

CHAPTER III

EXPERIMENTAL VALIDATION OF THE PROGRAMMED TEXT

Evaluation of programmed learning material for establishing its effectiveness has been considered an integral part of the process of developing such material. Strategies of evaluation adopted in evaluating programmes can be distinguished as internal and external (Hartley, 1972). Internal evaluation procedures include study of criterion test scores, analysis of errors on the programme frames and criterion test items, etc. Internal evaluation through such procedures is carried out as a part of the developmental and field testing of the programmed material. The main purpose of internal evaluation is to revise the programme frames and sequences on the basis of empirical results with a view to improving the effectiveness of the programmed material. Thus, internal evaluation procedures provide answer to the question, 'how well does the programme teach what it purports to teach?' Details of internal evaluation made with regard to the programmed text developed under the present study have already been discussed in the previous chapter.

External evaluation of programmed learning material essentially refers to the validation of the

programme against an external criterion. The external criterion generally set would be in terms of learning effected by the programme vis-a-vis some other method of instruction. Thus external evaluation of a programme involves experimental comparison of learning effected by the presentation of some instructional material through the programme, and the learning effected by the presentation of the same material through one or more other methods of instruction. Such a study, therefore, provides answer to the question, 'how well does the programme work as compared to some other method of providing instruction. In other words, the underlying purpose of such validation experiments is, mainly, to explore the advantages or otherwise the presentation of some instructional material in the form of a programme has over another method of presenting the same material, in terms of learning effects and administrative feasibility. This implies that such studies are to be carried out, generally, after the programmer is satisfied with the standard of the programmed material regarding its effectiveness studied through internal evaluation procedures. However, results of the comparison studies can also be beneficially used by the programmer for further revisions of the material. In fact, the programmer would receive continuous feedback as to the effectiveness of the programme, everytime a new group of

students use the programme. Details of a validation experiment conducted to study the effectiveness of the programmed text developed under the present study have been given in this chapter.

Research literature in the field of programmed learning abounds with reports of studies involving comparison of programmed learning and other methods of instruction. Some of the studies conducted to compare the effectiveness of programmed learning technique with that of the so called 'traditional' or 'conventional' technique have been briefly reviewed in the following. A brief discussion of the nature of these experiments and the validity of the conclusions drawn through them have also been presented.

Comparison Studies

Comparison experiment has been the most popular type of research work in the field of programmed learning. Particularly, the 1960's recorded an unprecedented proliferation of such studies undertaken in an attempt to establish the effectiveness of programmed learning as an instructional technique. Many of these studies have compared programmed learning with what is usually referred to as 'conventional' or 'traditional' method of teaching.

Hartley (1972) reviewing 110 studies comparing programmed instruction with 'conventional' instruction, reported that in 41 studies programmed instruction group was significantly superior to the other in terms of test results, however, in 54 studies results did not indicate any significant difference between the two methods; and, programmed instruction group did significantly worse than the other group in the remaining 15 cases. He has further pointed out that any individual study is limited in its implications, and that the overall picture shows a sorry state in terms of scientific rigour.

Following suit with their counterparts abroad Indian researchers also conducted a large number of studies comparing programmed instruction and 'conventional' or 'traditional' method of instruction. In fact, Kulkarni and Kapadia (1974) reviewing the research work done in the country at doctoral and institutional level concluded that the majority of the studies were comparison studies.

Shah (1968) conducted the first systematic research study in the field of programmed learning in India. She developed a programme in Mathematics on 'Solving Equations' for the class VI students and conducted an experiment using it. Two experimental groups were set up along with a control

group, consisting of in all 90 students drawn from an English medium school in Delhi. Students in one of the experimental groups learnt independently through programmed material, while in the other group class teacher helped the students when they read the programme. The control group was taught through the conventional lecture method. Duration of the experiment was five days. It was found that the group taught by programme alone did significantly better than the conventional lecture group on the post test. However, no significant difference was found between the mean post test scores of the two experimental groups.

Sharma (1968) gave a programme on 'sets' to a group of students belonging to class IX and taught a comparable group of students the same topic through the usual lecture method. Comparing the performance of the two groups on a post test, he concluded that although the difference was not significant, the study provided enough evidence to believe that the programmed instruction was as good as the usual lecture method.

