CHAPTER III
REVIEW OF LITERATURE*

Educational research in India is still in its infancy and the existing facilities for the same are very meagre. Inspite of the recommendations of the various Commissions and Committees for the introduction of new type tests, there has not been much research in this field. Of course individual workers and M.Ed. or Ph. D. students, here and there, have been making some contribution but due to the lack of satisfactory arrangements to publicise the results, only a few studies have been published so far. Since there is little co-ordination in the research work being carried out in the various institutions, the task of an investigator is made quite difficult for reviewing the related literature.

In foreign countries, like U.S.A. and England, as a result of large scale research in the field of evaluation and measurement, standardised objective tests and attitude scales are being widely used for varied purposes. One can find lists of such tests and scales in books on Evaluation and Measurements, journals on Education and Psychology, Books and on teaching of various school subjects and various Mental Measurement Year Books. But these tests and scales as they are meant for different geographical regions and are based on different syllabi and objectives, are of little use. They can at best serve the purpose of good guides in the principles and techniques of test constructions and Evaluation. Since no useful purpose

* For references in this Chapter, please see Bibliography.
can be served by mentioning the name of those tests and scales here, the investigator confines himself to mentioning the few studies that can be of some interest in this field.

As far back as 1890, Edgeworth* found that the marks allotted independently by 28 different examiners to a piece of Latin prose varied from 45 to 100 per cent.

In the early years of twentieth century in U.S.A. experiments were carried out to find the personal element which enters into every examiner's evaluation of a script. Starch and Elliott (1913) sent copies of a single geometry paper to the chief geometry teachers of 116 high schools requesting them to mark it in accordance with their usual practice. The marks awarded varied all the way from 28% to 92%.

Wood (1921) relates an anecdote of the six papers which were marked by six examiners. The first examiner wrote out a set of model answers to the questions to serve as his own guide in assessing the papers. Unfortunately, the 'model answer' script was left among the candidates' papers. He was subsequently awarded marks ranging from 40% to 90% by the other examiners.

Boyd (1924) selected 26 scripts from amongst a large number of essays on "A Day at the Sea Side" written by 11 year old Scottish school children, and had them marked independently by 271 teachers. The scripts were to be graded on 7 possible grades. In one instance 3 teachers marked a script Excellent, 31 Very Good +, 80 Very Good, 125 Good +,
28 Good and 4 marked it Moderate. Of course, none marked it unsatisfactory. He showed that part of the discrepancy was due to different standards of expectancy, but even when the 20% most lenient and the 20% most severe markers were excluded, the average essay was still awarded at least 4 different grades out of the 7 grades.

Barton (1925) submitted a questionnaire to 1513 pupils to find their opinion regarding a number of things about marks. 50% of the students believed that marks were a fair estimate of what they had done, and 50% of them believed that they were not a fair estimate. 55% of students did not get the highest marks in subjects in which they studied the hardest. 37% believed that awarding of low marks during the first month was a bluff.

Ruch (1929) reports the marks assigned by 91 teachers who graded three answers to a question in Geography. Answer number one was graded from 2% to 20%, answer number two from 0% to 2% and answer number three from 3% to 20%.

Of all the enquiries about examinations, one that was conducted under the auspices of Carnegie Corporation, the Carnegie Foundations and International Institute of Teachers College, Columbia University, is perhaps the first genuine attempt to highlight the element of subjectivity, unreliability and invalidity of the essay examinations. An International Conference on Examinations was held in May, 1931 at Eastbourne (England). The countries represented at the Conference were England, France, Germany, Scotland, Switzerland and United
States. Several enquiry committees investigating individually in each of the above mentioned countries were set up to go into the problems of examination reforms. Each of these committees received a grant for three years from the Carnegie Corporation and each of them reported independently to a Second International Conference on Examinations held at Folkestone (England) in June 1955. Hartog and Rhodes summarised the reports of the English and the French Committees in 'An Examination of Examinations' (1935). The main theme of the studies by both the Committees was 'the part played by chance in the verdicts given at different public examinations.' The findings of the two Committees were sufficiently convincing to prove that 'the element of chance in examinations subsists to a dangerous degree in all the subjects including mathematics.' The English Committee while investigating on the Marking of School Certificate History Script, found that when 14 experienced examiners remarked independently 15 scripts which had all received the same moderate mark from the examining authority by which they were supplied, these examiners allotted over 40 different marks to several scripts. It was further found that when these examiners remarked once more the same scripts after an interval of from twelve to nineteen months, they changed their minds as to the verdict of Pass, fail and Credit in 92 cases out of the total 210 i.e. in 44% cases.

