PG ED S2-02 (DSC)

TWO YEAR MASTER'S DEGREE COURSE

(CBCS in Accordance with NEP 2020)

SEMESTER - II

MEASUREMENT AND EVALUATION IN EDUCATION

EDUCATION

Indira Miri School of Education

KRISHNA KANTA HANDIQUI STATE OPEN UNIVERSITY

Subject Experts

1. Prof. Swarnalata Das, Dept. of Education, Gauhati University
2. Prof. Gayatree Goswamee, Dept. of Education, Gauhati University
3. Dr. Sadhana Goswami, Dept. of Education, Cotton University

Course Co-ordinator: Dr. Archana Adhikary & Dr. Indrani Kalita, KKHSOU

SLM Preparation Team

UNITS	CONTRIBUTOR
1, 2, 3, 4, 5, 6, 7, 8 & 9	**Dr. Mukul Saikia,** Darrang College, Tezpur
10, 11 &12	**Dr. Maramee Goswami,** Gauhati College
13	**Dr. Pranab Saikia,** KKHSOU
14	**Dr. Sradhanjali Pradhan,** Tezpur University
15	**Dr. Utpal Kalita,** R.G. Baruah College

Editorial Team

Content Editors	**: Prof. Polee Saikia,** Gauhati University
	: Dr. Mallika Kalita, KVBSASU, Nalbari
	: Dr. Nilima Deka, L.C.B College, Gauhati
Language Editors	**: Dr. Sukanya Kashyap,** USTM
Format Editor	**: Dr. Archana Adhikary**

Structure, Format & Graphics : Dr. Archana Adhikary, KKHSOU

Printed: December, 2023

This Self Learning Material (SLM) of the Krishna Kanta Handiqui State University is made available under a Creative Commons Attribution-Non Commercial-ShareAlike4.0 License (International): http.//creativecommons.org/licenses/by-nc-sa/4.0.

Printed and published by Registrar on behalf of the Krishna Kanta Handiqui State Open University.

Head Office : Patgaon, Rani Gate, Guwahati-781 017

City Campus: Resham Nagar, Khanapara, Guwahati-781 022; Website: www.kkhsou.in

CONTENTS

		Page No:
Unit 1 :	**Educational Measurement** Meaning of measurement, Types of measurement, Meaning of educational measurement, Nature of educational measurement, Functions of educational measurement, Scales of measurement.	**7-23**
Unit 2:	**Psychological Test** A brief history of psychological testing, Meaning and definition of psychological test, Nature of psychological test, test, examination and evaluation, Types of psychological test, Functions of psychological test, Uses of psychological test, Meaning of a good test, Practical criteria of good test, and Technical criteria of a good test.	**24-59**
Unit 3:	**Evaluation in Education** Meaning and definition of evaluation, Measurement, assessment and evaluation, Nature of educational evaluation, Relation between measurement and evaluation, Principles of evaluation, Steps in evaluation, Importance of evaluation in education, Tools and techniques of evaluation, Types of evaluation.	**60-83**
Unit 4:	**Test Construction and Standardization** Meaning of psychological test, Meaning and characteristics of a standardized test, Principles of test construction, Methods of test construction and standardization,	**84-98**
Unit 5:	**Test Items** Meaning of item analysis, Objectives or purpose of item analysis, Analysis and selection of test items, Determination of difficulty level, Determination of discriminating power, Distracter analysis, preparing different types of test items, Guidelines for construction of different types of test items.	**99-119**
Unit 6:	**Validity** Meaning and definitions of validity, Factors affecting validity of a test, Types of validity of a test, Methods of estimating validity of a test.	**120-134**
Unit 7:	**Reliability** Meaning and definition of reliability, Factors affecting reliability of a test, Methods of estimating reliability of a test, Relation between reliability and validity,	**135-150**
Unit 8:	**Educational Achievement and Its Measurement** Meaning of educational achievement, Measurement of educational achievement, Objectives of achievement tests, Types of achievement tests, Standardized achievement test, Teacher-made achievement test, Essay type, Objective type, and short answer type tests, Difference between standardized tests and teacher-made tests.	**151-173**

UNIT 9: Intelligence Test And Its Measurement **174-195**

Meaning and definitions of intelligence, Brief overview of the theories of intelligence, Measurement of intelligence, Types of intelligence tests, Uses of intelligence tests, Construction of an intelligence test.