Gupta (1968) conducted an experiment using the modified version of a programme on 'Force' developed abroad. The programme was given to 19 students of grade VIII and their performance was compared with that of another group

of 21 students of the same grade who were taught by the usual method. The criterion test results did not reveal any significant difference between the two groups.

Sharma (1968) conducted an experiment on 80 students of class IX who were divided into an experimental and a control group. Mean achievement score of the experimental group which was taught through programmed instruction was found to be significantly higher than that of the control group which was taught by the teacher through the lecture method.

Desai (1968) developed a programme on "the types of compounds" in Gujarati language and used it in an experiment involving 80 students of class IX who were divided into two matched groups. Based on the post test scores, she concluded that the programmed learning approach was more effective than the conventional approach.

Sharma (1968) developed a programme on "Earth's Rotation and Revolution" and used it in his experiment with students of class VIII. The sample for the study included 133 students of which 62 were from urban and 71 were from rural schools. Experimental and control groups were separately set up for the two samples. Mean difference in

the gain scores of the two groups was significant showing a clear superiority of programmed instruction over the conventional method of teaching.

Shah (1969) developed programmed learning material covering the whole syllabus in Algebra of std. VIII for using in her experiment. For the purpose of experimentation, two comparable groups were set up in four schools of Ahmedabad. It was found that the total mean score of the experimental group was higher than that of the control group indicating the superiority of programmed instruction over the conventional method of teaching.

Patel (1970) conducted an investigation with a view to finding out the relative effectiveness of programmed instruction and conventional method for teaching a topic in Mathematics to students of std. VIII. He separately conducted an experiment with a rural sample of 90 students taken from 5 schools in Dabhoi area and with an urban sample taken from two sections of a school in Baroda city. Findings of the study, reported separately for the two samples, revealed that in both cases programmed instruction and the conventional method groups did not differ in their mean achievement on the teacher made test as well as on the criterion test.

Mehta (1973) conducted an experiment involving 252 pupils of grade V belonging to six schools in the city of Baroda. The experimental group was given the programmed reading material prepared by the investigator whereas the control group was taught through the conventional teaching method. Results showed that students in the experimental group were significantly better than their counterparts in the control group, in terms of their scores on an immediate post test.

Two comparative studies were conducted with B.Ed. students. A programme on 'Action Research' was used by Kaushik (1968) in her study with B.Ed. students of Meerut University. She concluded that programmed instruction was superior to conventional lecture method in terms of student attainment on the criterion test. Mullick (1968) conducted an experiment on 128 B.Ed. students taking the correspondence course of Delhi University. The experimental group was given a programmed lesson on 'Rank Correlation'; the control group was given lessons on the same topic prepared on conventional lines. Based on the post-test performance, it was concluded that the programmed lesson was better than the conventional lesson.

In the studies reported above almost all the researchers have concluded that programmed instruction is

more effective than the 'conventional' method which has also been referred to as 'traditional' or 'lecture' method. However, it is doubtful whether such a conclusion would stand good on a close scrutiny of the details of these experiments. First of all, these studies except the ones conducted by Shah (1969) and Mehta (1973) are short term experiments and do not provide evidence to the effectiveness of programmed instruction when utilised on a long term basis in the regular instructional programme. Further, what the researchers have compared programmed instruction with is something loosely defined as "conventional method". The latter, as Lumsdaine (1963) has pointed out, has usually meant some unspecified combination of presentation by some instructor, plus perhaps some unspecified use of some text or other study material. Since the very method to be compared is non-specific and non-analytically defined, it does not allow for any meaningful conclusion being drawn about the comparative effectiveness of another method.

A more serious limitation of the above studies lies not in their design or execution but in the very purpose with which they have been conducted. Most of these studies have had the purpose of wider generalization

rather than the limited purpose of evaluating that particular programmed material. What Roshal (1960) has characteristically pointed out regarding such studies of comparison involving instructional films holds equally good in case of similar studies involving programmed learning material. What has been compared in such a study is some programme, not the best possible programme nor a representative (in the sampling sense) programme with some other instruction; not the best possible teacher, nor a representative teacher, but some teacher; also, neither the programme nor the compared medium (which is itself non-analytically specified) is chosen by a sampling technique which lends itself to determination of an estimate of error. Thus conclusions drawn from any study of the type reported above which involves the evaluation of a single instrument (programme) apply only to that particular instrument, and the generalisation of the results to other instruments of the class it claims to represent have, at most, the status of untested hypothesis (Lumsdaine, 1963).