The French Committee in a study of the philosophy examinations in Baccalaureat (France) observed that when one hundred scripts were marked by six examiners, over 80%
were passed by some and were failed by others. The results suggested that the correlation between any pair of examiners averaged only + 0.60. When a set of papers in Science were marked by three trained scientists and also by a student who had no scientific training, on comparing the markings it was found that the average correlation coefficient between the scientists was + 0.63 whilst the correlation between the student and scientists averaged + 0.51. The study does cast some doubt on the value of an examiner's knowledge of his subject.

Valentine and Emmett (1932) carried out comprehensive studies into the predictive validity of essay examinations. Valentine found that roughly 40% of all those awarded scholarships, at the time of the special place examinations, failed to justify their promise, since they eventually obtained only third-class honours or pass degrees. They were beaten by as many as 33% of the non-scholars, who achieved first or second class honours.

Notable contributions to literature on the technique and principles of Test Construction and Standardisation have been made by Adkins (1947), Bean (1953), Burt (1921), Carroll (1945), Cronbach (1949), Davis (1947), (1951), Ebel (1951), Flanagan (1939), Freeman (1953), Guilford (1954), Guiliksen (1950), Hoyt (1941), Hull (1928), Jackson and Ferguson (1941), Jordan (1953), Kelley (1927, 1939), Kuder and Richardson (1937), Lindquist (1951), McCall (1922), Mosier, Meyers and Price (1945), Ross (1941), Ruch (1924, 1929), Symonds (1927), Thorndike (1904,
Thorndike (1949), Thorndike and Hagen (1955), Travers (1949), Tylor (1934) Vernon (1940) Wood (1923) etc.

It would not be of much avail to incorporate their contributions to this field here. The same would be mentioned at the appropriate places while dealing with particular problems in the development of the test.

Amongst the critical literature on the subject by Indian writers, the following few are worth mentioning:-

Hartog is perhaps the pioneer in the field of critical literature on examinations in India. For his works 'Examinations in their Bearing on National Efficiency' (1911), 'Examinations and their Relation to Culture and Efficiency' (1918) and his articles like 'Some Problems on Indian Education' (1920) and Examinations in India (1932) are the only early critical writings on examinations in India.

Paranjpe (1937) laid bare the unreliability of the traditional examinations with special reference to those of India in his book 'Reliability of Examinations'.

A case for the introduction of new-type tests in India was well prepared by Menzel in 1939 in 'Suggestions for the use of New Type Tests in India'. Besides emphasising the need of these tests in India, the writer discussed the advantages and limitations of these tests at full length.

Examinations in India (1944) by Mukerjee is an exposition of the essential features of examination reform in the country. The author has tried his best to press the point that 'no reform of education in India would be possible
without a root-and-branch reform of the current examination system.'

'Examinations in India' (their defects and remedies) 1951, by Salamatullah is one of the most comprehensive studies on examinations in India. The author undertakes the survey of education in India under the British and leads to the educational needs and aspirations of the people of free India. After analysing the present system of examinations, he advocates the development of objective tests in India as a solution to defects thus discovered.

