CHAPTER -II

Achievement Tests.

Achievement Tests - A Brief Survey:

Although the measurement of achievement in Schools, is traced back to many centuries, the objective measurement dates, as already referred to, from the works of - Thorndike and his pupils. In 1918 Thorndike published a paper, the first of its kind, on measurement. It began with the dictum "Every thing that exists, exists in some amount, and that, if it exists in some amount, it can be measured". It counteracted the opinion of those who believed that ideas and emotions could not be measured.
The spelling and arithmetic tests of J.M. Rice, the Arithmetic tests of Stone, Elementary test of Courtis, Thorndikes Handwriting scale, Ayres Handwriting scale, Hellogas Composition scale, spelling scale of Buckingham and Starch's Reading Test gave the incentive to objective measurement of Achievement.

In 1911 Courtis\textsuperscript{3} reported the results of his test in the fundamentals of arithmetic, administered in the - Detroit School. The report attracted so much attention that Courtis was asked to participate in the New York survey and administer his test to 30,000 pupils. The report

\begin{enumerate}
\item Ballard: "Mental Measurements, page 2.
\item Ibid, page
\end{enumerate}
of the use of tests stimulated much interest among heads of institutions who were eager to compare the achievement of their pupils with the pupils in other schools. Curtis laid emphasis on the importance of norms. Thanks to his valuable efforts, a bureau was established in a number of States to cooperate with school men in administering tests and interpreting test scores. These endeavours led to the popularity of Achievement tests. From 1917 - 28 nearly 1,300 standardised and semi-standardised tests were developed. Surveying the advance the test movement was making, Buckingham observed that the "test making passed from an amateur to a professional basis". Walter Monroe in his book "Ten Years of Educational Research 1918-27", writes that the pioneer State of Educational Research is passed and quantitative production has been achieved. By 1927 there were developments in the new directions and this indicated a distinct advance. The early standard tests of Achievement were purely of a general or survey type. They were a measure of a pupil's attainment in a subject. They did not give detailed information required for remedial work. These tests gave way to tests of a specific type like diagnostic tests which give specific information regarding the pupils' strong and weak points. The tests were later organised into batteries consisting of survey tests in im-

2. W. Ross, "Measurement in To-days' Schools", page 43.
important school subjects in a single book-lets. In 1920
two such batteries appeared - one by Pinter and the other
by Monroe and Buckingham. In 1922 appeared the Stanford
Achievement Test. In 1929 the State University of Iowa
inaugurated a State-wide Every Pupil Testing Programme.
An innovation in Achievement testing was made by Lind-
ququist in 1942 with the Iowa test of educational develop-
ment. These tests, designed to measure the basic skills;
permanent learnings and attainment of other objectives of
education, were substituted for the strictly subject Cen-
tered tests of the original Iowa Every Pupil Testing Pro-
grame.

Achievement Tests in Science:

A good many achievement tests in science have been
standardized by various people in America and elsewhere.
A brief mention of a few may not be out of place here.

I Stanford Achievement Test.

The format, lay out, spacing and size of type of the
stanford Achievement Test are commendable. This test is
found in many forms. Each form consists of seventy items.
A great majority of items involves factual information.
There is no item which involves judgement and reasoning.
All the items are of multiple choice type. Many of these
are incomplete statements except a few direct questions.
Each item has three possible answers instead of four or five. There are also some defects. An item when completed correctly reads thus: "Humus is soil found chiefly under trees". The framing of this item is not correct because humus is one of the three constituents of soil, whether under trees or not.

II Co-operative Biology Test for High Schools:

This test is constructed by cooperative Test Division, Educational testing service. The test consists of 5 forms. Form Q consists of 120 items. There are five choices given for each item. About one third of the items test biological information. The remaining two thirds are basically factual. Form S is composed of 111 items. About 25 per cent of the items test the application of biological information and the interpretation of materials. The instructions are clear and simple. Data regarding test objectives, construction and validity studies are inadequate. The manual does not say how the questions were selected, what the item difficulty and validity indices are, on which population was the testing data obtained etc. It does not say also how it was validated and how its reliability was determined.