Studies of comparison may be carried out with two scientifically meaningful purposes. First purpose is to make an overall assessment of the programme to determine how much does the programme teach and to decide whether enough is learned through its use warranting its suitability

for instructional purposes when compared to an external standard, viz., learning effected by the other method of instruction being compared. Thus, a comparative evaluation experiment is essentially a method of validating the programmed material against an external criterion. However, in such comparisons, in order to draw meaningful conclusions about the effectiveness of the programme, it is essential to satisfy certain basic conditions regarding the procedure of experimentation. First, the method that is being compared to the programme, must be analytically defined in terms of the core software material and the mediating processes involved in its presentation. Secondly, there should be no discrepancy as regards the content input in the programme and the other method; as far as possible the sequence of presentation of different content points should also be same, as the order of learning of different points may affect the overall learning. Thirdly, actual instructional time under the two methods must be equated. Finally, as in any such experiment, it is essential that the subjects forming the treatment groups must be comparable in terms of the entering behaviour specifications. A second purpose of a comparative evaluation study is 'diagnostic'. For this purpose, in addition to the overall assessment, learning effects in relation to different specific content units would also be measured and compared. Such a measurement

of instructional effects would guide the programmer to decide whether to use or further revise particular content units of the programme.

Experimental Validation of the Programmed Text.

The following is a detailed account of validation experiment conducted to study the effectiveness of the programmed text developed under the present study. Learning effects of the programme have been compared to those of lecture method. An operational definition of lecture method as adopted in the experiment is presented before stating the specific objectives of the experiment.

Lecture method :- Lecture is an activity carried out to bring about desired changes in students' cognitive behaviour. The content material to be delivered by the teacher through his lectures is well structured as to the points included and the sequence of presentation. Also, the process of communication in the classroom allows for intermittent interaction between the lecturer and the students.

Objectives of the experiment :

- (1) To make unitwise comparison of the effectiveness of the programmed text with that of the lecture method.

- (2) To compare the effectiveness of the programmed text as a whole with that of the lecture method.
- (3) To find out the frame sequences in the programme that need improvement and, based on that, to prepare the final form of the programmed text.

Sample :

Sample for the study consisted of 69 B.Ed. students. The sample was randomly selected from a total of 136 students admitted to B.Ed. class of the M.S. University of Baroda during the year 1973-74.

Design of the Experiment :

Matched group design was adopted for conducting the experiment. The sample was divided into two matched groups, viz., an experimental group and a control group. The experimental group consisted of 35 students while the control group had 34 students. As one specific entering behaviour in relation to the content of the course could be assumed, matching was done on two general considerations, viz., educational qualifications and teaching experience. Students in the sample had varying educational backgrounds, as they come from different academic streams, viz., Arts, Science, Commerce and Home Science. Secondly, some of them were

postgraduates although the majority were graduates. It might be argued that students with different academic backgrounds would differ in their language facility and in their facility to carry out arithmetical operations. Thus educational qualification of the students was taken as one consideration in matching the two groups. Secondly, the sample included students who were fresh graduates as well as those who had the experience of teaching in school for varying lengths of time. This factor, it was thought, might affect the achievement of the students. Accordingly, teaching experience was taken as the second consideration for matching the two groups. Groups were matched on the above two factors, by matching the students by pair. Apart from controlling these two factors, differences in the mean achievement of students in the two groups due to intelligence were controlled by employing the technique of analysis of covariance.

Experiment :

The experiment was conducted for one full semester. In the beginning, students in both the groups were given orientation regarding the instructional procedures to be adopted for the particular course, viz., Educational Testing and Techniques of Evaluation in the two groups. It was brought home to the students that the procedures to be adopted

in the two groups formed the regular instructional work at the Faculty. They were informed that the criterion tests to be held at the end of each of the six units of the course and the comprehensive test at the end of the semester constituted the assessment scheme for the course.

