

**Institute of Distance and Open Learning
Gauhati University**

**MA in Education
Semester II**

**Paper 202
MEASUREMENT AND EVALUATION
IN EDUCATION**



Contents:

- Unit 1 : Scales of Measurement**
- Unit 2 : Test Construction and its Standardization**
- Unit 3 : Measurement of General Intelligence**
- Unit 4 : Aptitude Tests**
- Unit 5 : Educational Achievement and its Measurement**

Contributors:

Dr. Archana Adhikary Assistant Professor, Dept. of Education Tezpur University	Unit-1
Dr. Nabanita Sarmah Assistant Professor, Dept. of Education Kaliabor College, Nagaon	Unit-2
Dr. Utpal Kalita Assistant Professor, Dept. of Education R.G. Baruah College, Guwahati	Unit-3
Dr. Marami Goswami Associate Professor, Dept. of Education Guwahati College	Unit-4
Dr. Pradip Mochahary Assistant Professor, Dept. of Education Kaliabor College, Nagaon	Unit-5

Course Coordination :

Prof. Dandadhar Sarma	Director, i/c IDOL, Gauhati University
Prof. Polee Saikia	Professor, Dept. of Education, G.U.
Dr. Monoj Kr. Nayak	Asstt. Professor in Education, GU IDOL

Content Editor:

Prof. Kaberi Saha (Units: 2 & 3)	Professor, Dept. of Education, G.U.
Dr. Purabi Baishya (Units: 1 & 5)	Asst. Professor, Dept. of Education, G.U.
Dr. Nil Ratan Roy (Unit: 4)	HoD & Associate Professor, Dept. of Education, Tezpur University

Format Editor:

Dipankar Saikia	Editor, SLM, IDOL, Gauhati University
-----------------	---------------------------------------

Language Editor:

Prof. Anjali Daimari (Unit: 1)	Dept. of English, GU
Dr. Farddina Hussain (Unit: 2)	Dept. of English, GU
Dr. Nivedita Malini Barua (Unit: 3)	Asstt. Prof., Dept. of ELT, GU
Dr. Rajkumari Sujata Singha (Unit: 4)	Asstt. Prof., Dept. of English, Guwahati College
Prof. Padmini Bhuyan Boruah (Unit: 5)	Dept. of ELT, GU

Cover Page Designing:

Bhaskar Jyoti Goswami	IDOL, Gauhati University
-----------------------	--------------------------

ISBN: 978-93-84018-88-7

June, 2019 Reprint - June, 2020

© Copyright by IDOL, Gauhati University. All rights reserved. No part of this work may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise. Published on behalf of Institute of Distance and Open Learning, Gauhati University by the Director, Gauhati University Institute of Distance and Open Learning and printed at Gauhati University Press, Guwahati-781014.

SYLLABUS

202: Measurement and Evaluation in Education

- Unit 1 : Scales of Measurement:** Measurement in Physical and Social Sciences, Nature of Educational Measurement and Evaluation.
- Unit 2 : Test Construction and its Standardization:** Meaning and Nature, Principles and Methods of Test construction and standardization, Selection of subject matter, Item writing. Administration, Analysis and Selection of Test Items, Methods for Estimating Test Reliability, Test validation Procedures. Test Scores and Scales, Norms and their uses.
- Unit 3 : Measurement of General Intelligence:** Binet-Simon Tests, Weschler Tests of intelligence, Age scale and Point scale. Uses of General Intelligence Tests in Education.
- Unit 4 : Aptitude Tests:** Attitude Tests, Test of Interest, personality Tests through Subjective, Objective and Projective techniques, Test of Creativity.
- Unit 5 : Educational Achievement and its Measurement:** Objectives of Achievement Tests, Construction of Achievement Test, Diagnostic Test and Performance Test, Essay and Objective Type Tests.

Reference & Suggested Readings :

1. Bipin Asthana & R.N. Agarwal : *Measurement & Evaluation in Psychology and Education*
2. M. Goswami : *Measurement & Evaluation in Psychology and Education*

Unit: 1

SCALES OF MEASUREMENT

Content:

- 1.1: Introduction
- 1.2: Objectives
- 1.3: Meaning of Measurement
- 1.4: Measurement in Physical and Social Sciences
 - 1.4.1: Difference between measurement in Physical and Social Sciences
- 1.5: Scales of Measurement
- 1.6: Meaning of Educational Measurement and Evaluation.
 - 1.6.1: Nature of Educational Evaluation.
 - 1.6.2: Relationship between Measurement and Evaluation.
- 1.7: Summing Up
- 1.8: Reference and Suggested Readings

1.1: Introduction:

We often come across the concept of Measurement. It is an integral part of our living as in our daily life, we are involved in a number of activities which include measurement of one thing or the other, like-measurement of the blood-pressure, measurement of our weight, measurement of our electricity, time, water and so on. It can be said that in order to attain accuracy of an effort in our daily life or to compare objects, events and phenomena, measurement is carried out everywhere. We measure height, weight or area by using a universally acceptable scale of Centimeter, inch, yard, feet etc. and every one of us arrive almost to the same conclusion. All these measurement indicates the physical measurement, but it is not just confined to the physical world. In the field of education and psychology too, measurement is carried out. This unit will help you to know about measurement in the field of education and psychology.

1.2: Objectives:

After going through this unit, you will be able--

- *to know* the meaning of measurement
- *to know* about different scales of measurement
- *to explore* the measurement in physical and social science
- *to know* the meaning and nature of educational measurement and evaluation
- *to explore* the relationship between evaluation and measurement

1.3: Meaning of Measurement:

Measurement is the act of measuring something or someone in terms of unit. Measurement implies the act of determining the quantity of a property in terms of a suitable unit. The Sanskrit term of measurement is "Mapan" which clarifies its meaning of measuring. It provides the quantitative description of a thing or object or any activity. Thus, it is quantitative in nature. It is a process of quantifying something. It is a process of assigning status to something. It permits objective description as well as facilitates comparison. It is the means of comparison. It finds out the smallness and bigness or how much or how less of properties. In a very simple way, you can perceive measurement as the mechanism which answers the question- 'how much?'. One of the important facts about measurement is that it is always based on the result of tests.

STOP TO CONSIDER

- Measurement means a process of assigning status to something, permitting objective description as well as facilitating comparison. It is a means of comparison.

Definitions of Measurement:

The definition put forth by different educationists and psychologists will help you to comprehend the meaning of measurement-

According to **J.P.Guilford** , "Measurement means description of data in terms of number and this is in turn means taking advantage of the many benefits that operation with numbers and mathematical thinking provides."

In the words of James M. Bradfield, "Measurement is the process of assigning symbols to dimensions or phenomena in order to characterize the status of a phenomenon as precisely as possible".

Again according to Stevens, measurement is the assignment of numerals to object or events according to rules.

Basic Functions of Measurement:

The basic functions of measurement are -

- To assign some numerals to some object or phenomenon
- To determine quantitative aspects
- To provide status
- To compare.

SELF ASKING QUESTIONS

1. What are the primary functions of Measurement?

1.4: Measurement in Physical and Social Science:

The above mentioned concept and definition will certainly help you to grasp the idea about measurement. But a question that may arise in your mind here is whether it is possible to measure human mind like any other physical object. The answer to your question is yes but only to a certain extent. The concept of measuring human mind indicates the concept of measurement in psychology. Mind is measured in terms of the behaviour, i.e. the subject matter of Psychology. This will lead you to the concept of types of measurement.

It is a well known fact that measurement enters into all fields of human activity and all branches of Science. We know that science is concerned with knowledge. It refers to the body of knowledge systematically arranged. It is humanly impossible for any individual to master the entire knowledge with all its complexity and diversity. There are different sciences to deal with different branches of knowledge. The sciences are of two kinds-Physical science (Mathematics, Physics, Chemistry, Geology, Geography, Astronomy etc) and Social Science (History, Political science, Economics, Sociology, Anthropology, Psychology etc).

Measurement used to measure the different physical phenomena in different physical science and biological science is known as Physical Measurement. On the other hand, the measurement used to measure different humanistic and behavioristic phenomena in different social science and in applied science is known as Psychological or mental Measurement. Thus, measurement is categorized into these two broad categories or types.

STOP TO CONSIDER

Measurement is categorized into these two broad categories or types physical Measurement and Psychological or Mental Measurement. Measurement used to measure the different physical phenomena in different physical science and biological science is known as Physical Measurement. Measurement used to measure different humanistic and behavioristic phenomena in different social science and in applied science is known as Psychological or Mental Measurement.

Measurement in Physical Science:

You are aware that measurement in physical science indicates the physical measurement. It exists in the physical and material world. The most significant characteristics of this kind of measurement is that it is concerned with certain dimensions like-volume, size, length, weight, breath etc. It is to be clearly mentioned here that in case of physical measurement, it is possible to measure the whole amount like –the whole room, whole body, whole table etc. Therefore, it is absolute and definite in nature. Here, the units of measurement are fixed and constant throughout the entire scale, for which the measurement in physical sciences have same value everywhere. Physical measurement refers to a zero (0) point. For example, if you have to measure your length, you have to place the zero point of the meter at the edge of the table and then you have to read the scale point corresponding to your head, thus you will get your length.

SELF ASKING QUESTIONS**2. What is the primary characteristic of Physical Measurement?****Measurement in Social Science:**

Measurement that enters into all fields of human activities and human characteristics implies measurement in social science. Measurement in social science specifically emphasizes on the measuring psychological characteristics. Therefore, it is known as Mental or Psychological measurement. The units of measurement are fixed and constant throughout the entire scale, for which the measurement in physical sciences have same value everywhere. You should keep in mind here that psychological characteristics are analyzed and interpreted in terms of behaviour in different circumstances, for which mental measurement is indirect or relative and it is not possible to measure the whole amount of the particular mental trait. Most important characteristic of mental measurement is that it is subjective. For example, if you want to measure the attitude of your friend or any familiar fellow, you cannot measure his attitude directly just like measuring his/her weight and length. To measure attitude, you have to take help of an attitude test (there are different types of psychological test developed by the psychologists on the basis of their extensive research studies for measuring different mental attributes of people). This attitude test will again measure different aspects of behaviour of the particular people related to his/her attitude. Again, it is heartening to mention here that there are different factors which may influence the attitude testing like the testing condition, age of the subject etc. (you will come to know about psychological test and its allied

concept in details in the following units). Like the physical measurement, there is no zero point in mental measurement.

STOP TO CONSIDER

Physical Measurement exists in the physical and material world. The units of measurement here are fixed and constant throughout the entire scale, for which the measurement in physical sciences have same value everywhere.

In mental measurement psychological characteristics are analyzed and interpreted in terms of behaviour in different circumstances, for which mental measurement is indirect or relative and it is not possible to measure the whole amount of the particular mental trait.

1.4.1: Difference between measurement in Physical and Social Sciences

On the basis of the discussion held so far, you will be able to differentiate between the measurement in physical science and measurement in social science. This distinction will again clarify your concept regarding the types of measurement—

Physical Measurement	Mental Measurement
1. Physical measurement is direct and accurate	1. Mental measurement is indirect.
2. Physical measurement is objective in nature	2. Mental measurement is subjective in nature
3. It is absolute in nature	3. It is relative in nature
4. In physical measurement, units are referred to zero point.	4. There is no concept of zero point in mental measurement
5. It is possible to measure the whole amount of physical dimension in physical measurement	5. It is not possible to measure the whole amount of the mental trait in mental measurement.
6. Here, units have same value everywhere.	6. Here, units do not have same value everywhere; measurement may vary person to person.
7. It can provide direct comparison.	7. Direct comparison is not possible in mental measurement.
8. It is complete in itself.	8. It is not complete in itself.

SELF ASKING QUESTIONS

1. Differentiate between Physical Measurement and Mental measurement.

1.5: Scales of Measurement

This unit gives you idea about measurement. In the study of measurement, it is essential for you to be familiar with the scales, which are the part and parcel of measurement. It is well known to you that measurement is the means of describing data in terms of numerical values. The numerical value, in measurement is expressed on certain scales. These scales are pre defined and are based on mathematical and logical assumptions. There are different levels of measurement. Corresponding to different levels of measurement there are different scales.

There are basically four scales of measurement which are hierarchically arranged. These are—

1. Nominal or Classificatory Scale.
2. Ordinal or Ranking Scale.
3. Interval Scale.
4. Ratio Scale.

Let's discuss the scales of measurement-

1. Nominal Scale:

This is the lowest level of measurement. Measurement on this scale implies assigning different categories on the basis of some common characteristics. Here numbers or symbols are used to identify an Object, Person or Group and its Characteristics. For example, different players are given different numbers for the purpose of identification. Big classes of a school is classified into Sections A, B, C or A1, A2, A3 etc. for better management of the class. This scale is also known as classificatory scale. The primary function of this scale is to determine whether two persons are members of the same class or not. Again the main feature of the scale is that once a person leaves the group, he immediately loses the number. Some eminent persons are remembered with their numbers or symbols even after their leaving the group but this identification is not related to nominal scale. It is called nickname in literature. Classification of persons into males and females, objects on the basis of its colours, classification of people on the basis of their communities, numbers given to vehicles showing a particular District of State etc. are some other examples of nominal scale.

2. Ordinal Scale:

When various groups of the population are divided according to a certain order or sequence keeping in view the properties being

measured, it is called ordinal scale. That means, in ordinal scale, people are ranked or categorized on some dimensions. For this, ordinal scale is also known as ranking scale. Ordinal scale thus not only classify person or any object but also rank them in terms of the degree to which they possess a characteristic (like order from top to bottom or higher to lower). It is important to mention here that in this scale, intervals between the ranks are not equal. For example; rank first, second, third and fourth may be 80, 76, 60 and 50 respectively. Thus, there is difference of 4 between first and second, 16 between second and third and 10 between third and fourth and so on, thus the ranks are not equal. Classification of students into various groups – bright, average or poor, classification of the society into upper, middle and lower classes, ranking football, tennis or cricket players and teams as first, second, third etc. on the basis of their performance, assigning different grades to students keeping their particular attributes into account, arranging the students in a line according to their height from the tallest to the shortest etc. are some examples of ordinal scale.

In ordinal scale a transformation which does not change the order of the classes is completely admissible. For example, if a student getting distinction marks is given Rs.500 and another student scoring just first division is given Rs. 300 only as prizes, it means that the one with distinction is greater than the first divisioner.

STOP TO CONSIDER

Nominal Scale is the lowest level of measurement. Measurement on this scale implies assigning different categories on the basis of some common characteristics. Here numbers or symbols are used to identify an Object, Person or Group and its Characteristics.

Ordinal scale is also known as ranking scale. Ordinal scale classify a person or any object as well as rank them in terms of the degree to which they possess a characteristics (like order from top to bottom or higher to lower).

1. Interval Scale:

When a measurement conveys meaningful information with respect to distance or degree of difference between different points or the scale, it is called equal interval scale. That means, an interval scale involves classification, magnitude, as well as equal sized units. Equal distance of the scale indicates equal distance in the properties being

measured. i.e. units of measurement in this scale is constant and equal. Zero point (absence of the property) is also found on the scale (zero point in education does not, however, mean absence of the property). Zero point is only arbitrary in behavioural sciences but it is true in all physical measurement. Most of the tests used in psychological testing represent interval scale. Measurement of temperature, achievement or aptitude tests of objective nature etc. are some examples of interval scale.

2. Ratio Scale:

It is the highest level of measurement and has all the properties of nominal, ordinal and interval scales plus an absolute or true zero point. Origin of these scales is absolute zero. In this scale, measures are expressed in equal units. In ratio scale; the ratio of any two scale points is independent of the unit of measurement. In this scale all the nine postulates of measurement can be applied and hence all the statistical operations are possible without exception. Height, weight, length, loudness, vision etc. are common examples of ratio scale.

STOP TO CONSIDER

When a measurement conveys meaningful information with respect to distance or degree of difference between different points on the scale, it is called equal interval scale.

Ratio Scale is the highest level of measurement and has all the properties of nominal, ordinal and interval scales plus an absolute or true zero point.

SELF ASKING QUESTIONS

4. Discuss the different scales of measurement.

1.6: Meaning of Educational Measurement and Evaluation

In the study of Educational measurement, evaluation is placed at the centre position, because evaluation is inseparable part of educational measurement. The two terms-measurement and evaluation are like two sides of the same coin. As you know, measurement is the process of quantifying something. Measurement carried out in the field of education indicates educational measurement. Educational measurement sounds same with

measurement in social science. Education falls under the scope of a branch of social science. As such it is behavioural discipline. The psychological measurement that has already been discussed will help you to understand educational measurement.

Before going to discuss about the educational evaluation, let's get an idea about evaluation. It is an act of placing value on something. To evaluate means, to form judgment on the level of achievement. Evaluation is a systematic process of collecting and analyzing data in order to make decisions. Evaluation includes both qualitative and quantitative description of behaviour. It is important to note here that evaluation is not to determine whether something is good or worthwhile as opposed to bad or worthless. It is a continuous and systematic process.

STOP TO CONSIDER

Evaluation is a systematic process of collecting and analyzing data in order to make decisions. Evaluation includes both qualitative and quantitative description of behaviour.

Educational Evaluation:

Evaluation in education is a very old concept. Education is a continuous process. Therefore, it is essential in the process of education to evaluate periodically the students' abilities and readiness. In other words it can be said that education is also an investment in the sense that time, energy, and money are spent to bring about behavioural change in learners. That is why when a teacher performs his teaching works in his class, he also simultaneously finds out whether the changes in behaviour of learners are in concurrence with the learning objectives or not. If it is not so, then there will be no way to know what types of modifications are needed and what steps of teaching are needed. He also finds out whether the problem lies in his teaching strategies or techniques or it is there in the students themselves. Thus, teaching and evaluation run-hand in hand. Evaluation determines the behaviour of students on one hand and testifies to the teaching of the teacher on the other. This very purposeful behaviour of the teacher is called evaluation on education. Educational evaluation appraises the extent to which objectives of education have been achieved in a comprehensive manner. In one sentence evaluation is an educational programme concerned with the measurement of all the three domains of learning:-

cognitive, affective and psychomotor. Process as well as product is evaluated to arrive at a final conclusion. The results of evaluation are always expressed in terms of behavioural changes of the person being evaluated.

SELF ASKING QUESTIONS

5. What do you mean by Educational Evaluation?

1.6.1: Nature and Characteristics of Educational Evaluation: -

1. A Comprehensive process

All the three domains of human learning- cognitive, affective and psychomotor, can be accurately measured by evaluation. Apart from student's behaviour, teacher's behaviour (his task of teaching) is also evaluated.

2. A Systematic Process

Evaluation is a preplanned and systematic process which goes ahead keeping in view the predetermined objectives. By this process efforts are made to find out how much change has occurred in the behaviour of students and teachers and how much change is yet to be brought about by the task ahead.