'The subjectivity in examinations' is an interested study by Janak Batra (1951). She found that (1) marks varied from 21 to 77 in the case of one script, (2) the same examiner differed with himself after two months in the assignment of the position to the same group of students, correlation being .77, (3) the correlation between the essay type and an objective type test was only .48, (4) marks of the class teacher were quite different from those of the external examiners.

"An enquiry about examination (Calcutta University)" 1954 by Mukerjee is primarily a study of the matriculation examination of Calcutta University during 1912-45. The author has thrown a considerable amount of light on the entire problem of examinations for the guidance of the examining bodies like Universities and Boards. The investigator found that there was an appreciable but gradual decline in the quality of the results.

'A Study of the Examination System in India since 1835'
by Das (1957), traces the origin of the present examination system. By analysing the present system of examination, the study finds the system besides a severely defective technique of evaluation, further responsible for causing a host of educational and social problems of grave consequences.


National Council of Educational Research and Training (1963) published a list of 233 achievement tests constructed by students at M.Ed. and Ph.D. level in various school subjects. A scrutiny of the list shows that 69 of these tests are in the field of physical and biological sciences.


2- Bhargava, R.S. Construction of an Achievement Test in General Science for class VIII, M.Ed. Report - University of Rajasthan, 1956.


4- Dutt, N.K. Construction and Standardisation of an objective type achievement Test in General Science for class VIII. M.Ed. Report - Punjab University, 1958.


6- Nagar, B.L. Construction of an Achievement Test in General Science for class VIII, M.Ed. Report, Vikram University, 1958.


8- Parikh, R.C. Achievement Test in General Science for Standard VIII (Construction and Standardisation) M.Ed. Report, Baroda University, 1959.


10- Taneja, O.P. Construction of an Objective Type Achievement Test in General Science for the class VIII students of Delhi, M.Ed. Report, Delhi University, 1959.
The Attitude Scale:

In the field of Attitude Measurement, as early as 1925, Bogardus developed a technique for measuring attitudes towards different national groups. The items in his technique concern the desired social distance from a particular national group. He usually employed the following categories of social distance:

1- To close kinship
2- To my club as personal chums
3- To my street as neighbours
4- To employment in my occupation
5- To citizenship in my country
6- As visitors only in my country
7- Would exclude from my country, and so on

One of the earliest and still most widely used methods of constructing attitude scales was developed by Thurstone(1928) and his colleagues. The following major steps were employed by them in the construction of a scale:

1- The collection of a preliminary list of statements.
2- The evaluation of these statements by a group of judges
3- The elimination of ambiguous and irrelevant statements
4- The determination of scale values of the statements
5- The collection of normative data

The basic premises in the Thurstone method of scale construction are:

1- That a series of statements can be made to serve as the markers on a yard stick for the measurement of attitudes,
2- That each of these statements represents a specified degree of acceptance or rejection of a belief and 3- That these specified degrees of acceptance or rejection are equally spaced throughout the entire range of the attitude continuum.

Doubts have been often expressed regarding the personal attitude of the judges effecting the evaluation of the statements for an equal-appearing-interval attitude scale. Studies by Beyle (1932), Hinckley (1932), Ferguson (1935), Pintnor and Forlano (1937) and Eysenck and Crown (1949), and Hovland and Sherif (1952) proved that the attitude of the rater has little effect upon and is not related to his evaluation of the statements and that the assumption taken by Thurstone and Chave in their original monograph, that the evaluation of statements would prove to be independent of the attitude of the raters, is correct. However, the studies show that there are often significant differences in the absolute intensity of statements when judged by groups that hold different attitude. A prejudiced group and a non-prejudiced group may agree to the positiveness or negativeness of a statement, but they may differ in their judgements of how positive or negative the statement is.

Thurstone employed several hundred judges to evaluate the preliminary list of statements. Later studies by Nystrom (1933) Ferguson (1939), Rosander (1936), Uhrbrock (1934) and Edwards and Kenney (1946), however, indicate that reliable scale values can be obtained with much smaller groups of judges.
Uhrbrock (1934) reported a correlation of .99 between scale values for 279 statements obtained independently from two groups of 50 judges each.