UNIT 10: Measurement of Aptitude **196-223**

Meaning of aptitude, Definitions of aptitude, Nature / characteristics of aptitudes, Measurement of aptitude, Meaning of aptitude test, Types of aptitude tests, Multiple aptitude tests, General aptitude test battery (**GATB**), Differential aptitude tests (DAT), Test of primary mental ability (PMA), Measures of specific aptitudes, Mechanical aptitude test, Tests of clerical aptitude, Musical and artistic abilities, Measuring aptitude for art, Tests of scholastic and professional aptitudes, Importance of aptitude test.

UNIT 11: Measurement of Attitude **224-248**

Meaning of attitude, Definitions of attitude, Nature or characteristics of attitude, Importance of attitude testing, Measurement of attitude, Types of attitude test, Attitude scales, Thurstone's scale , Likert type scale, Semantic differential scale, Guttman scales, Master scale method of Remmers and Silance, rating scales, Rank order scales, Percentage of population scales, indirect scales, Projective techniques.

UNIT 12: Measurement of Interest **249-269**

Meaning of interest, Definitions of interest, Types of interest, Nature of interest, Measurement of interests, Techniques of measuring interest, Formal assessment techniques: Interest inventories, Strong vocational interest blank (**SVIB**), Kuder preference record (KPR), Thurston's vocational interest schedule, Informal assessment techniques, Importance of interest testing.

UNIT 13: Measurement of Personality **270-293**

Methods of assessing personality, Subjective methods, Objective methods, Projective methods.

UNIT 14: Measurement of Creativity **294-312**

Meaning of creativity, Criteria of creative personality, Measurement of creative thinking, Measurement of creative thinking for different levels of students.

UNIT 15: Emerging Trends In Evaluation **313-339**

Concept of question bank, Purposes of question bank, Construction of question bank, Uses of question bank, Concept of grading system, Methods of grading system, Advantages and disadvantages of grading system, Concept of semester system, Components of the semester system, Advantages and disadvantages of semester system, Use of computer in evaluation.

COURSE INTRODUCTION

This is the second course in Education for the second semester of two year Master Degree Level. This course primarily will deal with the basic concepts of measurement and evaluation in education. As we know that measurement and evaluation are integral parts of total education system, this course will introduce you with the major components of measurement and evaluation in the fields of education. The first section of the course will help you to know the different measurement and evaluation process and procedures. The later section of this course will introduce you with various test that can be done and administered on individual man or group of people. After going through this course, you will have a clear picture of measurement and evaluation and how they are applicable in the educational process.

The first Unit-**"Educational Measurement"** deals with the meaning and nature of educational measurement and different scales of measurement. The second unit- **"Psychological Test"** discusses the meaning, nature, function and different uses of psychological test in the field of education. The third unit- **"Evaluation in Education"**, being the most important in terms of its significance deals with the basic concept, nature, tools, principles and types of evaluation. The fourth unit **"Test Construction and standardization"** deals with the basic procedure of test construction through different stages and standardization. The fifth unit **"Test Items"** discusses about analysis, selection and preparation of different test items, which is an integral part of psychological test. The sixth unit-**"Validity"** and the seventh unit **"Reliability"** discusses the meaning, nature and testing of validity and reliability respectively. The eight unit **"Educational Achievement and its Measurement"** explains about different types of educational achievement test and its measuring procedure. The **ninth Unit-"*Intelligence test and its measurement"*** explains how to measure and construct intelligence test. It also discusses different types of intelligence test. The **tenth unit- "*Measurement of Aptitude*"** is the discussion of basic concept and measurement of individual's aptitude**.** The **eleventh unit**- "***Measurement of Attitude"*** deals with the meaning, nature, classification and measurement of individual's attitude. The **twelve unit** "***Measurement of Interest"*** explains the measurement of interest along with the basic concept and nature of interest. The **thirteen unit "*Measurement of Personality"*** being a very important topic of discussion deals with the different measurement techniques of human personality**.** The **fourteen unit "*Measurement of Creativity*"** explains about meaning, nature of creativity, different criteria of creative personality as well as measurement of creative thinking for different levels of students. The **fifteen** and **last unit** of this course "***Emerging Trend in Evaluation*"** is the explanation of different new trends in the field of evaluation like- Question bank, Grading system, Semester system in education.