Evaluation	Measurement
1. Evaluation is a wider and inclusive concept	1. Measurement is a part of evaluation process.
2. Evaluation is based on the objectives of education.	2. Measurement is based on the specific objectives of education.
3. Evaluation is subjective in nature.	3. Measurement is objective.
4. Evaluation includes both conventional tool (test) and unconventional tools (interview, checklist etc.)	4. Measurement is only based on conventional tools.
5. Evaluation is a continuous process.	5. Measurement is only concerned about the present status.
6. Evaluation is the initial basis for value judgment.	6. Measurement is the end exercise.
7. Evaluation is a comprehensive process.	7. Measurement is an isolated process.
8. Evaluation provides both qualitative and quantitative description of behaviour.	8. Measurement is only concerned with the quantitative aspect of person's performance.

Thus, we can differentiate between evaluation and measurement. Though wide distinction exists between the two, they are complementary to each other. Evaluation is incomplete without measurement; measurement is a part of the wide scope of evaluation. In the field of education and psychology, both play significant and respective roles.

STOP TO CONSIDER

Measurement and evaluation are interrelated concepts. But they are not the same. Evaluation is wider and comprehensive process, measurement is a part of it.

CHECK YOUR PROGRESS

- Q. 1. Discuss the different Scales of measurement with appropriate examples.
- Q. 2. Discuss the relationship between evaluation and measurement.

1.7: Summing Up:

Coming to the last part of this unit, it can be said that this unit tries to give you idea about measurement and evaluation in education. The unit has also thrown light on the different scales of measurement. Thus we can summarize the unit as-

- Measurement implies the act of determining the quantity of a property in terms of a suitable unit.
- As said by **J.P.Guilford**, Measurement is the description of data in terms of number.
- The basic function of measurement is -To assign some numerals to some object or phenomenon, to determine quantitative aspects, to provide status, to compare.
- Measurement is categorized into this two broad categories or types physical measurement and mental measurement.
- Measurement used to measure the different physical phenomena in different physical science and biological science is known as Physical measurement.
- Measurement used to measure different humanistic and behavioristic phenomena in different social science and in applied science is known as Psychological or mental Measurement.
- The most significant characteristics of physical measurement is that it is concerned with certain dimensions like volume, size, length, weight, breath etc.
- Again psychological characteristics are analyzed and interpreted in terms of behaviour in different circumstances, for which mental measurement is indirect or relative and it is not possible to measure the whole amount of the particular mental trait.
- Physical measurement and mental measurement differs from each other.
- Physical measurement is direct and accurate; on the other hand, mental measurement is indirect.

- There are basically four scales of measurement which are hierarchically arranged. These are-Nominal or Classificatory Scale, Ordinal or Ranking Scale, Interval Scale and Ratio Scale.
- Nominal Scale is the lowest level of measurement. Measurement on this scale implies assigning different categories on the basis of some common characteristics.
- On nominal scale, numbers or symbols are used to identify an Object, Person or Group and its Characteristics.
- Ordinal scale is also known as ranking scale. Ordinal scales classify persons or any object as well as rank them in terms of the degree to which they possess a characteristics.
- When a measurement conveys meaningful information with respect to distance or degree of difference between different points or the scale, it is called equal interval scale.
- Ratio Scale is the highest level of measurement and has all the properties of nominal, ordinal and interval scales plus an absolute or true zero point.
- Measurement carried out in the field of education indicates **educational measurement**. Educational measurement sounds same with measurement in social science.
- Evaluation is a key area in the study of measurement.
- It is a systematic process of collecting and analyzing data in order to make decisions. Evaluation includes both qualitative and quantitative description of behaviour.
- Educational evaluation appraises the extent to which objectives of education have been achieved in a comprehensive manner. In one sentence evaluation is an educational programme concerned with the measurement of all the three domains of learning:- cognitive, affective and psychomotor.
- The key characteristics of educational evaluation reflect it as a Comprehensive process, a Systematic Process, a Continuous Process and a Social Process.
- Measurement and evaluation are interrelated concepts.
- Both are supplementary and complementary to each other.
- Evaluation always involves the use of measurement as a tool for obtaining information that can be employed in decision making activities.
- Evaluation is a broad concept; measurement is a part of it.
- The basic difference between evaluation and measurement is that evaluation is subjective in nature, whereas measurement is objective.

1.8: References and Suggested Reading:

Agarwal, J.C (2004): *Essential of Examination System*, Vikash publishing House, New Delhi.

Asthana Bipin(2009): *Measurement and Evaluation in Psychology and Education*, Vinod pustak Mandir, Agra-2

Goswamee Maramee(2003): *Measurement and Evaluation in Psychology and Education*,

Sharma R.A(2004): *Essentials of Measurement in Education and Psychology: Measurement and Evaluation*, Surya Meerut.

Sidhu.K(2009): *New Approaches to Measurement and Evaluation*, Sterling publishers, New Delhi.

—x—

UNIT- 2

TEST CONSTRUCTION AND ITS STANDARDIZATION

Contents:

- 2.0 Introduction
- 2.1 Objectives
- 2.2 Meaning and Nature of Test Construction and Standardization
- 2.3 Principles and Methods of Test Construction and Standardization
- 2.4 Item Writing and Analysis
- 2.5 Methods for Estimating Test Reliability
- 2.6 Test Validation Procedures
- 2.7 Test Scores and Scales of Measurement
- 2.8 Norms and Their Uses
- 2.9 Summing Up
- 2.10 Questions and Exercise
- 2.11 References/Suggested Readings

2.0 Introduction

Measurement and evaluation are the integral parts of education process. To measure something we have to utilize some special tools. In this particular topic we have to make use of some tests which are constructed by the teachers and experts. But all tests cannot measure the educational aspects systematically. Only that test is appropriate which is standardized in the field of education.

2.1 Objectives

After going through this unit you will be able to:

- *define* the principles and steps of test construction.
- *outline* the characteristics of a good test.
- *explain* the methods of estimating test reliability and validity.
- *explain* about different types of scales and test scores.
- *describe* about norms and its uses.

2.2 Meaning and Nature of Test Construction and Standardization

A standardized test has some special qualities like reliability, validity, norms and scoring procedure. It comprises carefully selected items, having been given to a number of samples or groups under standard conditions and for which norms have been established after careful evaluation. Standardized test is quantitative in nature and are constructed by following systematic procedure. It is produced by some test publishing agency and test experts. Contrary to that teacher made test is a product of teachers for own class room efforts. The standardized test deals with larger segments of knowledge or skills than teacher made test; even possible effort is made to make the test highly valid, reliable and discriminating.

The essential characteristics of measuring tools are to minimize personal, variable, constant and interpretive errors. The test would be objective, reliable, valid and standardized. A standardized test consists of many objective type. It is the result of a test which is prepared according to principles and administered on a large population distribution over wide geographical areas representing various types.

Cliff W. Stone was the first in 1908 to produce standardized test to measure arithmetic reasoning. In the words of Anastasi, standardization implies uniformity of procedure in administering and scoring the test.

In standardized test besides reliability, validity and norms, there must be the standardized context and questions. The method of administration, scoring procedure and interpretation should also be standardized. So, a standardized test is one which is constructed by following various steps in order to establish reliability, variability, norms, etc.

2.3 Principles and Methods of Test Construction and Standardization

1. **Planning the Test:** Planning is the first and foremost step in constructing a standardized test. This step includes:
 - a) **Determining Objectives:** It needs to identify objectives, at first, for the test while is going to be constructed. The determining objectives should not be more than three or four. For different subjects objectives should be of various types. At last the evaluation is done on the basis of these objectives.
 - b) **Analysis of Curriculum and Source Material:** The content analysis is another important step in test construction. After determining the objectives, books and materials related to the topics should be analyzed. Chapters of books, synopsis, class notes, comments, questions of past year may be helpful in this regard. The following procedures are followed in source material:

- i) **Text Books and Note Books:** In this step the lecture, omitting a word, etc comes and therefore completion by students like can, usually, seldom words etc should not be used.
- ii) **Thought Question:** This may be another source for objective test items. Some modern texts are accompanied by objective tests in supplementary work book.
- iii) **Lecture Notes:** Lecture note may be considered as important source material where teachers own experiences and extended readings are included.
- iv) **Class Discussion:** Class discussions in which students as well as instructors participate are considered as an important source.
- v) **Job Analysis:** It is, particularly, helpful in finding items in case of aptitude test. Various aspects of many complex skills may be included as source material.
- vi) **Study of Test:** Study of various standardized test will help in construction of items for a new test.

2. **Preparing Final Draft:** It is the second step of test construction.

It is considered as blue print of entire test. This step is as follows:

- a) **Item Construction:** After determining objectives and analyzing source materials the construction of items started, the items may be true or false, alternative response, multiple choice, completion, etc. At the time of item construction some general principles should be kept in mind, like: the items belonging to same category should not be repeated. In the first draft the items should be more than that in the final draft and ambiguous questions should not be included in the test and it should be comprehensive enough.
- b) **Arranging the Items:** After constructing the items, arrangement of items should be followed in the following ways:
 - i) **Equal Difficulty Plan:** All items approximately of equal difficulty may be taken.
 - ii) **Increasing of Difficulty Plan:** In some test, like power test, items may be of increasing difficulty order.
 - iii) **Spiral Plan:** In this plan one item of various types is included. This will be followed by another item from various categories.
 - iv) **Logical Grouping:** It means grouping of items by subject wise and items should be arranged according to logic.

3. **Trying Out the Test:** After preparing the final draft, the test is tried out on a norm group. In this step, the following sub steps are followed:

- a) **Administration:** After preparing the final draft the test is administered on a sample where good, poor and average students are included. There should be proper arrangement of administering the test with clear instruction.
- b) **Scoring:** In this trying out process, the scoring procedure should be of simple one. One point may be given for every correct response and for that, answer key must be prepared beforehand.
- c) **Item Analysis:** This technique increases effectiveness of a test. It can be analyzed quantitatively with reference to their form and subject. Moreover, the length of a test can be reduced by item analysis.
- d) **Item Difficulty:** In this step, item difficulty is found out on the basis of pass percentage of students. This is helpful in finding out the difficulty level of each and every item.
- e) **Internal Consistency:** The internal consistency of items is done on the basis of total scores of the test. Internal consistency of items needs to be found out and it can be done by coefficient of correlation of the test item.
- f) **Finding out Discrimination Value:** The percentage of students doing each item correctly and incorrectly has to be calculated. This discrimination value can be found by the following methods :

$$D = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{N_1} + \frac{P_2 Q_2}{N_2}}}$$

1. **Preparing the Final Draft:**

After scoring from the item analysis the final test may be constructed. The items with good discrimination value may be taken and others may be eliminated. Proper instructions are to be given and required time should be decided. After that second administration of the test should be done. Then average percentage may be calculated. Consequently standard deviation, range and quartile deviation may also be calculated.

2. **Standardization and Interpretation:**

In standardization and interpretation, norms may be calculated like age norm, sex norm, grade norm, rural and urban norm, etc. Average performance of various groups will help in comparison of scores. After this

raw scores may be converted into standard score like Z and T score. The procedure adopted must be suitable for the test so that everyone can interpret the test easily.

If a test follows the above mentioned steps and procedure then it will become a standardized test and it can be applicable at anytime and anyplace.

STOP TO CONSIDER

A standardized test has some special qualities like reliability, validity, norms and scoring procedure. It comprises of carefully selected items, having been given to a number of samples or groups under standard conditions and for which norms have been established after careful evaluation.

A standardized test is one which is constructed by following various steps in order to establish reliability, variability, norms etc.

CHECK YOUR PROGRESS

- Give a definition about standardized test
- Mention the important principles of test construction and standardization.

2.4 Item Writing and Analysis:

Item writing includes objective items: True and false, alternative, matching, completion, etc, (*Referred Unit-V-Objective Type of Test*)

Test construction is one of the important aspects of measuring mental faculty of human beings. In constructing a test different stages have to be followed by the test constructor. Item analysis is one important stage of it.

Item analysis as a step of test construction provides many valuable information to the test constructors. Tests consist of different items. Test efficiency whatever its purpose depends upon caretaker in selecting the items. Item analysis helps the test user in their test evaluation. This procedure helps to differentiate between better and poorer items. It is a statistical technique which is used for selecting and rejecting the items of a test on the basis of their difficulty value and discriminative power. The major objectives of item analysis are the improvement of reliability, validity or both, and the achievement of better item sequences and types of score distributions. The primary purpose of item analysis is to improve the quality of the test items for better future administration. The main purposes of item analysis are:

- ❖ It is for improvement of total score's reliability and validity, and achievement of better item sequence and types of score distribution. It ensures high validity and reliability in a test, which depends on the characteristics of each item.
- ❖ Through items analysis qualitative, quantitative, content validity are evaluated. In qualitative analysis the content validity and the manners of stating items are evaluated. In quantitative analysis knowledge of item difficulty and item validity is sought.
- ❖ It can help the test user to evaluate the test more comprehensively.
- ❖ To reduce the length of a test. It also helps in bringing proper and suitable test items.
- ❖ To find out the difficulty level of items. It is that level where the higher mental group can complete the test items and lower group cannot.
- ❖ To find out the discriminating value of the test items. Discrimination value of the item is the percentage of examinee doing each item correctly as well as incorrectly. It helps in differentiating between persons having greater and lesser amount of knowledge.
- ❖ To find out the internal consistency of items.
- ❖ It is done to measure whether an item is to be accepted or rejected.

In test construction various criteria should be adopted by test constructors to select different items of the test. The important criterias are mentioned below:

- In item selection the test constructor has to select proper items from preliminary draft so that the test may be interesting as well as representative. The number of item in the preliminary draft should be more than that of the final draft. It helps the test user to bring suitable items which can represent the entire content of test.
- The items should be clearly phrased so that their context and not their form determine the responses. The words like always, exclusively, never are to be avoided from the items.
- The items belonging to same context should not be repeated in the same test.
- Items may be put in an ascending order of difficulty. It means simpler items should be placed in the beginning and more difficult items at the end.
- There should not be any sequence in responses. Responses should not have a sequential order so that one response should not lead to another.
- At the time of selecting the test item more importance should be given on learning rather than memory.

- The items should be selected from different areas so that it can represent the entire course.
- Attention must be paid to avoid chance error. No items should be such that it would be replied by referring to any other item or a group of items.
- Each item should carry equal marks. It saves the unnecessary time wasted by the students in order to complete the items beyond their capacities.
- The wording of items should be such that whole content determine the answer and not a particular part.

STOP TO CONSIDER

- Test construction is one of the important aspects of measuring mental faculty of human beings. In constructing a test different stages have to be followed by the test constructor.
 - Item analysis as a step of test construction provides many valuable information to the test constructors.
 - It is a statistical technique which is used for selecting and rejecting the items of a test on the basis of their difficulty value and discriminative power.

CHECK YOUR PROGRESS

- Mention two important purposes of item analysis.

Qualities of a Good Test:

A good psychological test has many qualities. If a test has all required qualities then that test is considered as good psychological test. The good psychological test have both technical and practical criterion. The practical criterion test must have face validity, ease of administration and interpretation, time and ease of scoring. The practical criteria is also a necessary quality. But so far the good test is concerned technical criteria are most essential. Following are the technical qualities of a good psychological test-

- ❖ **Validity:** Validity is an important consideration of a good measuring device. Validity of a test refers to the test's quality to measure what it intended to measure. It means a test is valid when it can attain the objectives for which it is designed. Thus, validity implies truthfulness. It helps us to judge whether the test measures the right thing for our purpose. If a test measures to a high degree of things it purposes to measure, it is valid. On the other hand, validity depends upon objectivity with which it measures the desired ability. Thus, validity is the foundation stone over which the entire super structure of testing is based.

According to Cronback, "Validity is the extent to which a test measures what it purposes to measure."

- ❖ **Objectivity:** Another quality of a good psychological test is objectivity. A good test is always objective. It is a pre-requisite for both reliability and validity. Objectivity is maintained when same scores are obtained by administering the test to a particular group of people on different occasions and getting it scored by the same scorer. In objective test scores are seldom affected by the physical, mental condition, etc. since each item of an objective test will have only one correct answer, there will be no possibility of subjectivity. Objectivity of a good psychological test has two aspects—objectivity of item and objectivity of scoring.

Objectivity of item construction implies that the items should be as simple as possible. Students should be able to interpret items correctly. An objective test item cannot have more than one definite answer. But words like perhaps, always, never, etc should not be used in the test items.

Objectivity of scoring implies that personal judgment of examiner is not affected in scores. It implies subjectivity factor is not affecting the scoring procedure of objective test whatever the examiner's judgment be, it must always and everywhere be same.

- ❖ **Reliability:** Reliability is another characteristic of a good evaluating technique. Reliability refers consistency of a test; it means the same result is measuring whenever it does measure. Thus, reliability refers to accuracy of measurement. A test is called reliable, when it shows the same result in repeated measure. It is the consistency of person's score on a series of measurement. Therefore, test reliability indicates the extent to which individual differences of the test scores are attributed to chance errors of measurement.

In the words of Anastasi, "Reliability refers to the consistency of scores obtained by same individuals when re-examined with the same test on different occasions or different sets of equivalent items or under variable examining condition."

A reliable test is expected to be free from chance error. However, reliability is a matter of degree and no test can be said to be 100% error free. Thus, scores obtained by reliable test are more or less stable.

- ❖ **Usability:** Usability refers to the practical aspect of the test in respect of administration, scoring interpretation and economy. By usability it means the degree to which the test or other measuring instrument can be successfully employed by class teacher or school administration.

Good psychological test must have the quality of usability. For that the instruction should be clear or definite and the scoring procedure must be simple and transparent with clearly stated. Moreover, it should be economic in cost and time and must be easy in interpreting the test value.

❖ **Norms:** Another important criterion of good psychological test is norms. Norms implies a particular standard. It is considered as a standard unit of a particular test. Norms is considered as a medium or average performance of pupil of certain age group in a standardized test. In the test reliability, validity, scoring procedure, norms table should be included so that the test can be easily administered by any one. If the norm table is associated then administrator can interpret the test scores easily.

STOP TO CONSIDER

- A good psychological test has many qualities. It has both technical and practical criterion.
- Regarding practical criterion the test must have face validity, ease of administration and interpretation, time and ease of scoring. But so far as the good test is concerned technical criteria are most essential.

CHECK YOUR PROGRESS

- What are the important criteria of a good test? Mention them.

2.5 Methods for Estimating Test Reliability

Among the good psychological test qualities reliability is a prominent one. A measurement procedure is reliable to the extent to which it provides constant result on repeated measurement. Consistency implies an individual is obtaining same score on repeated measurement. Test reliability tells to what extent the individual differences of scores happened. It tells us the extent to which true differences of traits can be attributed to these individual differences. Moreover, it refers to the accuracy of measurement. A reliable test is relatively free from personal errors because personal feelings, likes and dislikes may not affect the scores.