This is constructed for use at the end of a course
in high school. There are two forms of the test. Each of these forms consists of 75 items of multiple choice. The cover page contains directions and samples. The get up of the test is commendable. The questions are clear, direct and free from ambiguity. Importance is given to the understanding of Scientific phenomena and the ability to apply knowledge in the interpretation of situations and solution of problems. The reliability of the test has been calculated by the split half method. It has a reliability of 0.88.

Cooperative General Science Test:

This is constructed by Paul E Kambly and Carl A Pearson for the Cooperative Test Division Educational Testing Service. This test has been criticised by S. Richardson, Associate Professor of Education of the Ohio University, Columbia. This test is used as a measure of achievement at the end of course. The number of items in each form ranges from 85 in Form Z to 101 in form Q. With the exception of a few matching items in Form Q, all items are of multiple choice type. Each form is divided into two parts with 20 minutes allowed for the first part and 15 minutes for the second part. There is no apparent order in the listing of the items. There is considerable dependance on the re-call of factual information. To a very small extent principles of science are involved in
determining the correct response.

In item 34 of Form X, the key indicated that the correct answer is "water is a good conductor of electricity". Actually water is a poor conductor till some acid is added to it.

**Anderson Chemistry Test:**

This test is constructed by Kenneth Anderson. The working time for this test is 40 minutes. Directions and samples worked are found on the covering page. The test is in two forms. Each form consists of 80 questions of the multiple choice type, with five choices in each. This is a test not only of factual knowledge but also of the students' understanding of principles, familiarity with laboratory work, and reasoning ability.

**Item 47** in form A reads "Hydrogen gas was passed into a test tube containing copper oxide. Heat was then applied to the tube."

**Item 48** which one of the following statements - gives the principle that test explains the answer to item 47? - (6) Copper is comparatively an inactive metal (7) Metallic oxides often combine with an acid to form a carbonate. (8) Many oxygen compounds when heated with hydrogen with release oxygen. (9) Noticeable heat and light occur in many chemical reactions. (10) Some chemical com-
Pounds are able to absorb water from air. Answer 8 is considered correct but Riemann considers that item No. 6 also as an equally good answer. Though Riemann, (Professor of analytical Chemistry, Rutger University, New Jersey) has criticised item 48 in Form A (given below), it does not seem to be valid when we take into consideration the knowledge of students as per the syllabus. 

**Read General Science Test:**

In this test there are 75 items. The time is 40 minutes. The distribution of items by content areas is; Physics 42 per cent; biology 28 per cent; chemistry 4 per cent, and general 26 per cent. The sampling of the areas does not allow of adequate diagnosis of academic achievement. The wording is clear and exact. There is considerable reliance on the re-call of factual information and utilization of principles of science. Benjamin S. Bloom has highly commended this test from a statisticians point of view. Preliminary forms of the test were administered to 1,600 students in 12 high schools in 8 States near the close of the year. The mean IQ of the group was 102. The mean validity index (item total correlations) was computed and found to be .42 and .43 for the two forms. Items were selected so as to yield two forms of the test precisely balanced in difficulty of 53 per cent and composed of items known to be of significant discriminating power. The reliability coefficients
Objectives and content to be tested were drawn from an examination of widely used text books, state curricula, and samples of courses of study.

Even in India, as already referred to, attempts have been made in recent years to construct and standardize achievement tests in science. These have not been popularized. Some of the tests are in the regional languages, as for example, the one constructed and standardized by M.S. Chandavarkar, B.A., B.Sc., M.Ed., of the Kamatick University under the guidance of Dr. V.V. Kamat is in Kannada; and the other, an M.Ed. dissertation in general science submitted by Sri H.H. Pawar, B.Sc., LL.B., B.T., is in Hindi. The Psychological laboratory, Maha­­raja’s College, Mysore constructed an Achievement Test in Elementary Science for V Form some years back. The test consists of 60 items of the Multiple choice type. Each item has four responses. This test has not been standardized. The test constructed by Sri M.S. Chandravarkar and published by Macmillan and Co., is intended to measure their knowledge of children between the ages of 8 and 18 in General Science in Bombay-Karnatak. It consists of 100 items. There are 20 items of the simple Re-call type, 18 items of the True False type, 15 items of the Association type, 15 items of the Alternative Response, 15 items
of the matching type, and 15 Alternative Response type. This was administered to 4,626 pupils. Norms, reliability and validity have been found out. Incidentally he has found out also the interest of children in the various branches of science.