Rosander (1936) obtained correlations as high as .99 for scale values determined independently from two groups with as few as 15 judges in each group.

Edwards and Kenney (1946) found a correlation of .95 between the scale values for 129 statements obtained from a group of 72 judges with the scale values for the same statements based upon the judgements of 300 judges.

Lasswell (1938) studied the importance of the emotional relationship of children and their parents in the development of certain attitudes. His analysis of revolutionary agitators suggested that such individuals are likely to have deep, unconscious hostility towards the father. Krout and Stagner (1939) and Stagner (1949) supported this theory by supplying statistical data.

Remmers (1934) attempted to extend the usefulness of the Thurstone technique by preparing a type of scale that can be used to measure attitudes towards a great many objects of some designated class of objects. Some of the most famous Equal-appearing-interval scales edited by him and his collaborators are:

1- Attitude towards any group
2- Attitude towards disciplinary procedure
3- Attitude towards home making activity
4- Attitude towards play
5- Attitude towards peem
6- Attitude towards practice
7- Attitude towards proposed social action
8- Attitude towards proposed social activity
9- Attitude towards school subject
10- Attitude towards social institution
11- Attitude towards teacher
12- Attitude towards training
13- Attitude towards vocation
In each case, the name of the group, social activity, school subject or poem etc., must be indicated. But the same statements are used for any member of the appropriate attitudinal group.

Seashore and Hevner (1933) showed experimentally that if the opinion statements are memographed (instead of writing on the cards or slips as recommended by Thurstone) and the judges are instructed to encircle a number from 1 to 9 (on a nine point scale) before each statement the scaling results agree substantially with those obtained by the more time consuming card sorting procedure of Thurstone.

Sigerfoos (1936) in a study found that arranging the items in ascending or descending order of the scale value and using the median scale value of the items endorsed as the individual's score gave results as valid as arranging the items in random order and using the arithmetical mean of the scale values of the items endorsed.

Likert (1932) developed another major technique for the measurement of attitudes. The main steps in the construction of a scale by Likert's technique are:

1- Collection of a list of statements
2- Editing of the statements
3- Giving the edited statements, constituting the attitude scale, to the subjects whose attitudes are to be measured
4- Determination of the number and percentage of subjects giving the alternative responses of each statement
5- Determination of the scoring weights for the alternative responses to each statement
6- Scoring of the scale
7- Securing item-consistency data
8- Elimination of 'Inconsistent Items', if any
9- Rescoring the revised scale
10- Preparation of Norms
On the final scale of the Likert type, the subject marks each statement in one of the categories of strongly agree (SA), agree (A), Undecided (U), disagree (D), and strongly disagree (SD). A score of 5 or 1 is given if a subject marks 'strongly agree' on a positive or negative statement, a score of 4 or 2 is given for marking 'agree' on a positive or negative items, and so on to a score of 1 or 5 for 'strongly disagree' on a positive or negative statement. The individual's final score is obtained by summing the item scores.

Peterson and Thurstone (1933) in a number of experiments in which children were given attitude tests and then were allowed to attend a performance of a certain movie picture, and were retested a few weeks later to see whether or not any change in attitude had resulted, found that the changes produced were large and persisted for long period of time.

Remmers (1934) and his students successfully demonstrated that the teaching materials taking up only about 15 minutes of class time may produce significant changes in the child's attitude towards various social problems and that these changes still persist after a period of one year.

Stagner (1936) demonstrated that political preference is an important determiner of attitude towards fascism. Republicans being most pro-fascist, Democrats average and third party groups being most anti-fascist. Brown (1936 -p529) argues that these institutional influences are the true determiners of attitude and presents a view that attitudes are merely reflections of the persons membership pattern.
Newcomb and Svehla (1937) in their investigations reported that the correlation of mothers and fathers with sons and daughters for attitude towards the church are all above .55, for attitude towards War, they are all above .40.