In order to understand the test fully it is very essential to know what proportion the test variance is due to possible resources of variance. For this purpose one has to obtain two measurement of an individual and correlate the scores. Generally, four methods are applied for estimating reliability coefficient. They are—

v **Test-Retest Method:** It is the simplest method of estimating reliability. In this method a single test is administered twice on a sample with reasonable time gap. Thus, two sets of scores are obtained—scores of first and second administration. It means once a test is given to a group of students, their scores are noted. After sometime the same test is given to the same group of students and again scores are noted. After that, the coefficient of correlation is calculated between the two sets of scores. This reliability index is known as test-retest reliability coefficient. This indicates to what extent the individual retain their relative position as measured in terms of test scores over a lapse of time. Generally Pearson's coefficient of correlation formula is applied for obtaining coefficient between two sets of scores.

v **Split Half Method:** This method is sometime called a coefficient of equivalent. In this method the test is split into two equal halves usually by putting the odd numbered items for one group and even numbered items for another group. This usually makes two scores obtained from single test reasonably equivalent. One half of test includes the items like 1,3,5,7, etc. and another half includes items like 2, 4, 8, 10, etc. Thus, the test is administered twice—one for odd numbers and other for even numbers. Moreover, this method provides internal consistency of test scores. All items are arranged in increasing order of difficulty and administered once on sample. Thus, the coefficient of correlation between these scores is obtained and reliability of whole test is calculated by using Spearman Brown formula.

v **Parallel Test Method:** In this method the parallel form of the test can be administered to the same sample under similar condition. It is also known as alternative form of reliability. In this method two parallel forms of the test are prepared which are equivalent or parallel in content, objectivity, discriminating value and time are concerned. Though both forms are parallel but one is not the true copy of the other. Both forms are administered one after another. From these two forms of test two different sets of scores are obtained and the coefficient of correlation is calculated by using Pearson's method of correlation.

This method is appropriate for determining reliability of education and psychological test. But its practical difficulty is that it is difficult to set two parallel form of test at a time.

v **Rational Equivalence Method or Kuder Richardson Method:** This method is improved method of estimating co-efficient of correlation over earlier three methods. This method helps to compute internal correlation among the test items as a whole. As this method has been

developed by Kudar Richardson, hence it is popularly known as Kuddar Richardson Method.

It is assumed, in this method, that all items have equal difficulty value. Correlations between the items are equal; all items measure essentially the same ability and the test is homogeneous in nature and higher inter item consistency is expected to find out. e.g. - If a test includes multiplication and another test has addition, subtraction, multiplication and division, the first one will have more inter item consistency than later one. That is why similar types of items are included in the test.

The most common way of finding inter items consistency is through the formula of Kudar and Richardson. The formula is

$$r = \left(\frac{n}{n-1} \right) - \frac{\sigma^1 - \sum pq}{\sigma^2}$$

This method tries to find out actual coefficient of correlation between successful and unsuccessful individuals with each item.

STOP TO CONSIDER

Reliability refers to the consistency of scores obtained by same individuals when re-examined with same test on different occasions or different sets of equivalent items or under variable examining condition.

CHECK YOUR PROGRESS

Define reliability.

What are the methods of estimating reliability?

2.6 Test Validation Procedures

Validity is one of the important considerations of a good measuring device. A test is said to be valid when it measures what it intends to measure. So, validity provides a direct check on how well the test fulfills its functions. It depends upon the objectivity with which it measures the desired validity. Therefore, a good measuring device is considered as valid when it can measure the attainment of objective for which it is designed.

Methods of Estimating Test Validity

The methods of estimating test validity are also known as forms of expressing validity. The following methods are generally used to find out validity coefficient.

➤ **Correlation Coefficient Method:**

In this method the test scores are correlated with the criterion scores. The obtained coefficient of correlation gives the extent of validity index to a test. Though several methods of correlation are available but Pearson's method of correlation is most widely and popularly used. This formula is used in multiple correlations. Multiple correlations are one where more than two measures are involved. The technique of correlation depends on the nature of data obtained on test as well on criterion. The most popularly used method is

$$r = \frac{N \sum xy - \sum x \sum y}{\sqrt{N \sum x^2 - (\sum x)^2 (N \sum y^2 - (\sum y)^2)}}$$

➤ **Expectancy Table Method:**

In this method the test scores are correlated with the rating of supervisors. It provides empirical probabilities of the validity index. Norms are presented in a better way in terms of expectancy table which provides the user with an interpretation of raw score directly in terms of future behavior. It indicates the probabilities that a person achieving a given test score will behave in a certain way in some second situation. But the expectancy table is not generalizable to new situation. It is specific in case of every new situation. The high coefficient of correlation indicates higher validity of the test.

➤ **Cross Validity Method:**

Cross validated method is another prominent method of estimating test validity. Proper evaluation, usefulness and prediction of scores derived from one sample and to be validated on second sample of the subjects from the universe. This test is considered as cross validated for the universe. True cross validation would be a trial of selected items for new group. To properly evaluate the usefulness of a test, the prediction, evaluation must be derived from one sample of information and validated on second sample of subjects from the same universe. If this has been done it is said to be a cross validation.

Cross validation is determined by trying out previously developed and refined test on a completely new group as close as possible for that test.

➤ **Item Analysis Method:**

Another method of estimating test validity is the item analysis method. To be valid the test items of an achievement test it should have proper difficulty value and discriminating power. The difficulty value of an item is defined on the percentages of cases who fail to answer the items. Discriminating value implies the degree to which an item distinguishes

between high group and low group. If the number of persons in high group passing the items higher than the number of persons in low group passing the items is then said to be having discriminating value.

The processes by which the discriminating value and difficulty power of individual items of a test are calculated are called item analysis, which is discussed in **detail in item analysis section**.

STOP TO CONSIDER

Validity is one of the important considerations of a good measuring device. It means truthfulness.

A test is said to be valid when it measures what it intends to and when it fulfills the objectives of the test.

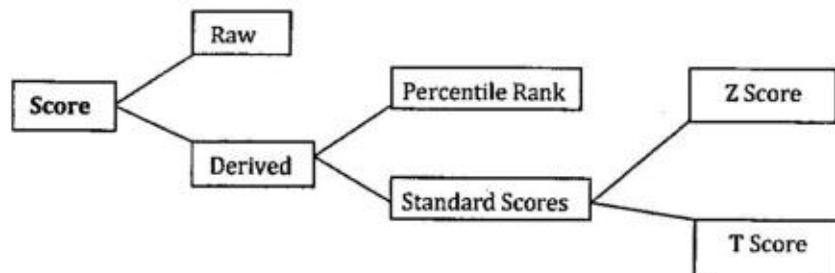
CHECK YOUR PROGRESS

Give one comprehensive definition on validity.

Write about few lines on item analysis method of estimating validity.

2.7 Test Scores and Scales of Measurement:

Raw A score means a numerical description of an individual's performance. Scores are of two types-



A raw score becomes significant only when it is comparable with other scores. In order to make them comparable and to interpret the raw scores are converted into derived scores. A derived score is a numerical description of a pupil's performance in terms of norms or performance of a standardized group.

Standard scores represent measurements from the Mean in the Standard Deviation unit. The difference of raw scores from mean in SD unit is expressed by standard score denoted by sigma score or Z score, is given by the formula—

$$Z = \frac{X-M}{\sigma} \quad \text{Where X-Raw Score, M-Mean, } \sigma\text{-SD}$$

Though the sigma score is most frequently used, it is sometimes awkward to have negatives or scores with decimals. Therefore, another version of a standard score, the T-Score has been devised by McCall, to avoid some confusion resulting from negative of Z score (below the mean) and also to eliminate decimal values. McCall suggested this scale with a mean 50 and SD 10 and named it 'T' after Thordike and Tolman. The formula-

$$T=50+10z$$

Hence, a raw score of 40 in σ score, where $M=50$ and $SD=5$ will be

$$\begin{aligned} & \frac{X-M}{s} \\ & \frac{40-50}{5} \\ & = -2 \sigma \end{aligned}$$

And the same raw score (40) or the σ score (-2 σ) in the T scale will be

$$\begin{aligned} & =T \text{ or } 40=50 + 10 \frac{X-M}{\sigma} \\ & =50+10 (-2) \\ & =50-20 \\ & =30 \end{aligned}$$

STOP TO CONSIDER

The sigma score is most frequently used, it is sometimes awkward to have negatives or scores with decimals. Therefore, another version of a standard score, the T-Score have been devised by McCall, to avoid confusion resulting from negative Z Score (below the mean) and also to eliminate decimal values.

CHECK YOUR PROGRESS

Mention the formula of finding out Z score.

Find out T value when raw score is 40 and σ is 20

Scales of Measurement:

In mental measurement, a scale may be thought of as a continuous or continuity along with items, tasks, problems and the like have been located in terms of difficulty or some other attribute. The units of a scale are arbitrary and depend upon the method employed by the investigator. Ideally, scale units should be equal, have the same meaning and remain stable throughout the scale.

Various scales are used to assess intelligence, interests, attitudes, opinions and the other personal, social adjustment of the individual. Scales

are methods of quantification or numerical method of describing observation of materials or characteristics.

Stevens classified four scales of measurement after various references passed on the scales. These are as follows:

- **The Nominal Scale:** The nominal scale is a primitive form of measurement. When a set of objects is differentiated among two or more categories on the basis of their qualitative differences that is called a nominal scale. Usually, a number or other symbol is used to represent all objects in a given category. Eg-in a group of people '1' is given to each male and '2' is given to all female members. The choice of numbers is quite arbitrary. The purpose of numbers is to make easy distinction between the two categories. It has no quantitative significance. Eg- pincode, car number plates, etc.
- **The Ordinal Scale:** The ordinal scale of measurement corresponds to quantitative classification of a set of objects. These sets or classes of objects are ordered in a series ranging from lowest to highest according to the characteristics to be measured. The ranking of students height and weight are the examples of ordinal scale of measurement. The common arithmetic operations like addition, subtraction, multiplication and division cannot be used with ordinal scale.
- **The Interval Scale:** The interval scale also involves quantification but with an added refinement or advantage. In this scale the differences between consecutive points are equal over the entire scale. The consecutive points are equal but there is no true zero points on it. Most of psychological test or inventories are based on interval scale. They have no real zero point. Eg- A student may get zero in arithmetic test, but this does not mean that his knowledge of arithmetic is zero.
- **The Ratio Scale:** The fourth and highest level of measurement is the ratio scale. This scale has all the characteristics of internal scale with an additional advantage of a true zero point. It is possible to indicate the complete absence of an attribute. Eg- the zero point on a centimeter scale indicates the complete absence of length or height. The number of ratio scale can be expressed in ratio relationship. Eg- the weight of 4 grains is one half of 8 grams. The height of 10 cm is one third of 30 cm and so on.

STOP TO CONSIDER

- In mental measurement, a scale may be thought of as a continuous or continuity along which items, tasks, problems and the like have been located in terms of difficulty or some other attribute.
- Scales are methods of quantification or numerical method of describing observation of materials or characteristics.

CHECK YOUR PROGRESS

1. According to Stevens what are the different scales of measurement?
2. What is interval scale?

2.8 Norms and Their Uses:

Marks awarded on a test are known as raw scores. The raw score obtained by an individual in test does not have much significance because raw scores cannot be interpreted properly. To facilitate interpretation, sound psychological test must provide table of norms.

The term "norm" has two meaning. Firstly, it is used in the sense of standard comparison and secondly, it means average performance or central tendency of a group. So, members of a particular group are not similar. This is true of all ages. Norms should be based on average performance or capacity of the whole group. So, "A norm is the average or typical score (Mean or Mdn.) on a particular test made by a specified population," (Freeman), for example, the mean intelligence test score for a group of 10 year old children, the mean score for a group of fifth grade pupils in arithmetic test. Reference to such a norm helps us to rank a students' performance according to own or other age or grade groups.

There are various types of norms and they are:

- **Age Norm:** Age norms are expressed in terms of average performance of particular age group. The fundamental assumption of this norm is that the characteristic being measured increases systematically with the age. It is based on age equivalence. An age norm is the average score on a test of a particularly age group. Age norms are widely used in achievement and aptitude test. Intelligence test also provides age norm in the form of mental age. Thus, age norms are obtained from a representative group of different age level and computing central tendency of the score obtained in a particular age group.
- **Grade Norms:** Grade norm is like age norm. But in this norm, measurement is based on grade level. Grade norms are the average achievement of a particular grade of people. It is calculated from the obtained score of a representative group of same grade level. For example: If the average number of problems solved correctly on a particular test by 4th grades is 30, then a raw score of 30 corresponds to a grade equivalent is 4.

- **Percentile Norms:** In percentile norms, a person's performance is compared with that of the group to which he belongs. Scores are placed in a percentile rank order. Percentile rank of a given raw score is simply the percentage of students in a given group who fall below that score. Reference to percentile ranks of all the individuals in a given group, permits one to evaluate the standing position of a pupil. If an individual's percentile score is 75, it means that 75% persons in the group fall below him.
- **Standard Score Norms:** Standard score simply represents the number of standard deviations by which a particular raw score deviates from the mean of the distribution. It is obtained by finding the difference between the raw score and the mean and dividing it by the standard deviation of the distribution. Thus, if the mean of a distribution is 100 and its SD 20, a person who has a raw scores, a raw score of 90, for instance, would correspond to a Z score of -0.50.

STOP TO CONSIDER

- A norm is the average or typical score (Mean or Mdn.) on a particular test made by a specified population."- Freeman.
- There are various types of norms used for educational and psychological tests.

CHECK YOUR PROGRESS

1. Give the concept of standard score norm.
2. Give one example of age norm.

2.9 Summing Up:

- Much of Educational and Psychological measurement is and will probably remain, at a relatively low level of elegance and precision. We must recognize this fact, using the best procedures available to us but always treating the resulting scores as a tentative hypothesis rather than an established conclusion.
- Tests are given for many different reasons. In order to achieve such diverse purposes, they need to be carefully planned. In classroom setting, this planning usually entails instructional objectives and /or a table of specifications.
- The use of objectives or a table of specifications may imply strict guidelines for constructing tests, but much of testing is determined by practical considerations. Test constructors must consider such things

as how much testing time is available, what item formats the students can handle and the development.

- Standardized tests are generally qualitative in nature and usually constructed by test specialists and experts. It is generally applied on a norm group which is one on which test is tried out. So, a test which is constructed by following systematic procedure is known as standardized test.
- A psychological test is a test which is being applied in the field of education and psychology to measure the psychological aspects of students. These tests are prepared by expert psychologists to measure a sample of behaviour of students. In this test objectives are determined beforehand to measure the psychological aspect. The good psychological test have both technical and practical criterion. The practical criteria include validity, reliability, objectivity and norms.
- Reliability refers to the consistency of scores which is reflected in the scores. A test is said to be consistent over a given period of time, when all individual retain their same score on different testing with the same test. On the other hand, validity refers to the truthfulness of the test.
- Norms are frames for references for the interpretation of test scores. They are based on the actual performance of a reference group which is representative of a whole specific population. It is based on average or the mean of the test performance of the group.

KEY TERMS

- ❖ **Standardization:** It implies the process of a test which requires some important characteristics i.e. validity, reliability, objectivity and norms.
- ❖ **Reliability:** Consistency of a measuring instrument obtained by some individuals where re-test with the test on different sets of equivalent items or other variables examining condition.
- ❖ **Validity:** The degree to which the test actually succeeded in measuring what it sets to measure is called validity.
- ❖ **Norms:** It is the proper way for expressing the test result in a meaningful way. Norms are computed as a standard and individual scores are compared with it.
- ❖ **Standard Scores:** Standard scores represent measurements from the mean in standard deviation unit. The difference of raw score from the mean in SD unit expressed by standard score denoted by sigma ("σ") score.

- ❖ **Scales of Measurement:** In measurement context, numerical values are expressed on well defined scales. On the basis of mathematical and logical assumptions various types of scales are used in psychological measurement purposes and corresponding to different levels of measurement there are different scales.

CHECK YOUR PROGRESS ANSWER-1:

1. "Standardization implies uniformity of procedure in administering and scoring the test."
2. The important principles of test construction and standardization are-planning, administering the test, trying out, etc.

CHECK YOUR PROGRESS ANSWER-2:

1. Two important purposes of item analysis are –
 - a) The main goal of item analysis is the improvement of total score's reliability and validity, and achievement of better item sequence and types of score distribution. It ensures high validity and reliability in a test. Validity and reliability, on the other hand, depends upon on the characteristics of each item.
 - b) Item analysis is done to find out the discriminating value of the test items. Discrimination value of the item is the percentage of examinee doing each item correctly as well as incorrectly. It helps in differentiating between persons having greater and lesser amount of knowledge.

CHECK YOUR PROGRESS ANSWER-3:

1. The important criteria of good test are-reliability, validity, objectivity and norms.

CHECK YOUR PROGRESS ANSWER-4:

2. Anastasi says, "Reliability refers to the consistency of scores obtained by same individuals when reexamined with same test on different occasions or different sets of equivalent items or under variable examining condition."
3. The different methods of estimating reliability are Test –retest method, Split half method, Parallel test method, Rational Equivalence Method or Kudar Richardson Method.

CHECK YOUR PROGRESS ANSWER-5:

1. According to Cronback, "Validity is the extent to which a test measures what it purposes to measure."

2. Item analysis method of estimating test validity is the item analysis method. In order to be valid the test items of an achievement test should have proper difficulty value and discriminating power. The process by which the discriminating value and difficulty power of individual items of a test are calculated are called item analysis. In order to do item analysis many steps are to be followed. Only by following these steps proper items can be selected.

CHECK YOUR PROGRESS ANSWER-6:

1. The formula of Z score $= \frac{X-M}{\sigma}$
2. $40 = 50 + 10 \frac{X-M}{\sigma}$
 $= 50 + 10(-2)$
 $= 50 - 20$
 $= 30$

CHECK YOUR PROGRESS ANSWER-7:

1. Stevens classified the scales of measurement according to various levels of refineness passed by the scales. They are-
 - a) Nominal Scale.
 - b) Ordinal Scale.
 - c) Interval Scale.
 - d) Ratio Scale.
2. The interval scale involves quantification but with an added refinement or advantage. On this scale the differences between consecutive points are equal over the entire scale. Most of psychological test or inventories are based on interval scale. They have no real zero point.

CHECK YOUR PROGRESS ANSWER-8:

1. Standard score simply represents the number of standard deviations by which a particular raw score deviates from the mean of the distribution. It is obtained by finding the difference between the raw score and the mean and dividing it by the standard deviation of the distribution.
2. The example of age norm -Deficiency of one year in knowledge of vocabulary or reading ability in case of achievement test.

2.10 Questions and Exercise

Short – Answer Questions:

1. What do you understand by test standardization?
2. When is the first try out of the test done?
3. Write short notes:
 - a) Test scores.
 - b) Interval scale.
 - c) Item writing.
4. Write about different uses of norms.
5. What are the characteristics of a good test?

Long Answer Type Questions:

1. Write about the process of test construction and standardization.
2. Explain the various methods of estimating test validity.
3. Give a detailed account on the methods of estimating test reliability.
4. What is a norm? Explain its different types.