The reliability of the test has been calculated by the split-half method. It is 0.89. The validity of the test has been calculated by comparing test scores with the examination marks. The co-efficient of correlation for various standards has been calculated. It lies between 0.37 and 0.66.

Sri Jagannath, B.Sc., B.T., M.Ed., Head Master, Sri Rama Krishna Vidyalaya has constructed in part fulfilment of his M.Ed. Course in the Institute of Education, Delhi, an achievement test in Physics, but it has not been standardized. He tried it on 200 students of the XI Standard of the city of Delhi. The test proposes to measure principles and basic concepts in Electricity and its everyday applications. There are 25 items of M.P. choice type, 35 items of T.F. type, 29 items of problem and completion and 5 diagrams. He has found out the norms, reliability and validity of the test. The reliability is 0.65 and validity is 0.71.

Sri H.Venkataramaiah, B.Sc., B.T., M.Ed., Lecturer, Teachers' College, Mysore standardized an achievement test
in 1958 in General science Part I (Physics and Chemistry) for High School III student of Mysore State. The test consists of 30 items of T.F. type, 16 items of Multiple choice type, 20 items of the matching type, 18 items of the completion type, 7 items of the reasoning type and two diagram tests. He has found out the norms, reliability and validity of the test. The test has high norms of 5.4.99 and it is highly reliable and valid. It was administered to 2000 students selected from various strata of society. The investigator has given the minutest details regarding the standardisation of a test in a very lucid manner. However, it would have been better if he had the test items on Biology also. This would have given a complete picture of the attainment of the boys in General Science as a whole.

Nature of Achievement Test.

The attainment or acquired ability of a pupil in a subject of study is termed Achievement. The Achievement test measures the ability to do use or understand a certain thing based upon the knowledge, method, attitude, interests and skills a pupil acquires from instruction or experience. The kind of achievement to be measured depends upon the nature of the subject and the
various objectives of its teaching or learning.

There are two types of Achievement tests, General or Survey tests and Diagnostic tests. "A General achievement test is one designed to express in terms of a single score a pupil's relative achievement in a given field of achievement". A diagnostic test on the other hand is designed to diagnose the deficiencies in learning and teaching in one or more areas of achievement. A battery of tests consists of different kinds of tests and each kind tries to find out the specific weakness, errors or gaps in attainment in a field of achievement, as for example a pupil's weakness in acquisition of knowledge, use of knowledge, (problem solving capacity) skill in performing experiments. Each of these tests is of the general achievement type and the whole test is diagnostic.

An achievement test may be constructed by a single teacher or a team of teachers for use in a class room. These are called informal tests. When a test is constructed carefully, systematically and scientifically in consultation with experts and experienced teachers and when it is tried and retried on a wide sample in a natural homely atmosphere and when certain characteristics, such as norms, validity, reliability, usability and predictability are determined, it is called a standardized test.

various objectives of its teaching or learning.

There are two types of achievement tests, General or survey tests and Diagnostic tests. "A General achievement test is one designed to express in terms of a single score a pupil's relative achievement in a given field of achievement. A diagnostic test on the other hand is designed to diagnose the deficiencies in learning and teaching in one or more areas of achievement. A battery of tests consists of different kinds of tests and each kind tries to find out the specific weakness, errors or gaps in attainment in a field of achievement, as for example a pupil's weakness in acquisition of knowledge, use of knowledge, (problem solving capacity) skill in performing experiments. Each of these tests is of the general achievement type and the whole test is diagnostic.

An achievement test may be constructed by a single teacher or a team of teachers for use in a class room. These are called informal tests. When a test is constructed carefully, systematically and scientifically in consultation with experts and experienced teachers and when it is tried and retried on a wide sample in a natural homely atmosphere and when certain characteristics, such as norms, validity, reliability, usability and predictability are determined, it is called a standardized test.

A standardized test has a wider objective than an informal test.

**Functions of Achievement Test:**

The value of a testing programme depends upon the extent to which its results have been helpful in improving learning, instruction, guidance and administrative practices in schools. The achievement test, an important type of measurement, has all the useful functions referred to above.