While going through this learning material, you will come across some boxes in the middle of the text. These boxes will explain you more about some important points and concept within the text. Apart from this, there will be some broad and short questions which have been included under **"ACTIVITY"** and "**CHECK YOUR PROGRESS"** in every unit. Activities will increase your thinking and reasoning capacity as questions put in **ACTIVITY** are not directly derived from text. Hints on the answer to the short questions have been put under section **"ANSWER TO CHECK YOUR PROGRESS"** .Besides, a few text-related questions have been out in **"POSSIBLE QUESTIONS/ MODEL QUESTIONS".** These questions will help to draw your attention on the probable topics for the examinations so that you can prepare for the examination with confidence and determination.

UNIT 1: EDUCATIONAL MEASUREMENT

UNIT STRUCTURE

1.1 Learning Objectives

1.2 Introduction

1.3 Meaning of Measurement

1.3.1 Types of Measurement

1.4 Meaning of Educational Measurement

1.5 Nature of Educational Measurement

1.6 Functions of Educational Measurement

1.7 Scales of Measurement

1.8 Let Us Sum Up

1.9 Further Reading

1.10 Answer to Check Your Progress

1.11 Model Questions

1.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- understand the meaning and types of Measurement
- define and explain the nature of Educational Measurement
- explain the functions of Educational Measurement
- explain the different levels or scales of Measurement

1.2 INTRODUCTION

Measurement is an indispensable part of human life. Almost all aspects of our daily life are touched by measurement of different forms. Ross & Stanley, in their book Measurement in today's schools wrote, "From birth to death almost every aspect of our daily lives is touched by measurement in its numerous forms. At birth the record of that important event is carefully made according to the nurse's watch. During the next few days measurements of the baby's weight and temperature are part of the daily routine of the hospital. Ever afterward, whether in school or outside,

watches, clocks, balances, thermometers, money systems, and other forms of measurement play prominent roles in the life of every human being." Thus, measurement, in its various forms, controls man's life to a great extent. Day-to-day activities like buying and selling, cooking, stitching, constructing, diagnosing diseases and using medicines, and numerous other such acts require measurement in its different forms. Hence, the concept of measurement is neither new nor difficult for us. In this unit we are going to discuss the entire concept of measurement in education.

1.3 MEANING OF MEASUREMENT

In simple word, measurement is the act or process of ascertaining the size, dimensions, amount, degree, or quantity of something. It is a process by which some characteristic of an individual or some object or event is presented in the form of some number or symbol according to specified rule(s). Measurement ascertains dimensions such as: big or small; much or less; long or short; high or low; of an object, event or other phenomena. Different units such as kilogram, meter, foot, inch, etc. and various tools such as weighing scale, thermometers, measuring tapes etc. are used for making different types of measurement. For example, in order to measure the physical dimensions such as the length, breadth and height of a classroom, tools like a measuring tape or rule is required and the measurement is expressed in units like meter, foot, or inches.

Psychometricians have put forward many definitions of measurement in order to explain the meaning of measurement. Some of the popular definitions have been presented below:-

- According to **S.S. Stevens**, "Measurement means assignment of numbers to objects or events according to rules".
- According to **James M. Bradfield**, "Measurement is the process of assigning symbols to dimensions in order to characterize the status of a phenomenon as precisely as possible".
- **Guilford** says, "Measurement means the description of data in terms of numbers and this, in turn, means taking advantage of the many

benefits that operations with numbers and mathematical thinking provide".

Thus, measurement means quantification of data. It involves the process of presenting some object, event, or their characteristics in symbolic or numeric form. Objects, events, or their characteristics are made more comprehensible by the way of comparing them with the units used for making measurement.