2.11 References/Suggested Readings

- Anastasi, A., and Urbina, S. (1997) *Psychological Testing*, 7th edition. New Delhi: Prentice Hall of Private Limited.
- Rath, R.K. *Fundamentals of Educational Statistics and Measurement*, 3rd edition. Orissa: Taratarini Publishers.
- Singh, Raj. (1994) *Techniques of Measurement and Evaluation*, 1st edition. New Delhi: CommonWealth Publishers.
- Thorndike, R.L. (1970) *Measurement and Evaluation in Psychology and Education*. New Delhi: Wiley Eastern Limited.
- Asthana, Bipin (2017/18) *Measurement, Evaluation and Assessment in Education*. Agra, New Delhi: Agrawal Publications. Agra.

—————x—————

UNIT- 3

MEASUREMENT OF GENERAL INTELLIGENCE

Contents

- 3.1 Introduction
- 3.2 Objectives
- 3.3 Meaning of Intelligence
- 3.4 Measurement of Intelligence
- 3.5 Types of Intelligence Test
- 3.6 Binet-Simon Tests
- 3.7 Wechsler Tests of Intelligence
- 3.8 Uses of General Intelligence Tests in Education
- 3.9 Summing Up
- 3.10 References and Suggested Readings
- 3.11 Answer to 'Check Your Progress'
- 3.12 Questions and Exercises

3.1 Introduction

Greatest creation of nature is human being. Human beings are enriched with a number of qualities and abilities. One such ability is general mental ability. All people of this universe have general mental ability. As individual differences are there among each and every individual, the general mental ability or intelligence of the individuals are also varied. Intelligence is very important for human survival. Measurement of intelligence is a very important area for psychologists as individual difference is there in intelligence. Psychologists have developed different types of tests to measure the intelligence of human beings. In this unit, we shall discuss the meaning of intelligence, the measurement of intelligence, the types of intelligence tests, Binet-Simon test and Wechsler test and the uses of intelligence tests in education.

3.2 Unit Objectives

After going through this unit you will be able to-

- *know* the meaning and definition of intelligence.
- *know* the concept of measurement of intelligence and its types.
- *develop* understanding about Binet-Simon test and Wechsler test.
- *understand* the uses of general intelligence test in education.

3.3 Meaning of Intelligence

Intelligence is a general mental ability through which people can adjust with the environment by solving different problems. For better adjustment and effective living, human beings have to adjust with the environment in a proper way. People are able to learn new things with the help of intelligence. As intelligence is a vast concept, there is no clear view about the meaning of intelligence. Different psychologists define intelligence from different perspectives. Some psychologists have defined intelligence as the ability to learn while and some have defined it as an ability of adjustment. Again some psychologists view it as a determiner of behaviour of an individual.

Some important definitions of intelligences are mentioned below:

- According to Terman, "An individual's intelligence is in proportion to his ability to carry on abstract thinking."
- According to Binet, intelligence is "judgement or good sense, initiative, the ability to comprehend and to reason well and to adapt one's self to circumstances."
- In the words of Stern, "Intelligence is the general mental adaptability to new problems and conditions of life."
- According to Cyril Burt, "Intelligence is the power of readjustment of relatively novel situation by organising new psycho-physical co-ordination."
- William James defined intelligence as "the ability to adjust oneself successfully to a relatively new situation."
- In the opinion of Prof. Henry E. Garrett, intelligence is "the abilities demanded in the solution of problems which require the comprehension and use of symbols, i.e. words, numbers, diagrams, equations, formulas."

- Jean Piaget opined that, "Intelligence is the ability to adapt to one's surroundings."
- In the words of Frank N. Freeman, "Intelligence is represented in behaviour by the capacity of the individual to adjust himself to new situations, to solve new problems, to learn."
- To quote Colvin, "An individual possesses intelligence in so far as he has learned or can learn to adjust himself to his environment."
- The comprehensive definition of intelligence is provided by American psychologist David Wechsler. In his opinion, "Intelligence is the aggregate or global capacity in the individual to act purposefully, to think rationally and to deal effectively with the environment."

The definitions mentioned above clearly reflect that the concept of intelligence is very vast. To summarise, it is said that intelligence is

- (i) **The Ability to learn.** With the help of intelligence, people can learn according to his/her own pace. The people who have high intelligence can learn very quickly.
- (ii) **The Ability to adjust.** People can adjust in a new situation very easily with the help of intelligence. Intelligent people can deal effectively with novel situations.
- (iii) **The Ability to carry on abstract thinking.** Only intelligent people can think in an abstract way. Abstract thinking means the use of concepts and symbols.

From the above mentioned meaning and definitions of intelligence, we can summarise the nature and characteristics of intelligence in the following way:

1. Intelligence is an innate natural power as it is termed as a native ability.
2. Intelligence helps people to adjust to an environment properly.
3. Education depends on intelligence but intelligence does not depend on education. All literate people are not intelligent and in the same way all illiterate people are not dumb.
4. Study, analysis and thinking of abstract concepts are based on intelligence. Without intelligence, it is not possible to think in an abstract way.

5. Other mental abilities of an individual like thinking, imagination, creativity are related to intelligence.
6. It is not easy to measure the intelligence of an individual.

3.4 Measurement of Intelligence

It is not an easy task to measure the native ability of an individual. Generally intelligence tests are developed to measure intelligence of an individual. The Main aim of these tests is to measure the ability or potential present in human beings for future success. Development of intelligence test is not new in the field of psychology. In the primitive age, intelligence in people was identified on the basis of the size of the head, weight of the body and height & figure of an individual. At that time, a person was called intelligent if he had a large head and athletic body.

Later on, this technique of measuring intelligence was changed. In the year 1879, Wilhelm Woundt established a psychological laboratory in Leipzig town of Germany. In this laboratory, lots of experiments on innate capacity of an individual were conducted by Cattle, Gilbert and Ebbinghaus. They tried to measure the intelligence of a person by measuring memory, sensory discrimination, muscular strength etc.

Francis Galton is another psychologist who established a laboratory in the year 1884. He tried to measure different psycho-physical aspects of human beings. American psychologist Cattell also tried to study the individual differences of individuals. In the year 1890, he first used the term 'Mental Test'.

In the first part of 20th century, Cyril Burt of England and Alfred Binet of France studied intelligence in a comprehensive way. They had tried to measure intelligence in a quite different way from those who had attempted to do so earlier. Burt could not prepare a detailed series of intelligence tests, but Binet prepared the most workable series of intelligence tests. In the year 1905, Binet developed the most acceptable scale of measuring intelligence. He along with his colleague Simon developed an intelligence test in Stanford University of France. This test was further revised in 1908 and 1911. After the death of Binet, this scale was re-standardised by Terman in 1916. There have been lots of revisions of this intelligence test. After that, many test of intelligence tests were developed by different psychologists to measure this general mental ability.

CHECK YOUR PROGRESS

Q 1: Define intelligence.

Q 2: In which year, Francis Galton established a laboratory?

Q 3: Who first used the word 'mental test'?

3.5 Types of Intelligence Test

On the basis of intelligence tests developed by different psychologists, intelligence tests are divided into the following heads:

1. On the basis of its form or content

- Verbal Intelligence Test
- Non-Verbal Intelligence Test
- Performance Test

2. On the basis of administration

- Individual Intelligence Test
- Group Intelligence Test

3. On the basis of scoring

- Age Scale
- Point Scale

These types are discussed below:

1. Verbal Intelligence Test:

Verbal tests are the oldest intelligence tests. When intelligence of an individual is measured by tests where oral or written language is used, it is known as verbal intelligence test. In this test, the testee express their thoughts, judgement, reasoning, comparison, analysis, problem solving by using language. Language is the most important component of this type of test. Oral or written instructions are given in this type of tests and testee also responded by using oral or written language. Some verbal intelligence tests are: memory test, vocabulary test, comprehension test, information test, association test, reasoning test etc.

Merits of Verbal intelligence Test: Certain merits of verbal intelligence test are mentioned below:

- Concepts and thoughts of an individual clearly expressed with the help of language.
- This test clearly measures the abstract thinking and reasoning of an individual.
- Confusion created during the application of these tests is cleared by the use of verbal or oral instruction.

Demerits of Verbal intelligence Test: Certain demerits of verbal intelligence test are mentioned below:

- This test does not measure the native intelligence of an individual.
- This type of tests is not applicable to young children and illiterate people.
- Validity and reliability of this test is not accurate.

1. Non-Verbal Intelligence Test:

Non-verbal test of intelligence are developed to minimise the drawbacks of verbal test. Nature of this test is opposite to verbal test. Different types of pictures, geometry, graphs are included in this test. This test is developed to measure intelligence of illiterate people, people with verbal deficiency and children. They can solve different problems related to the media mentioned above. Some non-verbal intelligence tests are: Chicago Non-Verbal test, Army Beta test, Roven's Progressive Matrices test etc.

Merits of Non-Verbal intelligence Test: Certain merits of non-verbal intelligence test are mentioned below:

- This type of tests is applicable to young children and illiterate people.
- This type of test is applicable for the people who have language difficulty.
- This test is helpful for measuring intelligence of physically challenged people.

Demerits of Non-Verbal intelligence Test: Certain demerits of non-verbal intelligence test are mentioned below:

- All mental aspects of an individual are not measured with the help of non-verbal test.
- Abstract concepts are not measured by this type of test.

1. Performance Test:

Performance test is one type of non-verbal test. Performance test is developed to measure those mental abilities which are not possible to be determined by using verbal test. The main content of this test is task performance and problem solving. In this test, intelligence is measured by the ability, skill and result of testee while performing a task. Language ability is not required in this type of test. This test is applied to all type of people especially illiterate people, children and people with verbal difficulty. This

test is also applicable for foreign people with different languages. In this test, oral language is used only in case of instruction. But in non-verbal test, instruction is given by providing a booklet. Performance test is conducted individually but non-verbal test is conducted in a group. Examples of some performance test are: Pinter-Patterson scale, Proteas Maze test, Alexander Pass-a-long test, Koh's Block Design test, Bhatia Performance test etc.

Merits of Performance Test: Certain merits of performance intelligence tests are mentioned below:

- This type of tests is very effective especially for young children and illiterate people.
- Cultural influence is not there in performance test. So, this test is applicable in any part of the world without modification.
- This test is used to identify different abnormalities in people.
- This test is also used in educational and vocational guidance.

Demerits of Performance Test: Certain demerits of performance intelligence test are mentioned below:

- Only limited mental traits are measured in performance test.
- This test is very expensive in terms of time.

2. Individual Intelligence Test:

Individual intelligence test means intelligence or general mental ability of a particular person is tested at a time. An individual test may be verbal, non-verbal or performance type. In some situations, it is needed to test the mental ability of people separately. To find the quality of a human being and to realise his/her personal problems, it is necessary to apply individual test. Individual intelligence test is divided into two types, viz. Individual verbal test and individual non-verbal test. Stanford Binet scale, Wechsler Bellevue Intelligence test, Pass-a-long test, Block Design test are some examples of individual test.

Merits of Individual Intelligence Test: Some good points of individual tests are:

- Results of individual tests are highly reliable and valid.
- Innate abilities of an individual are possible to identify with this type of test.
- A good rapport is established between tester and testee in this test.

Demerits of Individual Intelligence Test: Some demerits of individual tests are mentioned below:

- This type of test requires longer time and they are also very slow. This test is time consuming.
- Expert people are required to conduct this test.

2. Group Intelligence Test:

The tests which are administered on a group of people are known as group intelligence test. With the help of this test, it is easy to measure the intelligence of a large number of people. This test is very useful when we measure intelligence of a group of people in a short span of time. Group intelligence test is divided into two types, i.e. group verbal test and group non-verbal test. Army Alpha test and Army Beta test are some examples of group intelligence test.

Strengths of Group Intelligence Test: Following are the merits of group intelligence test:

- This test is less time consuming and economic.
- This test is very much quick and easy to administer.
- A large number of people can be studied in group test.

Weakness of Group Intelligence Test: Following are the drawbacks of group intelligence test:

- Good rapport is not established between tester and testee in this type of test.
- This test is not highly reliable and valid.
- It is not possible to observe the mental state of the testees in this test.

3. Age Scale:

There are some tests of intelligence where scoring procedure is different. On the basis of respondent's reaction, the mental age is used as scores in this test. That means where mental age of respondents is used as scoring of the test is known as age scale. In this test, mental age is given for every correct response. Instead of year, month system is used in mental age e.g. for 6 years, mental age is 6 and for 6 years 6 months mental age is 6.5. In this test, items of the test are classified on the basis of age. In Binet-Simon Scale, age scale is used.

Merits of Age Scale: Following are the merits of age scale:

- Test items are prepared on the basis of age of the respondents.
- Qualitative evaluation of respondents is possible in age scale.
- This scoring procedure is widely used all over the world.

Demerits of Age Scale: Age scale has some limitations. These limitations are mentioned below:

- Reliability and validity of this type of test is not satisfactory.
- It is very difficult to prepare test items on the basis of age of the respondents.
- It is not possible to measure intelligence of adult people effectively with the help of age scale.

3. Point Scale:

Point scales are those intelligence scales, where point or number is used instead of mental age. For each correct response, one specific number is given. These tests are homogeneous and unique in nature. Test items of these tests cover wide range and these tests are available for all age groups of people. Numerical evaluation of mental ability of individual is possible in these tests. At present, lots of intelligence tests are developed where point scale is used. One example of point scale is Wechsler Bellevue Intelligence Test.

Merits of Point Scale: Merits of point scale are mentioned briefly in the following ways:

- These tests are homogeneous in nature.
- Statistical analysis of results is possible in these tests.
- Point scale is related with numerical evaluation.

Demerits of Point Scale: Demerits of point scale are mentioned briefly in the following ways:

- Qualitative evaluation of respondents is totally ignored in this test.
- Continuous development of task or test items is not possible in this test.

Verbal, non-verbal or performance test, group or individual test and point or age scale- all these types of intelligence tests are widely used

at the present time. Specific type of intelligence test is used as per requirements.

CHECK YOUR PROGRESS

Q 4: Give one example of individual intelligence test.

Q 5: Write two differences between age scale and point scale.

3.6 Binet-Simon Tests

Alfred Binet is known as the father of intelligence testing as he developed the most scientific intelligence test. In the beginning of 20th century, Binet and his colleague Theadore Simon developed a scale of intelligence test which is also relevant today. At that time, French Government appointed Binet and Simon to study about mental disabled children. After that, Binet and Simon immediately developed a scale to measure intelligence in the year 1905.

➤ **Binet-Simon Scale, 1905:**

The first systematic scale for measuring intelligence was developed in the year 1905. There are 30 items in this scale. Test items are arranged in ascending order of difficulty. These items are helpful to measure high level of mental ability, which was not possible in previous time. Test items of this scale were prepared to classify students from general to lower level. Items of this scale mainly prepared on the basis of different aspects like judgement, comprehension and reasoning. As this scale was the first attempt by Binet and Simon, there are certain limitations.

➤ **Binet-Simon Scale of 1908:**

The scale of 1905 was criticised by some psychologists as there were certain limitations of that scale. So, Binet and Simon revised the previous scale after 3 years and developed the modified scale in the year 1908. In this scale, total numbers of items were increased from 30 to 59. These items were classified into different categories for the children of age group 3 to 13 years. Some items of previous scale were skipped in this scale and some new items were added. In this revision, the concept of 'mental age' was used.

STOP TO CONSIDER

Mental age: Mental age is a concept of intelligence testing. This concept was first used in Binet-Simon scale of 1908. Mental age of an individual can be estimated by test performance.

➤ **1911 Revision of Binet-Simon Scale:**

Binet-Simon revised the scale of 1908 was attracted by different psychologists of different countries like Belgium, Germany, England, Italy, Switzerland and USA. On the basis of the recommendations provided by psychologists of different countries, minor modification of 1908 scale was made. More items were added in this revision. Items of this revision were extended to adult people also but there were no test items for the age group of 11, 13 and 14 years.

➤ **Revision of 1916:**

For measuring intelligence, Binet-Simon scale was very popular in almost all the nations. All the nations revised and translated the Binet-Simon scale for use in their respective country. But, America faced problems in this regard. For that, L.M. Terman and his associates re-standardised the Binet-Simon scale in Stanford University of United States in the year 1916. This revision was known as Stanford Revision of Binet-Simon Scale. Total 90 items were there in this test for age group of children 3-14 years. Two new stages were also included in this revision i.e. Average Adults and Superior Adults. This revision was very popular in the world and reliability co-efficient of this revision was also very high (.80-.95).

Main focus of this revision was that, Terman used the concept of Intelligence Quotient (I.Q.) in this revision.

STOP TO CONSIDER

Intelligence Quotient (I.Q.): In the year 1911 Stern and in 1912 Kulkmann suggested the use of the concept Intelligence Quotient. This concept was first used by Terman in Stanford Revision of Binet-Simon scale in the year 1916. I.Q. is depending on mental age of an individual. Formula of calculating I.Q. is $I.Q. = \frac{MA}{CA} \times 100$ Where, MA = Mental Age CA = Chronological Age

➤ **Revision of 1937:**

For removing the drawbacks of 1916 revision scale, Terman again revised the scale with another psychologist Merrill in the year 1937. This scale was published in two forms- L Form and M Form. L indicates Merrill and M indicates Terman in this scale. There were altogether 129 items in each form of this scale. The reliability and validity co-efficient of this scale

was high very. Reliability co-efficient range is .90 -.98. This scale was time consuming as numbers of test items were very big.

➤ **1960 Revision:**

In 1960, Binet-Simon scale was again revised. All the good statements of L-Form and M-Form of 1937 revision were included in this revision. Age range of the respondents in this scale was 2 years 6 months to 18 years. In this scale instead of IQ, Deviation IQ (DIQ) was used.

➤ **1986 Stanford Revision:**

Fourth Stanford revision of Binet-Simon scale was published in the year 1986. Robert Thorndike, Elizabeth Hagen and Jerome Sattler were the psychologists behind the revision of this scale. In this scale, instead of Binet's Age scale, Point scale was applied. Test items of this revision were broadly divided into four sub heads namely Verbal Reasoning, Abstract/ Visual Reasoning, Quantitative Reasoning and Short Term Memory. Age group of the respondents in this revision was 2 years to 23 years.

➤ **Stanford Revision of 2003:**

Fifth edition of Stanford revision of Binet-Simon Scale was developed in the year 2003. This revision tried to measure multiple dimensions of abilities. There were ten sub groups in this scale including verbal and non-verbal domains. In this revision 2 years to 85+ years of respondents were included.

Binet-Simon Scale is very popular and effective even today. For that, different psychologists tried hard to minimise the drawback of this scale.