**Influence On Learning:**

An achievement test clearly indicates the type of learning or study procedures adopted by students. Even though Richardson and Stalanker are opposed to the belief that an achievement test would always have a pedagogic value, this should not be an obstacle in finding out the influence of examinations on learning. Meyer reported two studies on the type of examinations and memory, and type of examination and study procedures employed by students. The findings indicated that the type of examination does influence the nature of students' study. Johnson reported that testing stimulates learning and Spitzer concluded that testing helps in the retention of learning.

We may therefore conclude that measurement influences

2. Ibid, page 411.
learning.

**Measurement and Instruction.**

Any teacher is keen about the effects of his teaching. This he realises by the use of objective tests which throw some light both on the capacity of his students and his methods of teaching. The teacher would be able to gauge the good or bad points in his teaching methods. With the help of the test results he would also adopt the necessary remedial measures to achieve the necessary objective of the subjects taught.

**Measurement and Administration:**

The healthy running and progress of an institution on the right lines is the responsibility of the administrator. It is his duty to see that proper methods of instruction are employed and the pupils achieve the various objectives of instruction. This can best be evaluated by the use of well constructed objective tests. These tests indicate not only the achievement level of the students, but also diagnose whether his suggestions if any have been followed or not.

**Achievement Tests and Counselling:**

Counselling, as mentioned by John Darby,¹ "is the process in which information about the individual and about

---

¹ John Darby, "Educational Measurement", page 68.
his environment is organized and reviewed in such a way as to aid him in reaching workable solutions to a variety of adjustment problems in the normal range of behaviour.

School administrators are highly criticised for not identifying the abilities of pupils and for not directing them away from the fields in which they do not seem to have an aptitude. This can be easily overcome by the use of the new objective devices. The evaluation records of pupils show their worth and with the help of these - their energies can be directed towards right ends. They can be counselled about the vocation or career they should choose in accordance with their attainments of certain - skills, attitudes and concepts. They can also be guided with regard to the course of study they must choose for entry into a particular type of job or vocation.

**Achievement Tests and Educational Placement:**

Placing pupils who are normal for their group under the same group is a proper method of placement. Pupils alike in their chronological age, educational achievement, physiological, mental and social development must be placed under one group. The result of objective measurement can be used in determining the pupils placement in the group. Through the development of reliable grade and age norms based upon the achievement of groups of children in standardized tests, a valuable instrument for establishing a
grade line is made available. Achievement of children in standardized tests is made use of in admission into various kinds of colleges, in awarding scholarships and in the selection of personnel in the various branches of social activities. For example, the army alpha test was used for the purpose of selection of officers to man the armed forces.

**Principles for Achievement Test Construction:**

The objective test construction, as any other work must be based on certain cardinal principles. The constructor should know the purpose his test should achieve. He must know what to measure and having fully and clearly understood this, he must think of the ways and means of measuring it. If he should know what to measure he must have a clear knowledge of the objectives of teaching and learning a particular course of study. It is not enough if he knows the broad objectives of such a course of study. He must have a clear insight into the required modification in behaviour such a study is likely to bring. Lindquist in this connection observes that "if measurement is to continue to play an increasingly important role in education, measurement workers must be much more than technicians unless their efforts are directed by a sound educational philosophy. Unless they accept and welcome a greater share of responsibility for the selection and classification of educational objectives and unless they show much
more concern with what they measure as well as how they measure it, much of their work will go in vain. Educational measurements have a great influence on the educational process and its improvement. The ability of the test constructor to know the situations in which tests are essential and useful or their influence on educational practices is as important as the ability in test construction. His test must be valid and dependable. What his test measurement must be of importance and significance and its use must exercise a desirable influence upon the aims, habits, attitudes and achievements of students, teachers, counsellors and school administrators. The first and foremost thing a test constructor must do is to formulate the objectives of his test or what it proposes to measure. This implies what a particular field of study proposes to achieve. The objectives of teaching and measuring are interdependent. What then are the sources of getting the general objectives of teaching and learning a particular subject? According to Walter Monroe, there are four sources of formulating the objectives.

1) An analysis of the text book on the subject,
2) Analysis of the courses of study,
3) Analysis of teacher made tests,
4) Opinion of experts on the subject.

1. Lindquist: "Educational Measurement", page 158
The last of these sources is the most important one. Text books, courses of study, examinations and teacher made tests - all depend on expert opinion.