The experimental group students learnt through the programmed text. The programmed text was presented in the form of six programmed units corresponding to the six units of the course. Based on the tryout study average time required for completing each unit was estimated. Students studied the programme for each unit in scheduled classhours for the estimated length of time. A few students who could not complete the unit in specified classhours were allowed to study the remaining part at home. After each unit was completed students were given the programme for that unit. Students were administered the criterion test at the end of each unit. Feedback sessions were organised to discuss the performance of the students on the criterion tests immediately after each test.

The control group was taught the whole course through the lecture method. Structuring of the lectures was done by providing the teacher as well as the students with booklets containing the course content organised in the sequence in which the teacher would deal with it in

his lectures. Six booklets were prepared to cover the six units of the course. These booklets have been given in Volume II of the thesis. At the beginning of each booklet the objectives of the unit were specified in terms of questions a student would be able to answer after completing the particular unit. It was explained to the students that the booklets were not to be taken to replace teacher's lectures in the class; they were meant only to help them recall in an organised manner what would be delivered through the lectures. Booklet for each unit was given to the students, generally, after the particular unit was taught in the class. Although the content to be delivered through the lectures was planned before-hand and put in the booklets, the course of the lecture was made flexible enough to provide for intermittent interaction. At the end of each unit the criterion test for that unit which was the same as the one given to the experimental group was administered. Feedback sessions were organised to discuss the performance of the students on these tests immediately after each test.

Although the two instructional procedures were different proper controls were introduced so that there would be no discrepancy in the content input and the total instructional time for the two groups. Booklets given to the control group and the programmed units given to the experimental group were developed based on a common task analysis of the course content. This was done to bring about

a common base regarding the content input and also to ensure that the sequence of presentation of the content points in the two groups remained the same. The classhours devoted for each unit were equal for the two groups. As stated earlier, the experiment was conducted for the full length of the course carried out in one complete semester. At the end of the course a comprehensive test based on all the units was organised for both the groups.

Results and discussion :

Effectiveness of the programmed text was studied unitwise by comparing the criterion test scores of the two groups. Mean differences were tested for significance by using 't' test. Mean achievement scores for the two groups on the six criterion tests are presented in Table 3.1, on the next page.

It may be observed from Table 3.1 that on the tests for units I, IV, V and VI, the mean achievement scores do not differ significantly. For units II and III, the mean scores differ significantly at 0.01 level in favour of the experimental group which learnt through the Programmed Text.

Table 3.1

Mean Achievement Scores on Criterion Tests

	Group	M	S.D.	t-value	Significance
Unit I	Experimental (N=34)	17.40	5.79	1.64	Not significant
	Control (N=34)	14.85	6.84		
Unit II	Experimental (N=35)	23.51	3.24	3.63	Significant at .01 level
	Control (N=34)	19.59	5.33		
Unit III	Experimental (N=35)	29.37	3.25	2.74	Significant at .01 level
	Control (N=34)	25.97	6.36		
Unit IV	Experimental (N=35)	21.46	4.33	0.39	Not significant
	Control (N=34)	21.00	5.13		
Unit V	Experimental (N=35)	17.49	3.24	0.21	Not significant
	Control (N=34)	17.68	4.17		
Unit VI	Experimental (N=34)	12.94	3.73	0.61	Not significant
	Control (N=33)	12.39	3.57		

Trend in performance on the six criterion tests was studied by applying time series analysis. The trend ordinates were obtained by the least square method. The mean scores and the trend ordinates for the six criterion tests are presented in Table 3.2. The mean scores have been for this purpose converted into percentages.

Table 3.2

Mean Scores and Trend Ordinates
for the Six Criterion Tests

	<u>Experimental Group</u>		<u>Control Group</u>	
	Mean	Trend Ordinate	Mean	Trend Ordinate
Unit I	72	78.19	59	64.24
Unit II	78	76.65	64	65.61
Unit III	81	75.11	72	66.98
Unit IV	77	73.57	75	68.35
Unit V	73	72.03	74	69.72
Unit VI	65	70.49	62	71.09

From Table 3.2, it may be noted that there is a decreasing trend in the performance of the experimental group while there is an increasing trend in the performance of the control group. However, in both cases the slope (experimental group : 1.54 and control group : 1.36) happens to be very small showing that the difference is not much.

The unitwise comparison of criterion test scores made above by testing mean differences in achievement scores for significance and through time series analysis indicate that the Programmed Text used in the experiment is as effective as the lecture method in terms of learning effects.