3.1 Wechsler Tests of Intelligence

Apart from Binet-Simon Scale, Wechsler's Intelligence test is another individual intelligence test which is very popular and widely used test. American psychologist David Wechsler developed this test. He was basically a clinical psychologist in a hospital of Bellevue. This test is known as Wechsler Bellevue Intelligence Test. Wechsler was interested to study the people who have low intelligence capacity. This test was developed in the year 1939. Latest revision of this scale was published in the year 1981. Again in the year 1949, Wechsler developed an intelligence scale for children and name that test as 'Wechsler Intelligence Scale for Children or WISC' and in 1955, he developed another scale of intelligence specifically for adults which was known as 'Wechsler Adult Intelligence Scale or WAIS'. In the original Wechsler scale, point scale was used instead of age scale. There

were total 11 sub tests in this scale. Nature of these sub tests were verbal and performance type. These sub tests are mentioned below:

1) Verbal Scale:

There are 6 sub tests in verbal scale. These are:

- (i) **Information:** Total 27 test items are there in this test. These items are mainly related with knowledge and memory of the respondents.
- (ii) **Comprehension:** To measure general knowledge and judgement capacity of the respondents, there are 14 test items in this test.
- (iii) **Arithmetic Reasoning:** There are 14 items in this scale. These items measured the mathematical power, problem solving technique and attention of the respondents.
- (iv) **Digit Span:** Through this test, backward and forward memory of the respondents is measured.
- (v) **Similarities:** To find the similarities between two or more things, this sub test was developed. 22 test items are there in this test.
- (vi) **Vocabulary:** Vocabulary of the respondents is to be measured with the help of this sub test. Total 42 items are there in this test where words are arranged in easy to difficult way.

2) Performance Scale:

There are 5 sub tests in performance test. These are:

- (i) **Digit Symbol:** In this test, the respondents are required to use symbols as a substitute of digits.
- (ii) **Picture Completion:** Total 21 pictures are there. All the pictures are incomplete. The respondents have to complete the pictures in this test.
- (iii) **Block Design:** In this test, respondents have to prepare different blocks with the help of wooden or plastic cubes of different colours.
- (iv) **Picture Arrangement:** In this test, there are different pictures which should be arranged serially by the respondents to create a story.
- (v) **Object Assemble:** In this test, the respondents should collect different parts to make whole thing.

For completion of this test, almost one hour is required. Scoring process of this test very easy. All rules and scoring procedure is clearly written in the test manual. Reliability of this test was very high (.90) and validity of this test was .80 to .93.

Though this test is very popular and widely used, there are some limitations too. These are:

- Scoring conversion procedure is not effective in this test.
- Some test items are very old where revision is needed.
- This test is not based on any specific intelligence theory.

CHECK YOUR PROGRESS

Q 6: How many test items are there in Binet-Simon scale of 1905?

Q 7: Mention one advantage of latest revision of Stanford Scale of 2003. Q 8: Who is David Wechsler?

3.8 Uses of General Intelligence Tests in Education

In the field of education and psychology, intelligence test is widely used. Uses of intelligence test in education can be categories in the following way:

- 1) **Classification:** Intelligence test can be used to classify students on the basis of mental ability. Teaching-learning process faces lots of problems if different abilities of students are there in a single classroom. Teaching of this type of classroom is not effective for brilliant as well as dull students as teacher mainly concentrate on large number of average students. Therefore, intelligence tests are essential to classify the students on the grounds of mental ability.
- 2) **Selection:** Intelligence test is used to select efficient people for different type of tasks. Efficient students and teachers are to be select for admission in the institution, to provide instruction and guidance. Again for administrative works and co-curricular activities also different types of people are selected. Intelligence test is used to select these people for above mentioned tasks.
- 3) **Prediction:** Intelligence test is used for prediction also. This test provides prediction about the future achievement of the children. It also helps the students to select career for their future life as education is directly related with work life.
- 4) **Educational Guidance:** Intelligence test is also used for educational guidance. Intelligence test classifies students on the basis of their ability. This classification helps to provide educational guidance

according to their abilities. To have better idea about the capacity and abilities of the students, the guide or instructor widely applies different types of intelligence tests.

- 5) **Teachers' Self Analysis:** Intelligence test is also useful for self analysing and self evaluating the tasks of teachers. On the basis of the performance of the students teachers must change his/her teaching methods. So, in that case, intelligence test is important.
- 6) **Vocational Guidance:** Intelligence tests are widely used in vocational guidance also. As individual differences are there among all individuals, all are not equally capable for all types of vocation. Therefore, to select the best for vocation, intelligence tests are used. On the basis of the result of intelligence tests, vocational guidance is to be provided to the students.
- 7) **Research:** Intelligence tests are widely used in the field of research also. These tests are effectively used mainly in educational, psychological and sociological research. For solving different research problems of these areas, intelligence tests are used.
- 8) **Diagnosis:** To find out the educational problems of the students, intelligence tests are used. To find out the problems of students as well as to provide effective remedies, these tests are administered. In psychological laboratories, psychologists deliberately apply intelligence tests for diagnosis purposes.

For the above mentioned aspects, intelligence test is very essential. Now-a-days, standardised intelligence tests are widely used in teaching-learning situation.

3.9 Summing Up

- Intelligence is defined as the ability to learn, adjust and carry on abstract thinking.
- In the year 1905, Binet developed the most acceptable scale of measuring intelligence. He along with his colleague Simon developed an intelligence test in Stanford University of France.
- On the basis of its form or content, the intelligence test is divided into three types such as verbal intelligence test, non-verbal intelligence test and performance test; again on the basis of administration, it is

divided as individual intelligence test and group intelligence test. And on the basis of scoring, it is divided into age scale and point scale

- The first systematic scale for measuring intelligence was developed by Binet-Simon in the year 1905 which was later developed in the year 1908 and 1911.
- L.M. Terman and his associates re-standardised the Binet-Simon scale in Stanford University of the United States in the year 1916. This revision was known as Stanford Revision of Binet-Simon Scale.
- American psychologist David Wechsler developed Wechsler Bellevue Intelligence Test. This test was developed in the year 1939. Latest revision of this scale was published in the year 1981. In this scale, there were total 11 sub tests which were verbal and performance type in nature.
- Intelligence tests are widely used in education for classification, selection, prediction of students. These tests are also used for educational guidance and vocational guidance.

3.10 References and Suggested Readings

- Sharma, R. A. (2005). *Essentials of Measurement in Education and Psychology*. Meerut: R. Lall Book Depot.
- Goswami, Marami (2013). *An Introduction: Educational Measurement and Evaluation*. Guwahati: Ashok Book Stall.
- Rajat, D. S. (1970). *Measurement, Evaluation and Statistics in Education*. New Delhi: Raj Book Depot.
- Mangal, S. K. (2004). *Advanced Educational Psychology*. New Delhi: Prentice Hill of India Pvt. Ltd.
- Asthana, Bipin (2009). *Measurement and Evaluation in Education and Psychology*. Agra: Shri Vinod Pustak Mandir.
- Bhatia, K. K. (1995). *Measurement and Evaluation in Education*. Ludhiana: Prakash Brothers.

3.11 Answers to Check Your Progress

Answer to Q. No. 1: David Wechsler opined, "Intelligence is the aggregate or global capacity in the individual to act purposefully, to think rationally and to deal effectively with the environment."

Answer to Q. No. 2: 1984

Answer to Q. No. 3: Cattell

Answer to Q. No. 4: Wechsler Bellevue Intelligence test.

Answer to Q. No. 5: Two differences between age scale and point scale are:

- (i) Test items are prepared on the basis of age of the respondents in age scale; on the other hand in point scale, test items are prepared on the basis of measurement of tasks.
- (ii) Qualitative analysis is possible in age scale whereas quantitative analysis is possible in point scale.

Answer to Q. No. 6: 30

Answer to Q. No. 7: One advantage of Stanford revision of 2003 is that age group of the respondents range from 2 years to 85+ years.

Answer to Q. No. 8: David Wechsler was an American psychologist. He was basically a clinical psychologist in a hospital of Bellevue. He developed an intelligence test which was known as Wechsler Bellevue Intelligence Test.

3.12 Questions and Exercises

Short-Answer Questions

1. Who first used the term 'intelligence quotient'?
2. Write two merits of individual intelligence test.
3. Write two differences between verbal and non-verbal intelligence test.
4. What is age scale?
5. How can we classify students into different groups?

Long-Answer Questions

1. Explain the concept of intelligence.
2. Write an essay on measurement of intelligence.
3. Describe the classification of intelligence tests.
4. Explain the Binet-Simon scale of intelligence.
5. Discuss the Wechsler intelligence test.
6. Write the uses of intelligence tests in education and psychology.

— x —

UNIT: 4

APTITUDE TESTING

Structure

- 4.1: Introduction
- 4.2: Objectives
- 4.3: Meaning of Aptitude
- 4.4: Nature of Aptitude Test
- 4.5: Meaning of Aptitude Test
- 4.6: Uses of Aptitude Test
- 4.7: Types of Aptitude Tests
 - 4.7.1: Multiple Aptitude tests:
 - 4.7.1. (a): General Aptitude Test Battery (GATB)
 - 4.7.1. (b): Differential Aptitude Tests (DAT)
 - 4.7.1. (c): Test of Primary Mental Ability (PMA)
 - 4.7.2. Measures of Specific Aptitudes:
 - 4.7.2. (a) : Mechanical Aptitude Test
 - 4.7.2. (b) : Test of Clerical aptitude
 - 4.7.2. (c) : Musical and Artistic Abilities
 - 4.7.2. (d) : Measuring Aptitude for Art
 - 4.7.2. (e): Tests of Scholastic and Professional Aptitudes
- 4.8: Need and importance of Aptitude Test
- 4.9. Caution in the use of Aptitude Test Data
- 4.10: Summing Up
- 4.11: Answer to Check Your Progress
- 4.12: Unit End Activities
- 4.13: References and Suggested Readings

4.1: Introduction:

Generally, we observe that some persons are said to possess a certain specific ability in addition to intellectual ability. Because of these specific abilities, it is observable that, people differ from one another and within themselves in their performances in different fields of human activities. For example, we can find out that, Rupa profits from musical training while Smita almost having the same intelligence and getting the same training under similar conditions makes little or no progress in that area. The underlying cause of this difference is the musical aptitude of Rupa. This attitude is not ability rather a special ability. It helps to predict the probable development of certain abilities. Thus, by testing aptitude, we are able to know to a great extent whether an individual will become a good leader, a good musician or a good teacher. Aptitude can be measured through multiple evidence which includes past academic achievements of the pupil, occupational status of the parents, the observation of the teacher, the hobbies selected by the pupil, the interview, self-estimate by the pupil, a situation test and a standardized aptitude test. The most important of the above means is the aptitude test. Learners, in this unit we will discuss the tests used to measure aptitudes.

4.2 Objectives:

After reading this unit, the learner will be able to:

- *explain* the meaning and nature of aptitude
- *define* an aptitude test
- *classify* different aptitude tests
- *describe* multiple-aptitude tests with each example
- *describe* specialized aptitude tests with examples
- *discuss* the educational significance of aptitude tests
- *highlight* the needed cautions in the use of Aptitude Test Data

4.3: Meaning of Aptitude:

Before discussing aptitude let us discuss the meaning of aptitude. The word aptitude is derived from the word "Aptos" which means 'fitted for'. According to the Dictionary of Education, aptitude is a "pronounced innate capacity for or ability in a given line of endeavours such as a particular art, school subject or vocation." It is a condition, a quality, or a set of qualities in an individual of the probable extent to which he will be able to acquire some specific knowledge, understanding or skill in different activities under suitable training. Aptitude may be described as a special ability which helps an individual to acquire proficiency or achievement in a specific field. When we speak of an individual's aptitude for a given type of activity, we mean his

potentialities at present, as revealed by his performance on the selected test which has predictive value. When we say Reeta has an aptitude for music we mean that her present condition or ability reveals that there is a possibility of showing proficiency in music with proper training. Aptitude is something more than the ability. It is ability plus suitability of performance. A person may be a good scholar, because he possesses the high verbal ability, but not necessarily a good teacher.

An aptitude is not the same thing as ability or interest. It is more specific, measuring only certain aspects of functioning within a limited range. A person with a verbal ability cannot have aptitude for all the different tasks or vocations connected with verbal ability. The delimitation of the area is a result of the influence of the environment and personal cultivation for a period of time. Aptitude is a result of both heredity and environment. The primary mental ability which is a major factor herein is hereditary. The readiness to acquire knowledge or skill and the satisfaction in the activity is cultivated through the effect of the environment.

- The most appropriate explanation of aptitude is given by *Bingham*, which is included in Warren's Dictionary of Psychology. He defines aptitude as "a condition symptomatic of a person's fitness, one essential aspect is his readiness to acquire proficiency in his potential ability and another is his readiness to develop an interest in exercising his ability."
- *Freeman* has defined aptitude with the following words, "An aptitude is a combination of characteristics, indicative of an individual capacity to acquire (with training) some specific knowledge, skill, or set of organized responses, such as the ability to speak a language, to become a musician, to do mechanical work."
- According to *Warren*, "An aptitude is a condition or set of characteristics regarded as symptomatic of an individual's ability to acquire with training some knowledge and skill onset of responses such as the ability to speak a language or to practice music etc."
- According to *Traxler*, "Aptitude is a present condition which is indicative of an individual's potentialities for the future."

Thus, aptitude is an innate and inborn ability which finds expression through training in various fields. The knowledge of an individual's aptitude thus helps us to predict his/her future success in a particular field of activity, with appropriate training or experience.

Aptitude, in reality predicts what a person is fit to be:

- Achievement concerns the past, indicating what has been done.
- Ability concerns the present, indicating the powers now.
- Aptitude looks to the future, predicting what he may become.
- Aptitude is symptomatic or indicative of one's potentials.
- Aptitude opens the ways of interest and satisfaction in life,

4.4: Nature of Aptitude Test:

For a clear understanding of the true nature and concept of aptitude, the following points need to be mentioned.

- 1) Aptitude is not a general ability but a specific talent needed for different kinds of job.
- 2) It is an inborn and inherent quality which finds expression through practice and training in a particular area.
- 3) It is influenced by both hereditary and environmental factors. Hereditary provides the potential and environment helps to find expression to it.
- 4) It is a present condition but with a future reference.
- 5) It is symptomatic of one's ability for a particular job. It implies fitness and suitability for the activities in question.
- 6) Aptitude is the ability to learn a variety of skills or behaviours, not the type of skills learned. A person may have no present ability to sing but may have an aptitude for singing—which means that he has the capacity or ability to acquire proficiency in singing—provided he receives the proper training. Thus –
Aptitude = Ability + Probability of success.
- 7) An aptitude is not a unitary trait of human personality. It is pluralistic. For example, aptitude for science involves basic intellectual qualities like logical reasoning, abstract reasoning, arithmetical reasoning, certain temperamental qualities like interest in experimentation and initiative for invention, personality characteristics like persistence and hard work. Similarly, in order to be a successful architect, one must possess the cluster of abilities such as a keen sense of observation, a sense of aesthetic visual memory, abstract reasoning, and an ability to sketch free hand.
- 8) Aptitudes are not rigidly constant. They are somewhat variable and are affected to a certain extent by education and environment. But

a complete reversal in aptitude within the life span of an individual is practically unknown. An individual who is definitely superior in a certain kind of aptitude at the age of 6, will, as a rule, be above average in that kind of aptitude later.

- 9) Aptitude may be defined as an inherent or native ability that can be developed to its maximum through learning or other experiences. However, it cannot be expanded beyond a certain point, even by learning.

To sum up, aptitude is thought of as a natural tendency, special ability, or capacity or cluster of abilities. But it cannot appear or find expression unless the favorable or suitable opportunity is provided in the environment. It is the present ability of an individual to perform special work successfully.

Check your progress

1. Define aptitude.
2. Differentiate between aptitude and ability

4.5: Meaning of Aptitude Test:

An aptitude test measures the potential of a person to achieve in a given activity or to learn to achieve in that activity. Aptitude tests measure the degree or level of one's special ability in the same way as intelligence tests are employed for measuring one's general mental ability. The goal of aptitude testing is to predict what can be learned in the future. They are chiefly used to estimate the extent to which an individual would profit from a specific course or training. For example, a clerical aptitude test would be able to determine whether an individual would do well as a clerk after appropriate training and with the right motivation.

- ***According to Hull***, "An aptitude test is a test designed to discover what potentiality a given person has for learning some particular vocation or acquiring some particular skill."
- ***In the words of Bingham***, aptitude tests do not directly measure future accomplishment. They make no such pretense. They measure the present performance. Then, in so far as behaviour, past and present, is known to be symptomatic of future potentialities, the test data supply a means of estimating those potentialities. The estimate is necessary for terms of probabilities only.

So aptitude test or tests of special ability predicts whether an individual will be successful in a particular area or field, if trained.

4.6: Uses of Aptitude Test:

Aptitude tests may potentially be used.

- 1) To provide information to assist an individual in making educational and career decisions from competing alternatives
- 2) To help students in making choices relating to their future occupation according to their aptitude such as: teaching, social work, military etc.
- 3) To assist students in the proper choice of streams such as Science, Arts, and Commerce.
- 4) To select students for admission in colleges and technical institutes.
- 5) To help the counselor to render appropriate and exact counseling service according to aptitude.
- 6) To group together individuals with similar aptitudes for development and other educational purposes.
- 7) To help the school authority to reduce the number of probable failures.
- 8) To select workers for particular jobs
- 9) To supplement other psychological tests.
- 10) To encourage the development of special or potential abilities of an individual which the individual is not aware of.

Check your progress:

1. What is an aptitude test?
2. Mention three uses of aptitude test

4.7: Types of Aptitude Tests:

There are different types of aptitude tests. Test constructors have typically used two approaches of aptitude test. The first approach is to develop integrated batteries that measure a number of relatively broad abilities (usually six to twelve). This is a general aptitude test or multiple aptitude test approach. The other approach to aptitude testing is to develop individual tests for specific abilities. Thus, the two types of aptitude tests usually employed in psychological testing are—

- (a) Multiple Aptitude Batteries or General Aptitude Tests and
- (b) Specialized Aptitude Tests.

4.7.1: Multiple Aptitude Tests:

Multiple aptitude tests batteries are designed for the independent measurement of several kinds of mental operation and to provide a separate score for each. On the theoretical level, multiple aptitude approach represents the adoption of a group factor approach to abilities. On the practical level,

this approach recognizes that testing time will always be limited and those specific abilities can be clustered into groups. If there are a number of relatively broad, more or less independent, abilities, and if we want a relatively complete picture of an individual's abilities then a number of independent tests will be needed. So, a test battery must contain as many tests as their abilities to be measured.

A multiple aptitude test battery is used to assess the suitability of persons for different professions on the basis of scores in the relevant aptitude tests in the battery. Instead of a total score, they provide a suitable instrument for making intra-individual analysis through sectional scores. Most of the batteries of tests available for assessment of aptitude at school stage are in the form of test batteries consisting of the underlying abilities required for success in different occupations rather than a direct assessment of job aptitudes. Thus, an individual's potentialities in a number of academic and vocational areas can be analyzed by means of a series of relatively brief tests. Some examples of multiple aptitude test batteries are –

- The General Aptitude Test Battery (GATB),
- Differential Aptitude Test (DAT),
- Primary Mental Abilities (PMA)

4.7.1. (a): General Aptitude Test Battery (GATB) :

This battery was first published in 1947 by U.S. Employment service. This battery consists of 12 tests selected to measure nine aptitudes important for success in a wide variety of occupations. The factors covered by the GTAB are as under—

- (a) **Intelligence (G):** It is measured by the combined scores on three tests. Three-dimensional space, vocabulary, and arithmetical reasoning.
- (b) **Verbal Aptitude:** It is measured by vocabulary tests of synonyms and antonyms.
- (c) **Numerical Aptitude:** It includes both computational and Arithmetic Reasoning tests.
- (d) **Spatial Aptitude:** It is measured by three Dimensional space test, involving the ability to comprehend two-dimensional objects and to visualize the effect of movement in three dimensions.
- (e) **Form Perception:** It tests the ability to match drawing of tools in one part, and of geometric forms in another.
- (f) **Clerical Perception:** It requires matching of names.
- (g) **Motor Coordination:** It is measured by a simple paper-pencil test, requiring the subject to make specified pencil marks in a series of squares.