1.3.1 Types of Measurement

Measurement touches the life of everyone in different fields of day to day activities. However, the nature and purpose of all measurement is not the same. For example, the measurement of the dimensions of a room differs vastly in its nature as well in its purpose from the measurement of intelligence or achievement of an individual. Moreover, the former requires the use of physical sciences whereas the latter requires psychological tools. Hence, measurement may be classified into different types on the basis of their purpose, nature, and the tools or methods used. Following are the main types of measurement: -

Physical Measurement: Measurement that exists in the physical or material world is known as physical measurement. As the name suggests, this type of measurement is used to measure various physical or material things of the world. The dimensions such as weight, height, length, breadth etc. are measured using units such as gram, meter, foot, etc. which are of definite value and have universal standard. That is why physical measurement is simple and objective. Some important characteristics of physical measurement are:-

- Physical measurement starts from a true zero point. For example, when the length of something is measured to be ten feet, it is implied that the length of the object is ten feet from zero. Zero, in physical measurement, indicates the absolute

absence of the dimension being measured.

- Physical measurement is complete in itself; it is not dependent upon any other thing.
- Physical measurement is direct, definite, and objective.
- In physical measurement, the object or thing being measured can be measured completely.
- The units and the tools of physical measurement are definite, universal, and of the same standard.
- In physical measurement, the units bearing the same name have the same value and hence they can be subjected to direct comparison.

Psychological Measurement: Measurement that exists in the psychological or mental world and their characteristics is known as psychological measurement. This type of measurement is generally qualitative measurement. Unlike physical measurement, psychological measurement is indirect, because psychological characteristics or traits are abstract and hence cannot be subjected to direct measurement. Such traits can be measured indirectly only when they find expression in the behaviour of man. That is why psychological measurement is a complex process. Moreover, in the absence of definite and standard units of psychological measurement, this measurement is generally subjective and incomplete. Some of the important characteristics of psychological measurement are: -

- Psychological measurement does not have a true zero point. Zero, in this type of measurement, is not absolute but relative. For example, if an individual scores a zero in an intelligence test, that does not signify absolute or complete absence of intelligence in the individual.
- Psychological measurement is not complete in itself; it is relative and is dependent upon various other factors.
- Psychological measurement is indirect, indefinite, and subjective.
- Psychological measurement is incomplete, because the

psychological trait or characteristic being measured can never be measured completely. For example, in order to measure intelligence of an individual, only a sample of his behaviour can be measured and his total behaviour cannot be measured.

- The tools and units in psychological measurement are not of definite or universal standard. Hence psychological measurement does not have universal acceptability.
- Direct comparison between different measurements is not possible in psychological measurement since tools and units are not of definite or universal standard.

1.4 MEANING OF EDUCATIONAL MEASUREMENT

Education refers to the deliberate effort made to bring desirable change in student behaviour. In order to bring desirable change in student behaviour, selected learning experiences are provided to students using a curriculum in the form of teaching. At the end of teaching, the act of ascertaining or determining the changes (that are thought to have taken place as a result of teaching) in student behaviour is known as educational measurement. In order to ascertain the changes in student behaviour, characteristics of students, such as their knowledge, skills, abilities, and interests are assigned with numbers on the basis of an established set of rules. This process of quantifying the influence or effectiveness of learning experiences in terms of modification of student behaviour may be called educational measurement. Defining the meaning of educational measurement, Indian educationist **R. P. Taneja** wrote, "It is the process of quantifying any human attribute that is pertinent in the context of education."

In brief, educational measurement is the measurement of behavioural change in the field of education. For such measurement collection and systematic organization of data relating to the behaviour of an individual in the field of education is required. For this, the activities such as, construction of various educational tests and their standardization; development of other tools of measurement and their use; etc. are involved.

Measurement of educational achievement, intelligence, interest, ability, aptitude, and other traits of students using various tests and other tools falls within the scope of educational measurement. Thus, educational measurement may be understood as a part of psychological measurement. Because, educational behaviour of man is just a part of general behaviour and for the measurement of which methods and tools of psychological measurement is required. It is noteworthy in this context that educational and psychological measurement is not always accurate and definite unlike physical measurement.



CHECK YOUR PROGRESS

Q 1: What do you mean by measurement?

.....

Q 2: What is the importance of measurement in human life?

.....

1.5 NATURE OF EDUCATIONAL MEASUREMENT

Educational measurement is the measurement of behavioural change in the field of education. Measurement in education is important because educators need to ascertain what the learners already know or what they are capable of doing so that further learning can be planned. Moreover, curriculum or subject matter organization is based on the degree of mastery level of academic content for various grades. Hence, measurement is must for educators to verify that learners have acquired the mastery level of each grade and that they are ready to move on to the next grade or course content.

