :	I	:	A	B	C	D
	II	:	B	C	D	A
	III	:	C	D	A	B
	IV	:	D	A	B	C

Procedure	A	:	Every fifth word deleted
	B	:	Every second word deleted
	C	:	Every eighth word deleted
	D	:	Every twelfth word deleted

Each procedure had 30 items and thus each test format consisted of 120 items. A suitable lead was provided in each test format, and each deleted word was replaced by a line of ~~the~~ uniform length. Cloze procedures were differently ordered in different test formats in order to cancel out any unevenness resulting from a

particular order. This, Nihalani felt, also helps one to ensure that all procedures get equal attention and chance of being attempted because in a long passage parts towards the end tend to get less of testee's attention.

Sixty first year university undergraduates doing a compulsory course in General English at S.B. College, Aurangabad were tested on the cloze passage. Students were allowed one full hour for the test. The test was scored by 'contextually acceptable word' method.

The data were analysed by determining means ($\bar{X}$) and standard deviations (SD) for all procedures. $\bar{X}$ and SD for different procedures were as follows :

PROCEDURES

	Possible Score	A	B	C	D
$\bar{X}$	30	19.9	18.33	17.6	20.58
SD		9.15	10.91	10.10	8.70

With the exception of Procedure C, the mean scores for different cloze procedures were as expected - being highest (20.58) for D and lowest (18.33) for B. Furthermore, SD decreased as $\bar{X}$ increased. However, the mean score for Procedure C was only 17.6 (SD = 10.10). It is anomalous that the mean score for

Procedure C (the 8th rate deletion) should be less than those for Procedures B and A (2nd and 5th rate deletions).

Next, the significance of these findings was determined by applying the t-test for correlated samples. From the observed value of t at 5% level of significance with 59 degrees of freedom, it was found that the 8th rate deletion was significantly more difficult than the 12th rate deletion though the 5th rate deletion was not.

Regarding the 2nd rate deletion, it was pointed out that one would expect it to affect the performance significantly, but in this study it did not. This could be due to the fact that learners were tested on a taught passage. The experimenter was intrigued by the anomalous behaviour of Procedure C (8th rate deletion) as well. This was at sharp variance with Oller's findings (1973), and it was also counterintuitive that the 8th rate deletion should prove to be more difficult than the 12th rate deletion when the 5th rate deletion did not.

Nihalani's Findings :

Nihalani concludes that with these reservations the experiment supported Oller's (1973) findings. However, although there was some lowering of scores in 2nd rate deletion, the difference was not significant.

If we look closely at these conclusions, it would appear that there are some reservations about the catagorical conclusion that Nihalani's experiment supported Oller's findings (1973). From the t-test on Nihalani's results, it was observed that the 8th rate deletion was significantly more difficult than the 12th rate deletion, which means that the scores did not remain 'constant' between 8th and 12th rates of deletion. Moreover, the 2nd rate deletion did not affect the results significantly, which means that the scores tended to remain constant. All this means that there are some contradictions in maintaining that scores remain constant even if deletion rates are varied from second word to twelfth word in this particular experiment. Within the research design, Nihalani prepared four formats consisting of four sections A, B, C, D with different rates of deletions (keeping the number of items equal - 30) and ordering the four sections in four different ways. This was to ensure that all the four sections received equal attention, as in the usual course the later sections tend to get less attention of students. Nihalani's results in this regard show that this is largely true. In all but the second format, the scores are the highest in the section that occurs first, irrespective of the rate of deletion.

With these findings in view it was felt that one more experiment might yield some significant findings. Therefore it was decided to replicate Nihalani's experiment with two

modifications - one, using deletion rates of every fifth, seventh, ninth and eleventh words and preparing the test on a much easier but untaught passage; two, trying it out at two somewhat different levels of ESL learners, undergraduate and postgraduate.

Method

Research Design and Materials :

In our experiment, we followed the research design of Nihalani (1979). He had selected a taught passage for the cloze test, about 900 words in length. It was taken from Bhaskar and Prabhu English Through Reading, MacMillan, 1975, which was a prescribed text-book for first year university undergraduates doing a compulsory course in General English, at Marathwada University, Aurangabad. The four deletion rates taken up by him for experimentation were : every seoond, fifth, eighth and twelfth word.

While selecting our passage we were careful about the level of difficulty of the passage and the students' ability of reading comprehension. We selected a passage, 'Jawaharlal Nehru and His Qualities', written by Krishna Hutheesing, 838 words in length. It was taken from English Prose for Pre-University Classes, Somaiya Publications Pvt.Ltd., 1974, which was a prescribed text-book for first year university undergraduates, doing a compulsory course in General English, at S.N.D.T.University,

Bombay. (The passage is given in the Appendix-I). The four deletion rates taken up for experimentation were : every fifth, seventh, ninth and eleventh word. Using the same prose passage, four test formats, each comprising all the four cloze procedures differently ordered, were constructed. They were as follows :

PROCEDURES

FORMATS :	I	:	A	B	C	D
	II	:	B	C	D	A
	III	:	C	D	A	B
	IV	:	D	A	B	C

Procedure	A	:	Every fifth word deleted
	B	:	Every seventh word deleted
	C	:	Every ninth word deleted
	D	:	Every eleventh word deleted

For each format the first and the last sentence were left intact. Contractions, abbreviations and hyphenated combinations were all regarded as single words. Blanks indicating the deletion of words were of ten spaces in length. Each procedure had 25 items and thus each test format consisted of 100 items. Cloze procedures were differently ordered in different test formats in order to cancel out any unevenness resulting from a particular order on the same assumption that this helps to ensure that all

SHIVAJI UNIVERSITY LIBRARY KOLHAPUR

procedures get equal attention and chance of being attempted because in a long passage parts towards the end tend to get less of testee's attention. (The full text of all the four formats is given in Appendix-II).