(h) *Finger dexterity*: It tests proficiency in assembling and disassembling rivets and washers.

(i) *Manual Dexterity*: It measures hand dexterity by using both hands and then the preferred hand, in placing pegs in a pegboard.

Two of the nine factors measured by the Tests— Finger Dexterity and Manual Dexterity- utilize performance tests. The other seven factors are measured by conventional paper and pencil tests. The test requires approximately 2 and 1/2 hours of testing time. Raw scores are converted into standard scores for various aptitudes, measured by the tests. Norms for different groups have been prepared. Individual profile records are available for recording standard scores. It is a very good battery for vocational guidance of youth and adults.

4.7.1. (b): Differential Aptitude Tests (DAT):

Differential Aptitude Tests (DAT) have been prepared by George K. Bennett, Harold G. Seashore and Alexander G. Wesman and published by the U.S Psychological Corporation in 1947. Another revision of the test was published in 1972 (Form S and T). This was a comprehensive battery. It was developed principally for use in educational and vocational counseling of high school students. It can also be used for unselected adults. The DAT is probably the most widely used multiple-aptitude battery in the secondary school level. The battery consists of eight subtests of abilities required for success in different occupations. All eight tests of the battery were standardized on the same population. The battery includes the following tests:

(1) *Verbal Reasoning (VR)*: This test consists of verbal analogies. It is designed to measure the subject's ability to handle verbal concepts and relationships rather than his word knowledge. Thus, verbal reasoning utilizes a special form of analogical items that require students to use reasoning rather than mere associations.

(2) *Numerical Ability (NA)*: This test is designed to measure the understanding of numerical relationship and facility in handling numerical concepts. Items are essentially computational, rather than problem-solving type. Some of the items test only proficiency in four fundamental processes while others require an understanding of quantitative concept and relationships.

(3) *Abstract Reasoning (AR)*: This test attempts to measure reasoning without the use of words.

(4) *Spatial Relations (SR)*: This is the most ingenious test in the series. This test measures the ability to visualize mentally or imaginatively manipulate concrete materials. The test presents two-dimensional geometric figures

(variously shaded). The examinee selects the three-dimensional figures that can be made from the two-dimensional figures.

(5) Mechanical Comprehensive (MC): The test measures understanding of mechanical principles by means of a series of pictorially presented situations encountered in everyday life, involving mechanical and scientific principles. Thus, each item includes a picture illustrating a mechanical situation.

(6) Language Usage Part I: It is a spelling test in which the subject marks the incorrectly spelled words in a list.

(7) Language Usage Part II: It consists of sentences in which the examinee is required to distinguish faulty from correct grammar, punctuation, and word usage.

Language usage part I and part II are both achievement tests. These are included in the test because they represent basic skills that are necessary for so many academic and vocational pursuits.

(8) Clerical Speed and accuracy: Clerical Speed and accuracy test is the only test in the battery, placing primary emphasis on speed (the speed of performing a single perceptual task is measured). The subject is instructed to locate identical combinations of letters or numbers as quickly as possible. The items are intended to "approximate elements involved in many clerical jobs."

Three of the tests - verbal reasoning, numerical reasoning and abstract reasoning - measure broad intellectual abilities. The items are similar to items on general intelligence and scholastic aptitude and materials commonly taught in U.S. schools are used. Administration and scoring procedures are simple and straightforward. The authors recommend administering the entire battery of eight tests so that the pattern of scores can be considered when counseling the student. Total testing time is slightly over 3 hours, with time limits varying from 6 to 30 minutes per test. Administration of total battery can prove to be costly in terms of time but one can make selective use of certain subtests. For example, a student trying to explore whether s/he will have the required aptitude to go to engineering, may not be required to take tests like clerical speed, language usage, grammatical or verbal reasoning tests but may be required to take numerical abstract and spatial reasoning tests. Reliability of the test is high. Validity data are abundant. Normative and interpretive data are excellent.

The combined verbal reasoning and numerical ability scores provide an index of scholastic aptitude. In many ways, the DAT procedures are a model for interpretation of test scores. Normative data are presented for large representative samples of high school students with separate norm tables available for each grade (8-12) and sex grouping. Percentile equivalents of raw scores are provided for each group on each test. Percentile

scores are used to plot an individual profile on a normal percentile chart. Such a profile shows the subjects relative standing on different tests, all scores being expressed in comparable units and with reference to a common norm.

4.7.1. (c): Test of Primary Mental Ability (PMA):

Primary Mental Abilities is based on the theory given by Thurston. He arrived at this theory through the method of Factor Analysis. This analysis had led to the identification of certain mental abilities by the American Psychologists. Primary Mental Abilities was the first multifactor battery. Originally, it was known as the Chicago Tests of Primary Mental Abilities published by the American Council on Education in 1941-47. Another revision of the test was published by the Science Research Association. So this addition was named the SRA Primary Mental Abilities. In this revision, there were three scales: one for ages 5-7, another 7-11 years and the third one for 11-17 years. The current revision has five levels, with four or five tests at each grade level. The factors measured by the test include—

- 1) **Verbal reasoning:** the ability to understand ideas expressed in words or ability to deal with words and verbal concepts.
- 2) **Number facility:** facility in doing simple numerical tasks.
- 3) **Spatial relations:** ability to visualize and conceptualize. The test items are various simple designs and geometric figures, differently rotated.
- 4) **Reasoning:** the ability to think logically.
- 5) **Perceptual speed:** the ability to distinguish sizes and shapes.
- 6) **Word fluency:** the ability to produce words easily.

The verbal reasoning, number facility, and spatial relations test reasoning at older age levels. The tests are designed to measure readiness to learn both overall and in specific courses.

Check your progress:

1. What are the two types of aptitude tests?
2. What Multiple Aptitude tests?
3. What are the benefits of using aptitude test?
4. Write the names of three aptitude test battery.
5. Mention three subtests included in GATB.
6. How many subtests was included in DAT
7. Write the name of the only speed test of DAT

4.7.2: Measures of Specific Aptitudes:

In this approach, the test constructor has concentrated on measuring a specific ability rather than attempting to provide a broad picture of abilities. Multiple aptitude batteries measure abilities within the context of the single, integrated test battery. But specific aptitude tests measure the aptitudes of individuals in various specific fields of activities. These single aptitude tests include tests of mechanical aptitude, clerical aptitude, teaching aptitude, musical aptitude and so on. Such tests covering a group of related abilities necessary for performing in an occupation like becoming a mechanic, clerk, teacher, musician etc. Generally, these tests can be divided into the following sub-types according to the specific aptitude tested by these tests—

- (a) Mechanical Aptitude.
- (b) Musical and Artistic Aptitude.
- (c) Professional aptitude tests i.e. test to measure the aptitudes for a profession like teaching, clerical ability, medicine, law, engineering salesmanship, research etc.
- (d) Scholastic aptitude i.e. tests to measure the aptitude for different courses of instruction etc.

Some of them are discussed below—

4.7.2. (a) : Mechanical Aptitude Test:

Mechanical ability is an ability involved in manipulating concrete objects, such as tools and in dealing mentally with mechanical movements. It is not a single unitary function. It is a comprehensive form of the psycho-physical ability of an individual. It is a combination of sensory and motor capacities plus abilities such as manual dexterity, the perception of spatial relations, capacity to get mechanical knowledge, comprehension of mechanical relation, mechanical reasoning and problem-solving. Tests of mechanical aptitude are designed to measure the capacity and performance of a higher level of organization than those of sensory-motor capacity and dexterity (Freeman). A number of tests are available for measuring mechanical aptitude. Some of the popular tests may be described below.

(a) Stenquist Mechanical Aptitude Test :

Stenquist Mechanical Aptitude test was developed by S.L. Stenquist in 1923. It was intended to measure a person's ability to put together the parts of mechanical devices like a bicycle bell, a door lock, a mousetrap etc. The test was constructed for use among individuals covering the age range from children in the lower grades through adulthood. The test consists of three series of items. There are ten test items in each series. Total time allotted for the test was 30 minutes.

(b) Minnesota Mechanical Aptitude Test:

The Stenquist tests have been revised and extended at the University of Minnesota (1930) in America by G.D. Peterson. It was known as the "Minnesota Mechanical Aptitude Test." This test consists of the following four sub-tests —

- **Minnesota Mechanical Assembly Test:** These are essentially the same as Stenquist's tests, with new ones added. Thirty-three disassembled, mechanical contrivances are used. The test assessed a person in terms of rate and accuracy of work. The test has been found useful in predicting the success of junior high school boys in shop courses as well in a number of occupations such as mechanic and auto-mechanic, woodworker, tool-maker, ironworker etc.
- **Minnesota Paper Form board:** The test measures one's capacity to visualize and imaginatively manipulate geometric forms and to discriminate geometric patterns. It presents 64 problems. Each problem is composed of two or more parts of a geometric figure printed on a form-board. These parts when correctly assembled make a complete figure. The subject is required to identify the correctly assembled figure among several choices.
- **Minnesota Spatial Relations Test:** This test was designed in 1930. The test consists of a set of four standard form boards each of which has 58 cutouts of different forms and sizes. The subject is to make the proper placing of the cutouts of the board in their correct hole in the board.
- **Minnesota Interest Analysis Test:** Interests are significant for satisfaction and success in vocational fields also. Thus, this test is included in the battery to measure interest from a mechanical point of view. The test is used to measure speed and accuracy in discriminating sizes and shapes. It is suitable for testing adults as well as pupils in the upper elementary grades and in secondary schools. But it cannot measure the mechanical nature of problem-solving and capacity to manipulate small objects with precision.

(a) Bennett Mechanical Comprehension Test:

This test of mechanical comprehension presents a mechanical problem in the pictorial situation. This test is designed to measure understanding of the operation of mechanical principles in everyday situations. In this test, pictures and sketches are presented to the subject. Every picture provides a mechanical problem which demands the ability of comprehension of the picture situation. The subject is to answer certain questions arising out of the pictured situation. This test has been designed with three levels of difficulties——

- One form is designed for use with high school students, engineering school applicants and with relatively untrained and inexperienced persons.
- The second form is meant for use with engineering school candidates and
- The third form was devised for high school girls and women.

On the whole, the content of mechanical aptitude tests sample the following functions—visual motor integration, spatial visualization, perceptual speed, manual dexterity, visual insight etc. Besides, some tests measure specialized information, knowledge of some particular technique, arithmetical problem-solving ability, and technical vocabulary. Some of the functions are measured by means of apparatus tests, others by means of performance type materials and some others by means of paper and pencil tests.

4.7.2. (b): Tests of Clerical Aptitude:

Clerical aptitude like mechanical aptitude is not a unitary function. Different opinions are held regarding the term clerical aptitude. Super opines that it refers to the ability of routine clerical work. Bills points out clerical duties “include the gathering, classification, and presentation of data of all sorts and analysis and use of these data in planning, executing and determining the results of operation.”

According to Bingham, clerical aptitude indicates at least four different abilities as its components. These are——

- (a) **Perceptual ability:** Ability to observe words and numbers with speed and accuracy.
- (b) **Intellectual ability:** Ability to grasp the meaning of words and symbols and to make the correct decision regarding the questions they raise. This intellectual ability becomes increasingly indispensable as the clerical work increases in complexity.
- (c) **Various Mental Skills:** Ability to add and multiply, to spell correctly, to punctuate, and to use a wide variety of words and expressions.
- (d) **Motor Ability:** Ability to use various types of machines and tools like a typewriter, duplicator, cyclostyle machine etc.

For measurement of these abilities, the test materials and their functions include correct and systematic arrangements of symbols, numbers, words, letters and names, simple tests of arithmetic etc. For effective identification and measurement of clerical aptitude, test materials are devised according to the nature of the clerical jobs itself. Reading, writing, shorting, checking arranging, filling, stamping, coding etc. are the activities involved in this type of test. A number of tests are available for measuring clerical aptitude. Some popular clerical aptitude tests are——

(a) Minnesota Clerical Test:

This test measures perceptual speed and accuracy needed in the clerical job. This test was developed by M. Andrew. The test is useful for distinguishing those who shows promise for clerical work. There are two sub-tests in the test -

(a) Number comparison: In this test the examinee is given 200 pairs of numbers, each containing 3 to 12 digits. The numbers of each pair being exactly alike in some cases and differing very slightly in other cases. The subject is required to check the pairs which are alike.

(b) Name comparison: In this subtest, there are 200 names instead of numbers. The task is similar in both tests.

A very low score on the test indicates a lack of clerical aptitude, but a high score does not necessarily show a positive aptitude for clerical occupations, especially those involving responsibilities and variety of works. In such cases, the test result should be interpreted by considering the evidence from other sources.

(c) General Clerical Test (GCT):

This test has been designed as a general differential test for use with all types of clerical workers. This test aims at measuring three kinds of abilities considered necessary for clerical work. They are (1) Clerical speed and accuracy (2) Numerical ability (3) Verbal facility.

Its scope is wider than Minnesota clerical test. There are 9 sub-tests in this test with a working time of 40 minutes. These are—(1) Copy correction (2) Alphabetical filling (3) Arithmetic computation (4) Checking for errors in simple addition (5) Arithmetic reasoning (6) Spelling (7) Reading (8) Vocabulary (9) Grammar.

The test yields a total score as also subtest scores for clerical numerical and verbal ability. In short, clerical aptitude tests predict success in clerical and inspecting jobs, particularly routine ones. These tests might be called tests of perceptual speed and accuracy since they assess the ability to note details quickly and accurately.

4.7.2. (c) : Musical and Artistic Abilities:

These tests have been devised for discovering musical talents. Standardized tests of both musical abilities generally assess only component skills rather than attempt to measure complex performance skills. But a relatively small number of tests are available for assessing aptitude in music. Some of the important musical aptitude tests are described below.

(a) Seashore Measures of Musical Talents:

These measures are based on a scientific analysis of musical hearing appreciation and performance. It is intended for use from grade 4 through the college level. It measures skills such as discrimination of pitch, rhythm, timbre loudness discrimination and tonal memory.

- (a) Judgement of Rhythm:* The rhythm test requires the subject to compare two rhythm patterns and judge whether they are similar or different.
- (b) Intensity or loudness discrimination:* Judging loudness of pairs of notes with gradation differences.
- (c) Time discrimination:* Judging the duration of some notes which are played a second time.
- (d) Tonal memory:* In this test, paired tonal patterns are played to test the ability to perceive the difference.

The test items in the battery are presented on phonograph records. The subject listens and attempts to discriminate. He is required to mark his responses on an answer form supplied to him. The test is considered unfit for testing musical talents of little children. It is also incapable of measuring all components of musical aptitude, as it measures the auditory aspect.

(b) The Wing Standardized Tests of Musical Talent:

This test is intended for use for the ages of 8 years onwards. The wing tests measure the ability for perceiving musical relationships and aesthetic choices. It has seven aspects – chord analysis, pitch change memory, rhythmic accent, harmony, intensity and phasing. The wing tests are considered to be more relevant to the selection and training of individuals in that art than the Seashore tests.

4.7.2. (d) Measuring Aptitude for Art:

Art aptitude is a complex mixture of interest, abilities, energy, perseverance traits, and other uncommon factors. Relatively, a small numbers of tests are available for assessing graphic arts aptitude. Two important tests of art aptitude are ——

(a) The Meier Art Judgement Test:

It is intended to measure aesthetic judgement in a "global" manner. It consists of 100 pairs of representational pictures in black and white. One picture of each pair is a reproduction of a recognized masterpiece. The second one of the pair is alike in every respect, except one, and this one differentiation involves some aesthetic quality or principle, basic in all graphic art. Each individual is required to indicate his preference in each pair. The test has been standardized for junior high school, senior high school and college or adult levels.

(b) Graves Design Judgement Test:

The test contains exclusively abstract designs. It has 90 items — 8 contain three designs each and the rest contain only two. The designs are executed in black, white and grey. Some are line drawing, others are composed of squares, triangles, similar two-dimensional figures, and some others look like a reproduction of three-dimensional abstract sculpture perceptual function, rather than to make a separate analysis of elements that enter aesthetic judgment.

4.7.2. (d): Tests of Scholastic and Professional Aptitudes:

Many scholastic aptitude tests have been developed for the selection of students for admission to specific courses or professions like engineering, medicine, law, teaching etc. Examples of these tests are —

- (a) The American Council of Educational Scholastic aptitude test
- (b) Scholastic aptitude tests in the U.S.A.
- (c) Moss scholastic aptitude test for medical students.
- (d) Law School Admission Test.
- (e) Pre- Engineering Ability Test.
- (f) Minnesota Engineering Analogies Test.
- (g) Teaching Aptitude test by Moss etc.

A particular specific aptitude test is usually developed for a specific purpose or situation. The main advantage of specific aptitude test is that the ability can be measured more precisely. By concentrating on a single ability one can probe its various aspects in more detail. It is, therefore, a more valid indicator of that particular ability.

But this approach has some limitations also. It may lead to over specificity and lack of generality. Thus, if a test user wishes to measure several abilities, he will have to develop his own test battery. This may require developing a new battery for each situation. Since different tests will have been developed on different populations, normative data will not

be comparable. So there will be interpretation problems in comparing the scores of various measures.

Check your progress:

Write the names of two –

- (a) Mechanical Aptitude Test (c) Clerical Aptitude Test
(b) Musical Aptitude Test (d) Scholastic aptitude Test

4.8: Need and Importance of Aptitude Test:

Applicability of aptitude and its testing is quite obvious in the modern world. It is asserted that aptitudes are related to vocational success as intelligence is related to success in life in general. The need and importance of aptitude testing in educational and vocational guidance have been increasing day by day. Its need and importance have been felt on the following grounds:

- 1) Aptitude tests are helpful for identifying abilities of an individual. Their innate special inborn aptitude should be identified and necessary training should be given for use of manpower resources in today's world of specialization. Students have to make a decision regarding the choice of subjects in the high school study as the various types of courses need specific nature of abilities or aptitudes.
- 2) Vocational guidance largely depends on the use of these different kinds of aptitude tests. Students have different types of aptitude needed for different kind of vocations. It is necessary to identify the nature of their aptitude to which they are best fitted. Differential Aptitude test, Clerical Aptitude test, Primary Mental ability test etc. can be effectively used to identify specific aptitude of the secondary school students for a specific vocation.
- 3) Every individual is not fitted to any and every kind of job. An employer can use these tests for vocational selection. With these tests, he can reduce the number of probable failures among those who take up that job.
- 4) There are certain specialized courses such as Law, Medicine, and Engineering etc. where the selection of the trainees may be considered very essential before training is given to them. Hence, students can be selected for admission to technical or professional courses on the basis of aptitude test results. They may encourage the development of special or potential abilities of a given individual

It is thus clear that aptitude testing should get due importance for any purposeful guidance and counseling program to specialized academic and professional courses or selection for specialized jobs.