The famous behaviourist psychologist **E. L. Thorndike** opines that, "Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quantity as well as its quality. Education is concerned with

changes in human beings; a change is a difference between two conditions; each of these conditions is known to us only by the products produced by it - things made, words spoken, acts performed, and the like. To measure any of these products means to define its amount in some way so that competent persons will know how large it is, better than they would without measurement. To measure a product well means so to define its amount that competent persons will know how large it is, with some precision, and that this knowledge may be conveniently recorded and used. This is the general credo of those who, in the last decade, have been busy trying to extend and improve measurements of educational products."

In our day to day life we come across a wide variety of measurements of different things in our environment. However, measurement in education is not always definite and accurate unlike the measurement of physical phenomena. The nature of educational measurement is unique which is revealed through the following characteristics of it:

- Educational measurement refers to the quantitative presentation of different aspects of individual's behaviour and their characteristics in the field of education. In other words, educational measurement is the process of assigning numbers or symbols to individuals or their characteristics according to set rules.
- Educational measurement also refers to the process of testing, scaling, or appraising the outcomes of learning. It includes construction and administration of tools such as tests and scales, their standardization and validation, use of statistical techniques in the interpretation of obtained measures or test scores.
- Quantitative values used in educational measurement indicate the presence, absence, or the intensity of the trait being measured.
- Educational measurement does not start from a true zero point. Zero in educational measurement refers only to a relative position.
- The value of units used in educational measurement is neither constant nor definite. That is why educational measurement is not completely free from the subjective influence of person making the

measurement.

- Numerical data obtained by educational measurement does not convey definite meaning in their primary or crude form. For example, the score of 60 obtained by a student in an examination does not make proper sense unless total marks in the examination, minimum marks required to pass the examination, marks of other students in the same examination; etc. are known.
- Since observation or measurement of all of what a student knows or a student can do is not possible, educational measurement is based on samples of behaviours and hence it is incomplete.
- Educational measurement is indirect and is based on inference. Educational achievement or other characteristics of students cannot be measured directly and must be measured indirectly on the basis of their answers or behaviours in response to the questions asked or tasks assigned to them. Moreover, scores or performance of students in Maths or in English is considered to be the representations of their underlying problem-solving or communication skill. Such inference is conclusion based on the sample of behavior assessed during educational measurement.
- Numerical data obtained from different measurements cannot be subjected to direct comparison since measurement in education is relative and not absolute.

Thus, educational measurement is indirect, indefinite, and subjective unlike physical measurement which is direct, definite, and objective. Though constant attempts have been made to minimize the subjectivity and other limitations of educational measurement by experts, it cannot claim the accuracy and precision that characterizes physical measurement.



CHECK YOUR PROGRESS

Q 3: Write two differences between physical measurement and educational measurement.

.....

Q 4: Write any two limitations of educational measurement.

.....

1.6 FUNCTIONS OF EDUCATIONAL MEASUREMENT

Measurement in education performs various functions, including the main function of estimating change in student behaviour. Measurement in education helps in making decision about retaining or promoting students from one grade to another, besides assessing the effectiveness of teaching and learning. It assesses how effectively the educational objectives have been attained. Some important functions of educational measurement may be discussed as under: -

- **Instructional Function:** Measurement is an indispensable part of teaching and instruction. Measurement helps in the selection of objectives of instruction, their adequacy, and in the determination of their usability. Moreover, measurement is a must for assessing the progress of students towards the achievement of objectives. It also helps in determining the success or failure of teaching methods and strategies. Necessary revision and reform of curriculum is also aided by the process of educational measurement.
- **Selection:** Selection of students for admission into educational institutions and particular courses, decision regarding their promotion or retention, and selecting students for various curricular as well as co-curricular activities on the basis of their interests and aptitudes depend on measurement of various sorts.
- **Classification:** Various educational necessities call for classification

of students on the basis of their abilities, interests, or other criteria. Such classifications may be required for selecting students for admission into different courses or career, forming groups or sections of students of homogenous characteristics, making provision for special instruction or guidance; selecting students for advanced studies and such other purposes. Measurement may be used as a basis for making such classification of students.