Subjects :

The test was administered to two groups of students at two different levels, one at an undergraduate level and another at postgraduate level.

At the undergraduate level, forty-eight second-year university undergraduates doing a compulsory course in General English, drawn from Commerce faculty, at Kamala College, Tararani Vidyapith, Kolhapur were tested.

At the postgraduate level also forty-eight students doing their first year of post-graduation in English from Shivaji University were tested.

Administration of the Cloze Test :

While administering the cloze test, at the undergraduate level, the forty-eight subjects from Kamala College were divided into four groups, each group consisting of twelve subjects. Then they were given the four cloze test formats - one test format to each group of twelve subjects.

As the cloze test was unfamiliar to them, the nature of cloze test was explained. They were given two examples of cloze test from the taught lessons of their text-book. The passage selected for the cloze test was unknown to them, so the passage was read aloud in the class. The summary of the passage was told briefly in Marathi, their native language. It was not sure that though the passage was read aloud, the subjects could remember the content words (verbs, nouns and adjectives) at the time of filling the blanks. The loud reading was intended to convey the overall content of the passage. The instructions for doing cloze tests were read aloud, which were at the head of each test-paper. It was pointed out that only one word should be used to fill each blank. It was also clearly explained that this was an experiment : no personal results were to be recorded officially, so nothing was to be gained by cheating. A full hour (60 minutes) was given to them.

The same cloze test was then administered at the post-graduate level. Forty-eight students doing their first year of post-graduation course in English were tested. They also were divided into four groups, each group consisting of twelve subjects. The four formats of the cloze tests were given to these four groups. As they were students of English Language and Literature and as they were also familiar with cloze procedure nothing was explained to them, about either the nature of the cloze test or the passage. Only instructions were read aloud to

them. A full hour was given for the test to the subjects.

The test was scored by 'contextually acceptable word' method. Anderson (1971) suggests that in deciding which alternative may be accepted, it would be a good idea to have an objective reference book for the purpose (Collin's Gem Dictionary of Synonyms and Antonyms). This might serve the purpose in the case of content words, but the problem still remains in the case of structure words. In order to be as objective and reasonable as possible, the following procedure was followed.

All the alternatives that appeared to be acceptable to the researcher for those blanks where several alternatives were filled in by the students were listed. Each set of alternatives - both for content words and structure words was discussed with the supervisor, Dr. S.V. Shastri and a final decision made about acceptability. Thus a response was either accepted or totally rejected. The raw score sheets for each procedure and each format were prepared (These are given in the Appendix-III).

Results :

The data were tabulated and the histograms representing frequencies were drawn (These are given in the Appendix-IV). The results of the two level groups - undergraduate and post-

graduate - were subjected to the usual statistical analysis. The statistics thus arrived at are given below.

Undergraduate Group :

In the undergraduate group, the analysis of the scores showed that the modal (maximum) scores for each of the procedures A, B, C and D were 19, 20, 18 and 18 respectively. Thus the preliminary analysis revealed that in the case of C and D the subjects had done equally well and in the case of A and B they had slightly better.

The data were then analysed by determining means and standard deviations with the help of a statistician. $\bar{X}$ and SD for different test formats and procedures were as follows :

TEST FORMATS

	Possible Score	I	II	III	IV
$\bar{X}$	100	39	36.67	30.75	28.92
SD		19.94	12.82	9.38	7.74

PROCEDURES

	Possible Score	A	B	C	D
$\bar{X}$	25	7.85	8.58	8.98	8.83
SD		4.66	4.15	4.51	4.08

With the exception of C, the mean scores for different cloze procedures were as expected, that is, there was a slight improvement in the mean scores from Procedure A (7.85) to B (8.58), B to C (8.98), but from C to D (8.83) instead of an improvement as expected, there was a slight deterioration.

Next the significance of these findings was determined by applying the t-test for correlated samples. The formula used was as follows :

$$t = \bar{d} \div \sqrt{\frac{\Sigma d^2 - (\Sigma d)^2/n}{n(n-1)}}$$

with (n - 1) degrees of freedom

where d = difference between each pair of scores

$\bar{d}$ = mean difference

n = number of pairs of scores

The observed value of t at 5% level of significance with 47 degrees of freedom were :

	A	B	C	D
A	-	- 0.79	- 1.48	- 1.67
B	0.79	-	-0.79	- 0.06
C	-1.48	- 0.79	-	0.16
D	-1.67	- 0.06	0.16	-

The critical value of t at 5% level of significance with 47 degrees of freedom is 2.000 (Table E : Robson, 1973).