4.9. Caution in the use of Aptitude Test Data:

Aptitude is not the sole factor for success in a course or career. An aptitude test gives us a prediction of the probability only. There is never any certainty that predictions will always come true. The physical, social and emotional environment of the pupil is often different from the actual environment in the job or in college. This reduces the predictive value of aptitude test. Considering these limitations of aptitude tests, teacher and counselors select and use aptitude tests carefully, keeping in view the following aspects -

- 1) Aptitude test data must be used along with previous achievement data, present interests, leisure time activities, and work habits etc. in order to guide the students into various occupational possibilities.
- 2) The aptitude test samples certain abilities of the individual and helps to find what he/she can do now and how well it can be done further. On the basis of present performance, aptitude test estimate of future achievement may be inferred in terms of probability only.
- 3) It is worth noting that aptitude tests score only provide suggestions in career planning and do not automatically match individuals to suitable courses or studies, and occupations. There are factors like willingness, hard work etc. that contributes to one's success in a career which are not measured by aptitude tests.

To sum up, aptitude tests, coupled with other criteria like I.Q., interest, aspirations, personality traits etc. can help to a greater extent in avoiding wastage of human potential as well as material resources by placement of individuals in places and lines of work in which they are best fitted.

4.10: Summing Up

Aptitude is an innate and inborn ability which finds expression through training in various fields. The knowledge of an individual's aptitude thus helps us to predict his/her future success in a particular field of activity, with appropriate training or experience. Aptitudes are related to the professional and vocational success of an individual. There are different types of aptitude tests. Generally, two types of aptitude tests are usually employed in psychological testing: (a) Multiple Aptitude Batteries or General Aptitude Tests and (b) Specialized Aptitude Tests. Multi aptitude batteries measures abilities within the context of a single, integrated test battery. The objective is to obtain a number of separate scores in order to differentiate among several abilities within each individual. The General Aptitude Test Battery (GATB), Differential Aptitude Test (DAT), Primary Mental Abilities (PMA) are some examples of such test batteries. Such test batteries can give comprehensive information about the relative picture of the student's specific abilities. But specific aptitude

tests have been devised to measure the aptitudes of individuals in various specific fields of activities like: (1) Mechanical Aptitude. (2) Musical and Artistic Aptitude. (3) Professional Aptitude (4) Scholastic Aptitude. Aptitude test should be used with some special cautions.

4.11: Answers to Check Your Progress:

Check your progress-1

- (a) An individual's potentials are not equally strong. Individuals differ from one another in their potentials.
- (b) Inborn aptitudes find expression through training. They are influenced by both hereditary and environmental factors. They are specific in nature. It includes fitness and suitability. They can be pluralistic and not rigidly constant.
- (c) Aptitude is something more than the ability. It is ability plus suitability of performance.
- (d) $\text{Aptitude} = \text{Ability} + \text{Probability of success}$

Check your progress-2

- (a) An aptitude test, by definition, is any type of assessment that evaluates the talent/ability/potential to perform a certain task, with no prior knowledge and/or training.
- (b) Multiple aptitude test batteries are intended for the independent measurement of each of several kinds of mental operation and that provides a separate score for each. On a theoretical level, multiple aptitude approach represents an adoption of a group factor approach to abilities
- (c) The General Aptitude Test Battery (GATB), Differential Aptitude Test (DAT), Primary Mental Abilities (PMA)
- (d) Intelligence (G)/ Verbal Aptitude/ Numerical Aptitude/ Spatial Aptitude
- (e) Eight tests
- (f) Clerical Speed and accuracy
- (g) Thurstone

Check your progress-3

- (a) Bennett Mechanical Comprehension Test/ Minnesota Mechanical Aptitude Test / Minnesota Paper Form board
- (b) Minnesota Clerical Test/ General Clerical Test (GCT)
- (c) Seashore Measures of Musical Talents/ Wing Standardized Tests of Musical Talent

(d) Scholastic aptitude tests in U.S.A./Moss scholastic aptitude test for medical students.

Check your progress-4

- (a) Helpful for identifying abilities/ to help the students in the proper choice of courses, subjects and careers / used for the purpose of prediction for future success in careers etc.
- (b) Do not automatically match individuals to suitable courses or studies and occupations

4.12: Unit End Activities

1. What do you mean by Aptitude? What are the uses of aptitude tests? Show your acquaintance with any one of the tests of aptitude.
 2. What is meant by special ability? Give an account of any test that is used for assessing the mechanical ability of a student.
 3. What are aptitude tests? Discuss the value of aptitude tests in the field of education.
 4. What is aptitude test? Describe any one of the multiple aptitude test batteries.
- (5) Write notes on:
- (a) DAT (b) SRA: PMA
(c) Musical Aptitude test (d) Clerical Aptitude test.
6. What are the factors involved in the clerical aptitude? Which tests can be used for measuring clerical aptitude?
 7. What tests can be used for measuring art aptitude? Discuss.

4.13: References and Suggested Readings

- Anastasi, A. *Psychological Testing 4thed.* New York: Macmillan Publishing Co. Inc.
- Anderson, Scarvia B., Samuel Ball, Murphy, R.T. and et.al. *Encyclopedia of Educational Evaluation.* Jessey:Bass Publishers, 1976.
- Asthana, B. *Measurement and Evaluation in Psychology and Education.* Agra: Binod Pustak Mandir.
- Brown, Frederick. G. *Principles of Educational and Psychological Testing.* Holt: Rineheart and Winston, 1976.
- Goswami, M. *Measurement, and Evaluation in Education and Psychology.* Hyderabad: Nilkamal Publishing House, 2010

- Gronlund, Norman, E. *Readings in Measurement and Evaluation, Education, and Psychology*. The Macmillan Company, 1968.
- Freeman, F.S. *Theory and Practice of Psychological Testing*. Oxford and IBH.
- Mehrens, William, A., and Lehmann, Irvin, J. *Measurement and Evaluation in Education and Psychology*. Holt, Rinehard and Winston, 1978.

—x—

UNIT-5

EDUCATIONAL ACHIEVEMENT AND ITS MEASUREMENT

Contents:

- 5.0 Introduction
- 5.1 Unit Objectives
- 5.2 Educational Achievement and Its Measurement
- 5.3 Objectives of achievement Test
- 5.4 Construction of Achievement Test
- 5.5 Diagnostic Test and Performance Test
- 5.6 Essay and Objective Type Tests
- 5.7 Summing Up
- 5.8 References and Suggested Readings

5.0 Introduction

Educational achievement is an important aspect of broad educational process. Educational achievement means a student's performance in different subjects, which is generally measured at the end of academic year. In the present educational scenario, student's performance/achievement in course subjects is the main concern for students, parents and teachers. Academic achievement is important to students because it helps them to understand their level of performance. Based on it they can decide the future, if higher their academic achievement then it is opening wider prospect in job sector or to pursue better line courses and placement in better jobs; on the other hand if it is lower performance then limited scope in further studies and job placement. The academic achievement, however, it also indicates teacher's level of intelligence and effectiveness as well as overall effectiveness of the institutions. Thus academic achievement can be said as a "quantitative expression on student's accomplishment, in both quantity and quality, in course subjects at the end of the course". This accomplishment is measured through the test devices prepared either by concerned subject teachers or experts.

5.1 Objectives:

From this unit you will be able to:

- *define* educational achievement test and its measurement
- *define* the principles and steps of test construction
- *understand* the different types of achievement test
- *describe* and differentiate between performance and diagnostic test

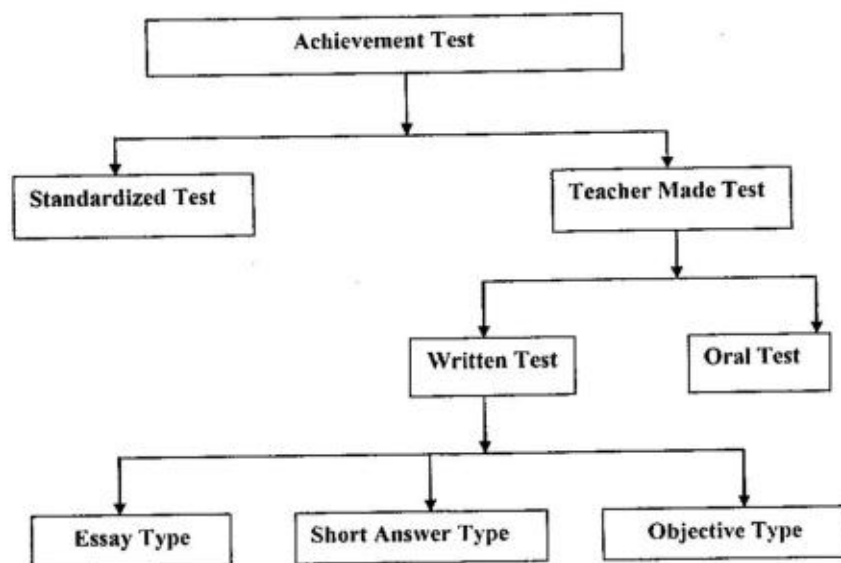
5.2 Educational Achievement and Its Measurement:

Achievement means one's level of learning attainments, accomplishments of instructional objectives, proficiencies acquired, etc. It is related to the students' growth and development in educational process. So, an achievement test is constructed to know the effect of particular educational instructions or training imparted to students' in a classroom. In our educational system there are mainly three broad aims, i. e. development of cognitive, affective and psycho-motor aspect. While we say achievement test it generally refers to measurement of cognitive aspects which has six sub areas, i. e. knowledge, understanding, application, analysis, synthesis and evaluation. This cognitive domain is tried to develop among students' through classroom transaction. So, Freeman defines a test of educational achievement as a test- "designed to measure knowledge, understanding, and skills in a specified subject or group of subject". Non-cognitive achievement is not given importance in achievement test. Further and in broader sense achievement test focus is laid upon child's performance and an achievement, learning outcomes, behavioral change and mental development of students.

Educational measurement of achievement has two main functions-

1. **Prognosis function:** It measures how much students have learnt from the taught areas or after the end of the session.
2. **Diagnosis function:** It measures how much students lack in the areas taught after the end of the session.

Teaching and testing the students are the main responsibilities of a teacher. So, in testing achievement of students main emphasis is given on content coverage or course. In performing achievement test a teacher may apply standardized test or teacher made test. These two tests are further divided as given following:



Principles of Test Construction:

1. The content should be cohesive and well integrated.
2. Weightage to content, objectives, forms of questions should be kept in mind.
3. A test should be clearly defined the learning outcomes.
4. A test should be concerned with all intended learning outcomes.
5. A test should measure a representative sample of instructionally relevant learning tasks.
6. A test should include all types of test items that are most appropriate for measuring the intended learning outcomes.
7. A test should be based on plans for using the results.
8. A test should provide scores that are relatively free from measurement errors.

Criteria of Achievement Tests (It detailed in Unit-II)

1. Reliability
2. Validity
3. Ease in administration
4. Low cost
5. Moderate time consuming
6. Acceptability by the students

7. Objectivity
8. Equilibrium
9. Specificity
10. Precision and clarity of tests items

STOP TO CONSIDER

- Educational achievement test is a design to measure knowledge, understanding, skills, etc. in a specified subject or a group of subjects.
- Educational achievement test are of two types teacher made test and standardized test.
- Teacher made tests are again subdivided into Essay type test and objective type test. Besides these objective type test is divided into short answer type and long type test.

CHECK YOUR PROGRESS-1

- Define educational achievement test.

5.3 Objectives of Achievement Test

The important objectives of achievement tests are mentioned below:

- To measure knowledge and achievement of students.
- To measure efficiency of teachers.
- To predict future performance of students.
- To develop patience and submission among students.
- To motivate students to study and work.
- To aid teacher in appraising his own teaching.
- To help in assessing the all round mental ability of the child.

STOP TO CONSIDER

- To motivate students to study and work.
- To predict future performance of students.
- To help in assessing the all round mental ability of the child.

CHECK YOUR PROGRESS -2

- Mention two important objectives of educational achievement test.

Standardized Test

A standardized test is one which is designed to measure what 'most' teachers cover, that is certain kind of common content areas. It is constructed by test construction specialists, usually with the assistance of curriculum experts, teachers, and administrators for the purpose of determining a student's level of performance in comparison to the performance of other students of similar age and grade. These tests often take long time to construct. It's called standardized because they are administered and scored according to specific and uniform procedures. A standardized test is a test which has some special qualities like reliability, validity, norms and scoring procedure; besides it must also have the standardized context and questions. The method of administration, scoring and interpretation should also be standardized. It comprises careful selection of items, having been given to large number of samples under standard conditions and for which norms have been established after careful evaluation. Standardized tests are quantitative in nature and are constructed by following systematic procedure. Contrary to that, teacher made test is a product of teachers for own class room efforts. Whereas, standardized tests deal with larger segments of knowledge or skills than teacher made test; even possible effort is made to make the test highly valid, reliable and discriminating. *(Details has been discussed in Unit-II)*

Some of important Standardized Achievement Test:

Following are some important standardized test batteries:

- **Stanford Achievement Tests:** This test was first published in 1922 and was revised in 1929. This test is used in American schools to measure academic knowledge of elementary and secondary school students. It consists of nine tests and takes 227 minutes. Norms have been provided according to age and percentile. The nine tests are- Paragraph meaning, word meaning, spelling, language, arithmetic computation, arithmetic reasoning, social studies, science and study skills.
- **Metropolitan Achievement Test:** The advanced battery for grades seven through nine consists of ten tests and takes 225 minutes. It revised edition was published in 1931 and then in 1937. Grade, age and percentile norms are provided. Ten tests included are-reading, vocabulary, literature, arithmetic fundamentals, arithmetic problems, spelling, English, history, geography and science.
- **California Achievement Test:** This battery is available in two forms: i) Intermediate level. ii) Advance level. Previously, this battery was known as Progressive achievement test. The intermediate form

takes 151 minutes and the advanced form takes 148 minutes. The tests include vocabulary of mathematics, science, social studies and general comprehension reasoning, fundamental operations in arithmetic and the English language. Grade, age and percentile norms are provided. A hand writing scale is added and can be used if needed.

- **Iowa Every- Pupil Tests of Basic Skills:** This battery is meant for grades five through nine. It takes 268 minutes. The basic skills included in the tests are: reading comprehension, vocabulary, map reading, use of references, use of index, use of dictionary, reading of graphs, charts and tables, punctuation, capitalization, arithmetic problems, fundamental arithmetic operations, spellings and vocabulary. Pupil norms for each test consist of age norm, grade norms, and percentile grade norms.

Iowa Test of Educational Development: These tests are prepared in Iowa University and was issued in 1942 for national use and are intended for grades nine through thirteen. These give a score for general educational and intellectual development. The whole battery takes 459 minutes to apply. The four types of tests included are: (i) Background Tests, (ii) Reading Tests, (iii) Vocabulary Tests and (iv) Sources of information Tests.

STOP TO CONSIDER-3

- Standardization implies uniformity of procedure in administering and scoring the test.

CHECK YOUR PROGRESS-3

- Mention one important standardized educational achievement test.

5.4 Construction of Achievement Test

- 1. Planning the Test:** Planning is the first and foremost step in constructing a standardized test. This step includes:
 - a) Determining Objectives:** First of all while planning, it needs to identify a test objective that's going to be constructed.
 - b) Analysis of Curriculum and Source Material:** The content analysis is another important step in test construction. After determining the objectives, books and materials related to the topics should be analyzed.

2. **Preparing Final Draft:** It is considered as blue print of the entire test. This step follows:
 - a) **Item Construction:** The items construction may be of true or false, alternative response, multiple choices, completion, etc. During item construction some general principles should be considered, like- the items belonging to the same category should not be repeated. In the first draft the items should be more than that in the final draft and ambiguous questions should not be included in the test and it should be comprehensive enough.
 - b) **Arranging the Items:** The arrangement of items should be of equal difficulty level, test items may be in increasing difficulty order, including one item of various types and grouping items subject matter wise and arranging according to logic.
3. **Trying Out the Test:** After preparing the final draft, the test is given for trying out on a norm group. In this step, it follows:
 - a) **Administration:** After preparing the final draft the test is administered on a sample where good, poor and average students are included. There should be proper arrangement of administering the test with clear instruction.
 - b) **Scoring:** In this trying out process the scoring procedure should be of simple one. One point may be given for every correct response. For that answer key may be prepared beforehand.
 - c) **Item Analysis:** This technique increases effectiveness of a test. It can be analyzed quantitatively in reference to their form and subject. Moreover, the length of a test can be reduced by item analysis.
 - d) **Item Difficulty:** In this step item difficulty is found out on the basis of pass percentage of students. This is helpful in finding out the difficulty level of each and every item.
 - e) **Internal Consistency:** The internal consistency of items is done on the basis of total scores of the test. Internal consistency of items needs to be finding out and it can be done by coefficient of correlation of the test item.
 - f) **Finding out Discrimination Value:** The percentage of students doing the each item correctly and incorrectly has to be calculated. This discrimination value can be found by following methods.

1. Preparing the Final Draft:

After scoring the item analysis the final test may be constructed. The items with good discrimination value may be taken and others may be eliminated. Proper instructions are to be given and required time should be decided. After that second administration of the test should be done. Then average percentage may be calculated. Consequently standard deviation, range, quartile deviation may also be calculated.

2. Standardization and Interpretation:

For standardization and interpretation, norms may be calculated like age norm, sex norm, grade norm, rural and urban norm, etc. Average performance of various groups will help in comparison of scores. After this raw scores may be converted into standard score like Z and T score. The procedure adopted must be suitable for the test so that everyone can be interpreted the test easily.

If a test follows the above mentioned steps and procedure then it will become a standardized test and it can be applicable at any time and place. *(For more details referred Unit-II)*

STOP TO CONSIDER

- **Item Difficulty:** In this trying out step item difficulty can be found out on the basis of pass percentage of students. This is helpful to finding out the difficulty level of each and every item.
- **Internal Consistency:** The internal consistency of items is done on the basis of total scores of the test. Internal consistency of items needs to be finding out and it can be done by coefficient of correlation of the test item.
- **Discrimination value:** The percentage of students doing the each item correctly and incorrectly has to be calculated. This discrimination value can be found by following methods.

CHECK YOUR PROGRESS - 4

- What is discriminating index?
What is internal consistency

5.5 Diagnostic Test and Performance Test

The term diagnosis has been derived from the medical profession. It means identification of disease by means of patient's symptoms which patients express in doctor's query. Then doctor prescribes medicine as remedial treatment after careful analysis of the information provided by patients. Similarly, in the field of education teachers apply diagnosis technique to find out students' strength in specific taught areas. Because difficulties in learning occur frequently at all stages and among pupils of both high and low level mental ability. The teacher, then, analyzes the causes of weakness of students and provides remedial measures as per necessary. So a diagnostic test can be stated "As a test which is used to find out student's strength and weakness in certain course of study". And it is designed to analyze individual performance and provide information on the causes of difficulty. But the teachers need to be cautious to use the diagnostic data with great care for designing remedial programs because the tools and techniques used in mental measurements are not exact, objective and precise like those used in sciences.