For studying the effectiveness of the Programmed Text, as a whole, two indices, viz., mean scores on the comprehensive test and mean of the combined scores representing the performance on the six criterion tests were obtained. To obtain the combined criterion test score, for each student the actual score on each test was converted into percentages; mean of these percentages over the six criterion tests represents the combined criterion test score for the particular student. The two groups were compared in respect of their mean comprehensive test scores and the mean combined criterion test scores. Comparisons were made after adjusting these mean scores, by analysis

of covariance, for intelligence of the subjects measured using the Raven's Standard Progressive Matrices. The adjusted means for the two groups are presented in Table 3.3. Differences in the adjusted means were tested for significance using 't' test.

Table 3.3

Mean Achievement Scores Adjusted for Intelligence

		Mean	S.D.	T-value
Comprehensive Test Score	Experimental Group (N=34)	62.64	9.93	0.168
	Control Group (N=34)	63.42	10.37	
Combined Criterion Test Score	Experimental Group (N=34)	73.99	10.09	1.901
	Control Group (N=34)	68.28	12.23	

It may be noted from Table 3.3 that neither the difference in mean comprehensive test score nor the difference in the mean combined criterion test score is significant, indicating that achievement through the programmed text and through lectures does not differ. In

other words, it shows that the programmed text as a whole is as effective as the lecture method adopted in the control group in terms of learning effects.

Final Form of the Programmed Text

Itemwise error analysis of students' performance on criterion tests was carried out. Appropriate modifications were made in the programme frames, and frame-sequences based on the error analysis data. It was observed that there was much scope for revision in the units V and VI of the programmed text. It might be noted that the results of unitwise comparison of the programmed text also revealed that the programme effectiveness had decreased towards the last two units. More frames were added to the last two units, particularly to the unit VI : Elementary Statistics in Education. As a result of the revision the total length of the programme was considerably increased. A concise description of the final form of the programmed text is given below. The complete Programmed Text in its final form is given in Volume II of the thesis.

Title : Educational Evaluation - A Programmed Text

Target Population : B.Ed. students - Graduates with basic ability to read and understand simple English and with facility to perform fundamental operations of Arithmetic involving decimal numbers.

Content : The programmed text covers the course on 'Educational Testing and Techniques of Evaluation' prescribed for the B.Ed. classes of the M.S. University of Baroda. The text is divided into six units.

Unit I - Educational Evaluation and Measurement-A

Unit II - Educational Evaluation and Measurement-B

Unit III - Essential Characteristics of a Good Instrument of Evaluation

Unit IV - Major Evaluation Tools and Their Uses.

Unit V - Teacher-made Achievement Tests

Unit VI - Elementary Statistics in Education

The programmed text contains six criterion tests corresponding to the six units.

Style : Linear (but not a traditional Skinnerian type)

Language : English

Length : The frames are numbered separately for each unit. Number of frames in the different units are -
Unit I : 114; Unit II : 114; Unit III : 143;
Unit IV : 141; Unit V : 128; and Unit VI : 172.

The course is covered alongwith other courses of the B.Ed. programme in one full semester.

A Strategy for Utilisation of the Programmed Text in the Instructional Work of the Faculty

The results of the validation experiment indicate that the Programmed Text as instructional material for B.Ed.

students is as effective as the lecture method adopted. It suggests that the Programmed Text may be utilised in the regular instructional work of the Faculty for providing instruction to B.Ed. students in the course on 'Educational Testing and Techniques of Evaluation'. It should be noted that the 'Lecture Method' adopted for providing instruction to the control group of the experiment is not just the same as the method in which instruction is provided in the usual course. The lectures, here, are highly structured, which ⁱⁿ fact ensures that all the necessary content points are delivered, with proper emphasis, by the lecturer. The course of each lecture, that is, the sequence of material is prespecified on the basis of a scientific task analysis of the course contents and objectives. Enough provisions are made for interaction among the students and the lecturer. The booklets provided to the students serve as guides for their reference and help them remember what they have learnt in an organised way. Contrary to these procedures, in lectures given in the usual course, decisions regarding specific emphasis to be laid on different content points, sequencing of instructional points, time to be devoted for teaching different points, etc., are left to the vagaries of the individual lecturer. It may not be too wrong to assume, here, that a well planned and structured lecture of the sort adopted in the experiment is not the method

usually adopted by the lecturers; and also, that this method would be more effective than that of the usually adopted unstructured lectures. This further strengthens the conclusion that the Programmed Text, whose effectiveness was found to be comparable to the structured lectures adopted in the experiment, is effective enough to be adopted as instructional material for the particular course.