- **Comparison:** Educational measurement is essential for making comparison between students on the basis of their achievement, ability, aptitude, intelligence, and other traits or characteristics. Such comparison may help to find out individual differences and the causes of such differences among the students. Educational measurement acts as a basis for making comparison between various schemes of instruction, teaching methods, and strategies etc., in order to find out their relative success or failure.
- **Prediction:** Many decisions in education involve prediction as its basis or foundation. Educational measurement is used as a basis for making predictions as to the success or failure of students in particular course, career or other related areas of activities. Current data obtained by measurement indicate the potential success of students in the future.
- **Diagnosis:** Every individual student coming to an educational institution is different from others which make them unique. Different students have strengths and/or weaknesses in different areas. Measurement is important to diagnose such strengths and weaknesses so that proper measures may be taken for their development. Measurement plays crucial role in helping the selection of any course or career on the basis of students' strengths, and for providing them special training or remedial measures to offset their weaknesses.
- **Guidance:** Guidance is essential for the success of students in their educational, vocational, and personal lives. Lack of proper guidance may lead to failure and frustration in the lives of people.

Guidance is vital to live a successful and complete life, keeping aside the problems of life. Measurement of interests, abilities, and other psycho-physical characteristics of students is indispensable for providing effective guidance to students.

- **Administration:** Measurement is very important for making various administrative decisions in any educational enterprise. Measurement forms the basis for many educational decisions, policy formulation, assessing the effectiveness of administrative measures, preparation of progress report, admission, promotion, and other administrative decisions.
- **Research:** Research is very important to find out the effectiveness of curriculum and teaching methods; suitability of educational objectives and the level of their achievement; effectiveness of educational administration; and to find out means for necessary improvement. Measurement in education helps immensely in the conduct of research for finding solutions to educational problems.
- **Development:** Improvement and development of educational practice and related issues is assisted by measurement in education. Educational measurement is very important for obtaining feedback about the success of teaching-learning and teaching methods, for collecting data to improve learning experiences and teaching strategies, and for making effective schemes of education to achieve the objectives of instruction.



CHECK YOUR PROGRESS

Q 5: What are the main functions of educational measurement?

.....

1.7 SCALES OF MEASUREMENT

Measurement is the process of assigning numbers to objects or

events according to specified rules. As Guilford said, "Measurement means the description of data in terms of numbers and this, in turn, means taking advantage of the many benefits that operations with numbers and mathematical thinking provide". However, in the process of measurement, i.e., in assigning numbers, all statistical or mathematical operations are not always applicable to the measurements (numbers) made. It is here, that the concept of scales or levels of measurement becomes relevant and important. Scales of measurement refer to ways in which numbers (measurements) are defined and categorized. Each scale of measurement has certain properties or attributes which determine its applicability in certain mathematical or statistical analysis. Thus, scales of measurement, also known as the Levels of measurement, is a classification that describes the nature of information within the numbers assigned to variables. Stanley Smith Stevens, who was the Director of the Psycho-acoustic laboratory in Harvard University, developed the classification with four levels, or scales, of measurement. These levels or scales include: nominal, ordinal, interval, and ratio.

- **Nominal scale:** When numerical data or numbers in measurement are used as simple identifiers or names, they represent a nominal scale of measurement. Examples of nominal scale includes: registration numbers of vehicles, PINCODE, phone numbers, etc. Similarly, in a research study where gender is a variable, coding Females as 1 and Males as 2 is another example of nominal data. Here, numbers are used only to represent categories of data.

  Nominal scales are known as qualitative scales, and measurements made on qualitative scales give qualitative data. Numbers used in nominal scales are assigned as labels only and they have no specific numerical value or meaning. That is why mathematical computations cannot be performed on nominal data. From a statistical point of view this is the lowest level of measurement.
- **Ordinal scale:** When numbers used in measurement represent an ordered series of relationships or rank order, they represent an ordinal scale of measurement. Students ranked in a class as first,

second, or third represent ordinal data. Thus, ordinal scale gives information about the rank or order of the measures, but it still does not provide information about the relative difference between the successive intervals on the scale. For example, when three boys are ranked as 1st, 2nd, and 3rd on the basis of their intelligence/ achievement, height/ or any other trait; we cannot say that the difference between 1st ranked and the 2nd ranked is proportionate to the difference between the 2nd ranked and the 3rd ranked boy. Such scaling cannot provide information about the difference between the intervals and they represent a rank ordering of the intervals only. However, this scale is widely used in educational and psychological research.