As all the observed values of t are numerically smaller than the corresponding critical value of t, one is led to infer as follows : The 5th and 7th rate deletions were not significantly difficult.

Postgraduate Group :

In the postgraduate group, the analysis of the scores showed that the modal (maximum) scores for each of the procedures A, B, C and D were 25, 24, 23 and 23 respectively. Thus the preliminary analysis revealed that the subjects had done equally well on all the procedures and slightly better on Procedure A.

The data were then analysed by determining means and standard deviations with the help of a statistician. $\bar{X}$ and SD for different test formats and procedures were as follows :

TEST FORMATS

	Possible Score	I	II	III	IV
$\bar{X}$	100	48.59	40.25	41.33	49.75
SD		20.77	8.19	16.01	9.30

PROCEDURES

	Possible Score	A	B	C	D
$\bar{X}$	25	10.39	11.27	12.15	11.17
SD		4.78	4.80	4.93	5.05

The highest value for mean is 12.15 for Procedure C and the smallest value for mean is 10.39 for Procedure A.

With the exception of Procedure C (12.15) the mean score for different cloze procedures were as expected, that is, there was a slight improvement in the mean scores from Procedure A (10.39) to B (11.27), B to C (12.15) but from C to D (11.17) instead of an improvement as expected, there was a slight deterioration.

Next the significance of these findings was determined by applying the t-test for correlated samples. The formula used was as follows :

$$t = \bar{d} \div \sqrt{\frac{\Sigma d^2 - (\Sigma d)^2/n}{n\ (n-1)}}$$

with (n-1) degrees of freedom

where d = difference between each pair of scores

$\bar{d}$ = mean difference

n = number of pairs of scores.

The observed value of t at 5% level of significance with (48-1) = 47 degrees of freedom were as follows :

	A	B	C	D
A	-	- 1.37	- 1.90	- 1.29
B	- 1.37	-	- 2.11	- 0.63
C	- 1.90	- 2.11	-	0.99
D	- 1.29	0.63	0.99	-

The critical value of t at 5% level of significance with 47 degrees of freedom is 2.000 (Table E : Robson, 1973).

As all the observed values of t except for Procedure C (-2.11) are numerically smaller than the corresponding value of t, one is led to infer as follows :

1) Though the 9th rate deletion did make the test more difficult than each of the remaining three deletion rates, it didn't affect performance.
2) The 9th rate deletion was more difficult but 5th and 7th rate deletions were not.

Discussion :

1) It was expected that increasing the value of N in rates of deletion would affect the performance significantly. But it was found in the study that it does not. This is in accordance with the findings of earlier researchers Oller (1973) and Nihalani (1979).

2) However, the unexpected lower performance in Procedure C is almost identical with what Nihalani discovered. The C procedure had 9th rate deletion whereas D procedure had 11th rate deletion. One would expect that the performance on D procedure would be better than that on C procedure. But in both the studies (Nihalani's and ours), it was discovered that the scores deteriorated instead of improving.

It must be recalled that Nihalani used a taught passage for constructing the test and so the equally good performance on 2nd rate deletion was explainable. In our test, we did not use such a low rate of deletion; nor did we use a taught passage.

3) In Nihalani's experiment some subjects had left the C procedure unattempted where it occurred in the last position, Format IV. In our experiment, no such thing happened. However, it was noticed that in all the formats the procedure that came first in order produced the best scores, except Format II, in which the best scores were produced on C procedure which was second in order. This was found to be true of not only our undergraduate group but also our postgraduate group. What is more, this turned out to be true of Nihalani's scores also (See Appendix-V).

From these findings, it appears that there is a need for paying greater attention to research design and for doing further

8618 BARR. BALASAHEB KHARDEKAR LIBRARY
SHIVAJI UNIVERSITY, KOLHAPUR.

A

research in view of the repeated findings that change in rate of deletion does not affect general performance. As we have seen in both Nihalani's and our experiment, some intriguing, counterintuitive results are produced. This is particularly so in the case of C procedure in both experiments where the N is 8 (Nihalani) and 9 (ours). Again, the C procedure has produced poorer scores than the D procedure where the N is 12 (Nihalani) and 11 (ours).

Counter to all expectations the Format II, produces the best results with the C procedure and not B. The C procedure has deletion rates of 8 (Nihalani) and 9 (ours). Given the present available empirical data, there seems to be no explanation for this peculiar behaviour.

At this stage, it may be useful to take an altogether different line of approach to the problem. From the bulk of theoretical and empirical information, it appears that two questions affecting the research design used so far in these experiments are worth looking at - equivalence of cloze test formats and the effect of intersentence clues on performance. These two together seem to suggest that the possibility of filling in every blank in a cloze procedure is uniquely determined, which means that N is largely irrelevant. It is

the word that is blanked out and the context in which it occurs and not the frequency of intervals at which it happens to get blanked.

In view of all this, I would like to review the literature relevant to these points in the next chapter.