When experiments are conducted in special extra-class situations, they do not suggest any practical strategy for implementing the new approach. However, it might be observed that in the present case, the experiment was conducted according to the regular schedules of instructional work specified by the Faculty. Thus, implementation of the new approach, namely, learning through the Programmed Text, should be considered feasible as it would not pose any administrative problem of scheduling or work distribution. The investigator, however, suggests that use of the Programmed Text may not be adopted as the sole instructional technique for providing instruction in the course. Perhaps, the deficits in learning indicated by the criterion test performance can be made up by enriching the instructional work with certain other instructional techniques. A suitable instructional strategy may be evolved which would involve the utilisation of the techniques of discussion,

library work and practical work in addition to the Programmed Text. In such a strategy, basic knowledge related to different concepts in the course would be given through the Programmed Text; knowledge provided through the Programmed Text would be augmented by organising library work; discussion sessions would provide a forum for exchange of ideas and for seeking clarification of concepts covered through the Programmed Text and library work; and practical work sessions would help develop various skills needed by a teacher to carry out educational evaluation in a school. In fact, the Faculty is already conducting an experiment to evolve such a strategy involving the use of the Programmed Text developed under the present investigation and other techniques mentioned above. The experiment is being conducted as an institutional project and does not form a part of the present investigation.

References

- Desai, U.R. (1968) 'Programmed learning versus traditional approach in teaching of Gujarati', IAPL Newsletter, Vol. 3, No. 1.
- Gupta, M.C. (1968) 'Adaptation of programmed learning material developed abroad', IAPL Newsletter, Vol. 3, No. 1.

- Hartley, J. (1972) 'Evaluation'. In Hartley, J. (Ed.) Strategies for Programmed Instruction : An Educational Technology, Butterworths, London.
- Kaushik, S. (1968) 'An experimental study of a programme in action research', Unpublished M.Ed. dissertation submitted to Meerut University.
- Kulkarni, S.S. and Kapadia, G.G. (1974) 'Programmed learning : a trend report'. In Buch, M.B. (Ed.) A Survey of Research in Education, Centre of Advanced Study in Education, M.S. University of Baroda, Baroda.
- Lumsdaine, A.A. (1963) 'Instruments and media of instruction'. In Gage, N.L. (Ed.) Handbook of Research on Teaching, Rand McNally and Company, Chicago.
- Mehta, S.J. (1973) 'An investigation into the effectiveness of programmed material in English in developing reading ability', Unpublished Ph.D. thesis submitted to the M.S. University of Baroda.
- Mullick, S.P. (1968) 'An experiment on a programmed learning lesson in a correspondence course', IAPL Newsletter, Vol. 3, No. 1.
- Patel, D.P. (1970) 'To compare the relative effectiveness of programmed lesson vis-a-vis conventional method in teaching directed numbers', Unpublished M.Ed. dissertation submitted to the M.S. University of Baroda.
- Roshal, S.M. (1960) 'The instructional film'. In Finch, G. (Ed.) Educational and Training media : A Symposium, National Academy of Sciences - National Research Council. Washington. D.C.

Shah, M.S. (1968) 'A programme on solving equations'
IAPL Newsletter, Vol. 3, No. 1.

Shah, M.S. (1969) 'To develop auto-instructional programmes
in Algebra for standard VIII and to find out their
effectiveness in relation to different variables,
Unpublished Ph.D. thesis submitted to Gujarat
University.

Sharma, J.M. (1968) 'Construction of a programmed unit on
'sets' for class IX students', IAPL Newsletter,
Vol. 3, No. 1.

Sharma, M.M. (1968) 'A comparative study of outcomes of
teaching of Algebra by conventional classroom
methods and method of programmed instruction,
IAPL Newsletter, Vol. 3, No. 1.

Sharma, R.A. (1968) 'A study of achievement in Geography
through programmed instruction, IAPL Newsletter,
Vol. 3, No. 1.