If text books are not written with due care taking into account the purpose of such a venture, examinations or tests constructed on the basis of these text books measure just what is taught and not what should be taught. They do not take into account the modification in the behaviour of the pupil resulting from such instruction.

Objectives resulting from an analysis of the courses of study have an advantage over the previous type as they attach more importance to application of facts, principles and generalizations than factual knowledge.

An analysis of the teacher made tests gives us an idea of the opinion of a number of teachers about the importance to be given to a topic of study in a given field. Unless these tests keep in view the why and how of things, they are not helpful.

Dr. Benjamin S. Bloom, the college examiner of the Chicago University and expert on evaluation, conducted a number of evaluation workshops while in India in the year 1958.

He suggests - that a teacher or a test constructor
should first define the objectives of teaching a particular subject and should break down these into proper changes of behaviour expected in each pupil. To effect these changes of behaviour, learning situations have to be provided. After this comes the stage of test construction.

This means that an achievement test constructor focuses his critical attention on the objectives of education. This results in a systematic analysis, classification and restatement of the educational objectives. Mere informational tests have been highly criticised by experts on measurement. Lindquist states "Good testing, as well as good teaching should penalize rote learning rather than place a premium upon it. A good test in this respect is one in which, among other things the constructor has assiduously avoided the use of textbook language or of stereo-typed and catch phrases or pat verbalization likely to be acquired by the rote learner".¹

The Iowa tests of Educational development constructed by Lindquist have the following titles for the individual tests (1) Ability to do quantitative thinking, (2) Ability to interpret Reading material in the social studies, (3) Ability to interpret reading material in the natural sciences, (4) Ability to interpret literacy material, (5) Correctness in writing, (6) Understanding of the basic

social concepts, (7) Background in the natural sciences, (8) General Vocabulary, (9) Use of sources of information. These titles give us an idea of the important educational objectives and the types of tests in an overall evaluation programme.

Some testing techniques.

It is quite beyond the scope of this work to have a complete and exhaustive discussion of the various types of tests. However, it is worth devoting a few pages to a brief treatment of these so as to give an idea about their uses and limitations, and the various pitfalls that are to be carefully guarded against in constructing the test items.

1) Alternate Response:

This is one of the most popular forms of the simple recall type. Alternate response items are those in which only two alternatives are presented to the pupil for his response. More often than not, it takes the form of True false statements requiring the pupil to establish the correctness or otherwise of the given statement. Sometimes it requires the pupil to mark merely a 'Yes' or 'No' against each of the statements and sometimes involves the selection of the correct or the better one of the two responses presented as possible comparison in a given —
situation. However the 'true - false' reigns supreme as the most widely used alternate response type. It involves a very simple method of response in the aligned answer positions of the test items. It is widely applicable in all subjects fields. It is generally believed to be easy to construct, however in actual practice, the elimination of ambiguities is often difficult, if not impossible, to accomplish. Yet, they can be used satisfactorily in many situations if they are constructed carefully enough to keep them free from ambiguity. They have the advantage of affording a wide coverage in a short period of time and serves a very useful purpose in the measurement of a functional type of instructional outcome. On the other hand guessing is more of a problem for this than for any other type and hence of little diagnostic value.

The following suggestions may be borne in mind while constructing the alternate response items.-

1) Avoid double negative statements for they serve no useful purpose and are often likely to needlessly confuse the pupil.

2) Do not use statements that are partly true and partly false. They add nothing to the test and intentionally or unintentionally take the form of 'catch' items.
3) Do not use specific determiners such as always, never, none, only, nothing etc. except with great care and as sparingly as possible.

4) Require answers in a simple but highly objective form.

5) Have a random distribution of true and false items and let there be no definite proportion of true and false items.

6) Make the crucial element in the question as obvious as possible to avoid ambiguity.

7) Avoid the use of textbook language.

8) Avoid the use of general terms such as large, small, great, well-known, many, few etc., when fine distinctions are involved, or when the meaning is not obvious.

Completion:

This is another form of the simple recall type and very commonly used. A completion item typically consists of a sentence or a paragraph from which key words have been omitted and blank spaces provided. The pupil is required to complete the statement by placing the proper words in the blank spaces. This is applicable to almost all achievement fields, easy to prepare and is likely to encourage more thorough study habits. On the other hand
it is over rated from the stand point of requiring understanding rather thanrote memory, not highly objective unless great care is taken in construction and scoring, subject to over mutilation and time consuming.