Salient Features of Diagnostic Test:

1. It starts where the formative test ends.
2. It's a means by which a student's profile is examined and compared against certain norms or criteria.
3. It focuses on student's educational weakness or learning deficiency and identifies the gaps in pupils.
4. It is more intensive and acts as a tool for analysis of learning difficulties of student's.
5. It is more often limited to lower ability/low proficiency.
6. It is corrective in nature.
7. It helps teachers to identify the trouble spots and discover those areas of students' weakness that are unresolved by formative test.

Performance Test:

Measurement of intelligence is a prime concern of psychological tests. There are so many tests which are generally used in measuring intelligence. Initially, some tests were used to measure intelligence which was known as verbal intelligence. This test can be applied only in those individuals who are literate and who have knowledge of language. But these tests cannot be applied in case of illiterate individuals. That is why it was strongly felt to develop another type of intelligence test which came to be known as performance test.

Performance tests are typical form of non verbal test. These have also been constructed to overcome some practical difficulties and

inconvenience of testing intelligence with the help of verbal test. As the name implies this type of test is one in which the subject has to perform and manipulate different concrete materials without much use of language ability. These tests are usually prepared for groups such as persons being tested in a non native language, illiterate people, young children, mentally retarded, deaf and dumb, etc.

Performance test consist of variety of items. Some of most common types of items are form board in which examinee must place cut out objects in corresponding hole on the board. Maze learning, block design test, pass along test are some of the renowned performance test.

Performance test have some significant characteristics. They are:—

- a) Performance tests are generally used for those groups of students who have language barriers, physical deformities, difficulty in heaving and speaking. In this test reading or writing is not needed, so it can also be applied in case of illiterate people. Moreover this type of test can also be applied to individuals speaking foreign languages, youngers and severely mentally retarded people to measure intelligence. The main character of this type of test is non-using of language.
- b) The contexts of this type of tests are in material objects. Eg= in Koh's Block design test, some blocks are needed to design according to design given on the card. Subject has to complete it with the help of blocks given to them. The intelligence is measured on the basis of performance of the subject. The time taken by the subjects to complete the test indicates their level of intelligence.
- c) In these tests there is no use of written language. Instructions are given to the subject in oral form. Sometimes some special signs are used to impart instructions to the subjects. On the basis of these signs the subject has to complete the task and on the basis of performance shown by the subject intelligence is measured.
- d) In performance tests some apparatus are very essential. The subject has to complete the task with help of any apparatus by following the instructions given to him. That is why apparatus is the primary concern of performance test.
- e) The subject's responses in performance tests are assessed in terms of what he does rather than what he says. It means responses are assessed on the basis of activities completed by the subjects. No verbal responses are considered as a sign of performance and examiner observes the activities done by the subject. Similarly behaviour is observed during the time of activity and notes it down if possible.

STOP TO CONSIDER

- Performance test is one in which language is used only in the instruction or not at all when directions are given.
- Diagnostic test is a test designed to real specific weakness or failure to learn in some subject of study such as reading or arithmetic.

CHECK YOUR PROGRESS - 5

1. Name one performance test.
2. Mention two important characteristics of performance test.
3. Mention two important features of diagnostic test.

Teacher Made Test:

Teacher made tests are normally prepared and administered for testing classroom achievement of students, to know how far the specific objectives have been fulfilled, diagnose the learning difficulties, arrange specific remedial measures, evaluating the method of teaching adopted by the teacher, award grades and other curricular programmes of the school. Teacher made tests are based on content and objective specific to teachers own class or school. It may deal either with a specific limited topic or skill or with larger segments of knowledge and skill. Generally, the teacher-made tests are less reliable in comparison to the standardized test. It has a limited use, to a particular class or a single school. However, it is one of the most valuable instruments in the hands of the teacher to know students and solve his purpose.

Teacher made tests are basically used for formative evaluation. A Teacher made test may be a full-fledged achievement test covering the whole course of a particular subject or a unit test covering one unit of study. The teacher made tests are evaluated by the same teacher, who makes the test paper. While constructing a test paper, it needs to follow certain principles. These principles are establishing reliability and validity of course contents, fully following principles in constructing a test paper is called standardized test and constructing a test paper without fully following principles and prepared by a teacher is called Teacher made tests.

Features of Teacher-Made Tests:

1. The items of the tests are arranged in increasing order of difficulty.
2. These tests are prepared by the teachers and can be used for prognosis and diagnosis purposes.

3. The test covers the whole content area and includes a large number of items.
4. The preparation of the items conforms to the blueprint.
5. Test construction is not a cooperative endeavour.
6. A teacher-made test does not cover all the steps of a standardized test.
7. Teacher-made tests may also be employed as a tool for formative evaluation.
8. Preparation and administration of these tests are economical.
9. The test is developed by the teacher to ascertain the students' achievement and proficiency in a given subject.
10. Teacher-made tests are least likely to be used for research purposes.
11. They don't have norms, whereas, providing norms is quite essential for standardized tests.

Written and Oral Test Type Test:

The written test is a comprehensive exam that assesses the examinees' subject knowledge based on criterion-referenced test. Examinees are measured against the skill and knowledge represented by each test item. In this test there are many advantages than the oral test. Since the examinees express test response in written form, so skill in writing tends to become a part of the score given to response. It is seldom appropriate for use with young children or with the physically handicapped who lack in writing skills. Moreover, it is of little use in measuring the oral communication skills. The written response lacks most of the unique advantages of the traditional oral examination, except for a few instructional objectives which are amenable to written tests. The written response form of the oral examination is used by the classroom teacher more often than the oral response form.

The oral test or examination, means a spoken or verbal form of test. It is a face to face taking test. In this test examiner presents the questions in spoken form and the examinee or student has to express their subject knowledge in same manner. This test is especially popular in language course but can be used to assess understanding level in almost any course. The basic advantage of oral test is that students get immediate feedback which allows them to learn as they are tested. However, the amount of time required and record keeping is the main problem in this test.

Formats of Oral Test:

1. Interview Style: The examinee is asked on general topics.
2. Clinical Style: Questions are asked specifically and regarding diagnosis and treatment purposes to a particular patient.

3. Cognitive Style: In this type it requires problem solving around specific cases.
4. Role-Playing Style: Students needs to assume roles and to perform within this role.

5.6 Essay and Objective Type Tests

Essay Type Test:

Test, measurement and evaluation are the important parts of teaching-learning process. Test gives us information on achievement of instructional objectives by students'. To measure and ensure the level of achievement obtained by students' in this objectives teacher applies different types of test, viz. essay-type test, short-answer type test and objective-type test.

The word essay originated from the France word 'essayer' which means 'to try' or 'to attempt'. The essay test refers to any written test that requires the examinee to write a sentence, a paragraph or longer passage. William Weirsama defined essay test as "a test that requires the student to structure a rather long written response up to several paragraphs." This test gives full freedom to the students to express own thought which they present through selection and organizing. So, it provides a better indication of pupil's real achievement. It is, thus, used to assess skills in organizing and summarizing the information and explaining events or places. In other words, the answer of the pupil reveals the structure, dynamics and functioning of pupil's mental life.

Characteristics of Essay Test:

1. The length of the required responses varies with reference to marks and time.
2. It demands a subjective judgment; so it varies from person to person.
3. Most familiar and widely used.

Forms of Essay Test:

- (a) **Restricted Response:** In this test usually limits are given both in the content and the response by restricting scope of the topic to be discussed. The questions are followed with certain word limits. So, it is useful for measuring learning outcomes requiring interpretation and application of data in a specific area. But this test has limitations like-students opportunity to express their desired responses is limited. So their originality of expression, thinking, organization and integration are not come out fully and freely.
- (b) **Extended Response:** This test allows students freedom to select information that they think is pertinent, to organize the answer in

accordance with their best judgment, and to integrate and evaluate ideas as they think suitable. It does not set limits on the length or exact content to be discussed. For freedom student's creativity, originality, cognitive ability, etc are facilitating students to express and similarly it can be measured from this test.

Advantages of Essay Test:

1. It is relatively easier to prepare and administer than multiple-choice test.
2. It helps to assess an examinee's ability to organize and present own ideas in a logical and coherent manner.
3. It can be successfully employed to all subjects.
4. Logical thinking and critical reasoning,, systematic presentation, etc can be best deployed by this type of test.
5. It helps students to improve their writing skills such as writing speed.
6. It encourages creativity by allowing their own unique way.
7. The students can show their initiative, the originality of their thought and the fertility of own imagination as they are permitted freedom of response.
8. The responses of the students need not be completely right or wrong. All degrees of comprehensiveness and accuracy are possible in this type of test.
9. Essay Test largely eliminates guessing and provides less scope for adopting unfair means during test.
10. It is more economical than objective type test.

Disadvantages of Essay Test:

1. Essay tests questions don't cover the whole course content as comprehensively covered in objective type tests.
2. This test possesses relatively low validity and reliability for covering course content, chances of biasness and subjectivity of scoring, good verbal ability even in the absence of relevant points, mood of examiner, first impression, improper comparison of answers of different students', and ambiguous wording of the questions for which student's may interpret differently by various students.
3. It requires excessive time on the part of students to write and similarly with examiners in assessing. So, it is a time consuming and laborious task.
4. It can be assessed only by a teacher or competent professionals.

Though the essay test has lots of demerits but it can be minimized by following certain precaution in framing questions and during assessing.

Short Answer Type Test:

Short-answer type tests are open-ended tests that requires students to create an answer. This test is commonly used in examinations to assess the basic knowledge and understanding (lower cognitive levels) of a topic. The form of answers may such as complete the sentence, fill up the missing words, short descriptive or qualitative answers, diagrams with explanations, etc. the answers are usually short, varying from one word to a few lines. This test is used to improve the test reliability, validity and sampling capacity. This test has the following distinctive features:

- (i) It takes less time to read, answer and assess.
- (ii) It permits larger sampling of content.
- (iii) It tends towards greater objectivity in scoring.
- (iv) It is more reliable and valid than essay type test.
- (v) It is precise and specific as to the scope and length of answers.

Advantages of Short Answer Type Test:

1. Objectivity of scoring can be better ensured in this test compared to essay test.
2. Greater coverage of course content.
3. It is more reliable and valid than essay test.
4. There is less chance of guessing the answer by the students.
5. It is easier to prepare and administer.

Disadvantages of Short Answer Type Test:

1. In short answer questions test the assessor need to be very careful on the type of answers expected while setting questions, because it is open-ended questions, so students are free to answer any way they choose which may lead to difficulties in grading if the question is not worded carefully.
2. It is typically used for assessing knowledge level only, so students may often memorize the answers with rote learning.
3. It is difficult to address students' reasoning tasks, such as decision making and critical thinking because they don't have opportunity to elaborate.
4. Accuracy of assessment may be influenced by handwriting or spelling.
5. There can be time management issues in this test.

Objective Type Test:

An objective type test item is a tool which is free from all subjective elements. It is characterized by its total objectivity and unambiguity in scoring. It has one and only one correct response. Full marks is given for the correct response, which is unique, and no mark is given for an incorrect response. Another characteristic of an objective type test item is that it is completely structured. Complete structuring of task limits student to the type of response called for. It prevents bluffing and avoids influence of writing skills. The scoring of an objective type test item is not only objective but also quick, easy and consistent, which is its greatest advantage.

Types of Objective Test:

- **True or False:** This objective type test is presented in a simple declarative statement and the pupils respond indicating that the statement is true or false. It is applicable to all learning areas.
- **Matching Type:** The matching test format consists of two columns, and learner needs to correlate correctly one or more entries from one list with one or more entries from the other. It evaluates the student's ability to recall information under time constraints.
- **Multiple Choices:** In this test students have to select one correct answer from several choices offered to them.
- **Enumeration:** In this test two or more responses to an item will be there.
- **Labeling:** It's an objective type test in which the names of parts of diagrams, map, drawing or picture are to be indicated.
- **Identification:** This is an objective type of test in a form of completion test which is defined, describe, explained or indicated by a picture, diagram or a concrete object and the term referred to is supplied by the students. Example: Direction: Identify the following-Question: The best and the most accurate measure of variability.
Answer: Standard Deviation
- **Completion Type:** This test includes series of sentences where certain important word has been omitted and the testee is to write a number or a word or a phrase so as to complete the statement. And each blank counts one point.
- **Simple Recall:** In this objective type of test that sometimes require the students' to supply an answer to direct question and sometimes require him/her to complete a statement where a word or phrase has been omitted.

STOP TO CONSIDER

- Teacher made test are divided into two types essay type test and objective type test.
- There are two forms of Essay type test —Restricted Response and Extended Response.
- Objective test are of various forms which includes True or False, Matching type, multiple choices, Enumeration, Labeling, Identification, Completion, Simple Recall Etc.
- Objective type tests are further sub-divided into oral type and written type test.

CHECK YOUR PROGRESS- 6

- Mention two disadvantages of essay type test.
- What are the different forms of objective type test?

5.7 Summing Up:

- ❖ Educational Achievement is an instrument designed to measure relative achievement of students. It is an indispensable instrument of teaching-learning process. In the words of Gronland, an achievement test is 'a systematic procedure for determining the amount a student has learned through instruction.'
- ❖ There are various dimensions that can be used to classify achievement test. Achievement tests are broadly classified into two categories- Standardized achievement test and Teacher made test. Teacher made tests are further divided into oral test and written test. Written type test is again divided into essay type, short answer type and objective type test.
- ❖ Standardized achievement tests are commercially produced test on the basis of certain procedure to meet the demands of objectivity and accuracy. These are constructed by test publishers and are designed to use in a wide variety of situations.
- ❖ Teacher made tests are designed to operate within the restricted situation of a given classroom. Teacher, who makes the test, uses his discretion in matters of the scope of the test area and choice of task types and items.

- ❖ Essay type test demand lengthy types of questions which are influenced by halo effect, intra examiner variability, inter examiner variability in scoring. In other side, objective type test is objective based because scoring is objective which demand answer in a sentence or a word or a mark.
- ❖ A performance test is one in which language is used only in the instructions or not at all when directions are given. The chief characteristics of the performance test are that it does not require the use of language or number to take the test.
- ❖ In diagnostic test, score or mark is not assigned for correct answer by wrong answer, considering in view of the sequence of the content to identify the cause for the wrong answer.

KEY TERMS:

Achievement test: An Achievement test is designed to measure the extent to which an individual has 'achieved' some information or mastered certain skill, usually as a result of specific instruction provided in a classroom or training programme.

Teacher made test: Teacher made tests are designed specially according to the content and objectives of particular unit of study. They are more suitable to the assessment of objectives of a particular course. These are not governed rigidly by any process. Teacher who makes the test uses his discretion in matters of the scope of test area and choice of task types and items.

Diagnostic test: In diagnostic test, the main interest is the performance on individual items or small groups of highly similar items. In diagnostic test, score or mark is not assigned for the correct answer but wrong answer provides the basis for the cause of his failure.

Performance test: The main aspect of performance test is that it does not require the use of language or number for the completion of the task. In performance test, the subject has to manipulate or make something eg ; picture completion, making design and cube construction etc.

CHECK YOUR PROGRESS ANSWERS:

Check your progress answer 1:

1. Freeman defines educational achievement as a test, "designed to measure knowledge, understanding, and skills in a specific subject or group of subjects."

Check your progress answer 2:

1. The two important objectives of educational achievement test are-
 - (a) To measure knowledge and achievement of students.
 - (b) To help in assessing the all round mental ability of the child.

Check your progress answer 3:

1. One of the important standardized achievement tests is California Achievement test.

Check your progress answer 4:

1. Discriminating Index: It is the percentage of students doing the each item correctly and incorrectly.
2. Internal Consistency: The internal consistency of items is done on the basis of total scores of the test. Internal consistency of items needs to be determined and it can be done by finding out the coefficient of correlation of the test item.

Check your progress answer 5:

1. One important performance test is Pass Along test.
2. Two important characteristics of performance test are-
 - a) Non-use of language.
 - b) In performance tests some apparatus are very essential. The subject has to complete the task with the help of any apparatus by following the instructions. That is why apparatus is primary concern of performance test.
3. Two important features of diagnostic tests are-
 - a. It's a means by which a student's profile is examined and compared against certain norms or criteria.
 - b. It focuses on student's educational weakness or learning deficiency and identifies the gaps in pupils.

Check your progress answer 6:

1. Two disadvantages of essay type test:
 - a) Essay tests questions do not cover the whole course content as comprehensively covered in objective type tests.
 - b) It requires excessive time on the part of students to write and similarly with examiners in assessing. So, it is a time consuming and laborious task.

2. Different forms of objective type test are-
 - a) True and false
 - b) Multiple choice
 - c) Matching types
 - d) Completion type
 - e) Simple recall.

Questions and Exercise:

Short –Answer Questions:

1. What do you mean by educational achievement test?
2. Mention different objectives of achievement test.
3. Write short Notes on:
 - a) Performance test.
 - b) Diagnostic test.
 - c) Essay type test.
 - d) Objective type test.
4. What is teacher made test?
5. Write few lines about the oral and written type test.

Long Answer Questions:

1. Describe about the different standardized achievement test.
2. Differentiate between performance test and diagnostic test.
3. Describe about different types of objective type test with examples.
4. Differentiate between teacher made test and objective type test.

5.8 References and Suggested Readings

1. Patel, N. Rambhai, "*Educational Evaluation-Theory & Practice*", Himalaya Publishing House, Dr. Bhalerao Marg, Mumbai, 400004.
2. Panda, P. K., "*Measurement, Evaluation and Assessment in Education*", Agarawal Publications, Jyoti Block, Sanjay Place, Agra-2.
3. Sharma, R.A., "*Essentials of Measurement in Education and Psychology*", Begum Bridge Road, Meerut-250001
4. Anastasi, A., and Urbina, S. (1997) *Psychological Testing*, 7th edition. New Delhi: Prentice Hall of Private Limited.
5. Rath, R.K. *Fundamentals of Educational Statistics and Measurement*, 3rd edition. Orissa: Taratarini Publishers.

6. Singh, Raj. (1994) *Techniques of Measurement and Evaluation*, 1st edition. New Delhi: Common Wealth Publishers.
7. Thorndike, R.L. (1970) *Measurement and Evaluation in Psychology and Education*. New Delhi: Wiley Eastern Limited.
8. Asthana, Bipin (2017/18) *Measurement, Evaluation and Assessment in Education*. Agrawal Publications. Agra, New Delhi.
9. "Academic Achievement of Adolescents in Relation to Study Habits," Reeta Arora, The International Journal of Indian Psychology, Vol-3, Issue-2, January-March. 2016. www.ijip.in

—x—

[The page contains extremely faint, illegible text, likely bleed-through from the reverse side of the document. The text is too light to transcribe accurately.]