Studies by Ferguson (1941) and Edwards and Kenney (1946) have shown that the Likert technique of attitude scale construction does not obviate the need for a group of preliminary judges as claimed by Likert.

Farnsworth (1943) found that scaling of opinions may shift with time and changing cultural influence. He was able to detect the cultural effects by a comparison of mean scores. Certain opinions on war were judged more pacifistic in 1940-41 than they had been some 10 years earlier.

Ferguson in another study found that the group status as a factor in the evaluation of statements could be disregarded. He found that the ego-involved assistant manager - group, the superior status group of managers and the inferior status group of agents, all rated the statements from which an assistant managerial evaluation form was to be constructed, in the same way.

Crespi (1944) developed a scale for measuring attitudes towards 'Conscientious Objectors' to tap extreme negative feelings, as there are various degrees of dislike beyond wanting to have no social contact with a group. His scale employed statements like the following:

1- I would accept conscientious objectors only so far as having them for speaking acquaintances

2- I don't want anything to do with conscientious objectors
3- I feel that conscientious objectors should be shot as traitors, and the like.

Edwards and Kenney (1945) found that construction of Thurstone scales required about twice as much time, exclusive of the time spent by the judging group in sorting the items, as did the Likert Scales.

Barclay and Weaver (1962) found that the construction of two alternate forms of a Thurstone Scale of 21 items consumed 43.2% more time than construction of two alternate forms of a Likert Scale of 21 items. Whereas Murphy and Likert (1937) claim that the Thurstone Scales are ‘exceedingly laborious’ to construct, Bird (1940 – p. 160-161) feels that in the interest of constructing refined measuring instrument time is not an important factor. He remarks that:

"The claim of greater or lesser laboriousness seems to have been put forward without due regard for all processes in scaling techniques; but in the interest of constructing refined measuring instruments, time can be neglected. There is much to be said in favour of a psychologist’s refining his instrument before actually applying it to experimental group. The argument that the method of summated ratings is less laborious limps badly".

Vernon (1958 ~ p. 154) also feels that though the Thurstone method is laborious, it does yield scales which 'go down' well with most subjects and the fact that they need check only a few statements instead of giving a graded response to every item, is a greater advantage.

During Second World War, Guttman developed an interesting attitude scaling procedure in connection with studies of the
The objective of the method is to test directly whether a collection of items can be scaled on one attitude continuum. The criterion of scalability is that if a subject endorses a more extreme item, he should endorse all less extreme items also. The response pattern found in the perfect Guttman Scale is exactly what is obtained if people are rank ordered on a physical continuum. The principal practical drawback to the use of the Guttman Scaling procedure is that almost no collection of statements will meet the strict lawfulness of the Scalability criterion. In the few cases, where this scalability criterion has been approximately met, the items prove to be so closely related in content as to constitute near rewording of the same statement.

In India only the following few attempts have been made by students at M.Ed. level to construct attitude scales for various purposes:

1- Chaudhury, D.C., Construction of an attitude scale to measure attitude towards Basic Education. M.Ed. Report, Utkal University, 1959.

2- Mahalanebish B.L.(Smt), The attitude of College Girls towards Teaching as a career (Construction of the Scale only) - M.Ed. Report, Delhi University, 1957.


4- Sarojini (Kum) An Attempt to Construct an attitude scale to measure attitude towards Teacher Training by the Thurstone Method (Hindi), M.Ed. Report, Gorakhpur University, 1959

Besides this efforts have been made by Dorai (1963), Maheswari (1961), Pawar (1952), Purandare (1961), Riar (1960), Sandhu (1960), Swan (1950), Verma (1961), and Vimla (1950) to find the correlation between attitude towards a school subject and attainment in the subject. But none of these have attempted to find correlation between the attitude of students towards learning of a subject and achievement in that subject.