The following suggestions are offered for the construction of completion items.

1) Make each blank call for the completion of a single idea.
2) Avoid too many blanks.
3) Make all blanks the same lengths to avoid giving clues.
4) Avoid text book wording.
5) Avoid clues afforded by the requirement of grammatical consistency.
6) Make the statement sufficiently complete to enable the pupil to interpret the item correctly.
7) Omit only the key word or words.
8) Provide positions for responses ordinarily, at the end of the sentence.

Multiple Choice.
The multiple choice type is a form of the recognition item type and has come to be 'the most popular form for standardized testing of recent years; and increasingly coming into wide use for informal objective testing as well'. A multiple choice item usually consists of an
incomplete statement followed by from three to five plausible alternatives that will complete the statement. The pupil is expected to choose the correct or the best response and to indicate his choice by an answer appearing in a column at the left or the right side of the test paper. It may be in the form of a question rather than a statement or may consists of three to five words, symbols or numbers from which the correct or the best one is to be chosen by the pupil. The multiple choice and its various forms are widely adaptable to different types of content as is the case with true false type. It is readily adaptable to the measurement of "discriminative power, inferential reasoning, interpretative ability, reasoned understanding, generalising ability and other types of outcomes deriving from the pupils ability to apply and use facts". But multiple choice items are not as easily constructed as some other types of tests because of the various technical problems involved requiring great care in the drafting of items.

Some of the important points to be remembered while constructing multiple choice items, are:

1. Use only one form of multiple choice in the same section of the test.
2. Use at least four or five possible responses in order to minimise chance successes.
3) Do not mix items with varying number of possible responses in the same test if the scores are to be corrected for guessing.

4) Make the alternatives plausible.

5) Have the alternative answers at the end of the statement.

6) Do not ordinarily use "a" or "an" to introduce the alternative answers.

7) Distribute the correct responses with approximate equality among possible answer positions.

8) Use a random occurrence of the correct responses.

9) Require the answers in a highly objective form.

**Matching.**

Matching exercises may be considered as combinations of multiple choice items in such a manner that the choices are compound in manner. This type of test requires the matching of items placed in two or more columns. To prevent guessing extra items may be placed in the response column. The pupil is usually required to write the number of the matching item in a space provided for it.

Matching exercises are widely applicable, easy to construct, fairly free from guessing and economical of space and time. But they are not suitable for measuring reasoning, understanding and judgement. They are suscep
to clues and more adapted to measuring factual memory.

The following suggestions will be helpful in the construction of the matching type.

1) Provide only one correct matching for each item and make the items mutually exclusive.

2) Take care to see that there is consistency of grammatical form.

3) Maintain the consistency of classifications. Each of the two lists should contain items that are of the same category.

4) Make the matching sets neither too short nor too long from ten to fifteen pairings are one probably optimum for balanced matching groups.

5) Arrange the items in random order in each test.

6) Have all the matching items on the same page.

7) Require the answers in a highly objective form.

Classification Test.-

To a casual observer, this seems to be very much like the multiple choice. In fact, most of the rules which apply to the construction of multiple choice items apply to this type of test as well. But a careful examination of the mental processes involved reveal that it is quite different from the multiple choice. The
classification test requires the pupil to discover a common relationship among four out of five words, which is not shared by the fifth which is to be eliminated from the group. This form of test is suitable for measuring the capacity for reasoning and judgement.

**Analogies**

This type of test involves to a high degree, and rational thinking, reasoning/judgement and the education of relationships and correlates. It consists of three words, two of which bear a definite relationship themselves. Not only the pupil is required to discover the relationship that exists between the two, but has also to find the correlate for the third. It discourages, as in the case of classification, rate learning and guessing and puts a premium on effective thinking.

**Enumeration**

Enumeration is a form of the simple recall type in which the candidate is asked to enumerate properties, uses, and such other things. Enumeration test usually requires from three to five responses. This type of test is not much reliable as it lacks objectivity. However it is easy to construct and can be used to test factual knowledge.
Diagram Test.

Diagram test items are based on pictures and graphical representation. But they are not commonly used owing to various reasons. Good pictorial items can be used to measure useful skills, detection of defects and critical observation.