- **Interval scale:** An interval scale represents quantity, and equal difference between successive intervals on the scale, in addition to the information about their order. A temperature scale is an example of an interval scale. On a Fahrenheit or Celsius scale, a degree represents the same amount of heat wherever it occurs on the scale. The difference between a temperature of 30 and 18 degree Celsius is the same as the difference between 42 and 20 degree Celsius. However, zero represents an additional point of measurement in an interval scale and it does not represent the absolute lowest value. For example, zero degree Celsius on a temperature scale does not signify absolute absence of temperature.
- **Ratio scale:** The ratio scale of measurement also represents quantity and equal difference between intervals with an absolute or true zero. Normally, physical measurements, such as height and weight, represent ratio data. Such measurements invariably refer to definite quantity, where units are equal, and zero means absolute absence of the dimension being measured. All types of statistical measures are applicable to ratio scales.

Following characteristics of the scales of measurement in the **table 1.1 shows the difference between them:**

Table 1.1: Difference among the four Scales of measurement

	Differentiates between variables/measures	Ranks/orders measures	Indicates the degree of difference between measures	Possesses absolute or true zero
Nominal	✓	×	×	×
Ordinal	✓	✓	×	×
Interval	✓	✓	✓	×
Ratio	✓	✓	✓	✓

TO LEARN MORE

Stanley Smith Stevens, who was the Director of the Psycho-acoustic laboratory in Harvard University, developed the classification with four levels, or scales, of measurement. These levels or scales include: Nominal, Ordinal, Interval, and Ratio.



1.8 LET US SUM UP

- Measurement is the process of assigning numbers to objects or events according to rules. It is the act of ascertaining the size, dimensions, amount, degree, or quantity of something.
- Measurement may be classified as physical measurement and psychological measurement on the basis of their purpose, nature, and the tools or methods used.
- Educational measurement is the process of determining the changes in student behaviour that are thought to have taken place as a result of teaching. In other words educational measurement is the process of quantifying the influence of learning experiences in terms of modification of student behaviour.
- Educational measurement, which may be understood as a part of psychological measurement, involves collection and systematic organization of data, construction of tests and their standardization, development of other tools of measurement and their use.
- The scope of educational measurement include measurement of educational achievement, intelligence, interest, ability, aptitude, and other traits of students using various tests and other tools.

- Unlike physical measurement, educational measurement has certain limitations. Educational measurement is indirect, indefinite, and subjective unlike physical measurement which is direct, definite, and objective. Though constant attempts have been made to minimize the subjectivity and other limitations of educational measurement by experts, it cannot claim the accuracy and precision that characterizes physical measurement.
- Measurement in education performs various functions, including the main function of estimating change in student behaviour. Measurement in education helps in selection, classification, comparison, prediction, diagnosis, guidance, besides assessing the effectiveness of teaching and learning.
- There are mainly four scales (viz: nominal scale, ordinal scale, interval scale, and ratio scale) of measurement that refer to ways in which numbers are defined and categorized. Each scale of measurement has certain properties or attributes which determine its applicability in certain mathematical or statistical analyses. Thus, scales of measurement, is a classification that describes the nature of information within the numbers assigned to variables.



1.9 FURTHER READING

1) Anastasi A. (1988): Psychological Testing, McMillan, New York
2) Asthana Bipin (2007): Measurement and Evaluation in Psychology and Education, 9th Ed. Vinod Pustak Mandir, Agra
3) Cronbach, L. J. (1990): Essentials of psychological testing,5th Edition, Harper Collins Publishers: New York
4) Freeman, F.S. (1962): Theory and Practice of Psychological Testing, Holt, Rirehant & Winston, New York
5) Pradhan Hemant (1996): Essentials of Educational Measurement, Evaluation and Statistics, Saha Publishers and Distributers, Sambalpur
6) Stanely, Julian C. & Ross, Clay Campbell (1964): Measurement in today's schools, 4th Edition, Prentice Hall

1.10 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Measurement means quantification of data. It is a process by which some characteristic of an individual or some object or event is presented in the form of some number or symbol according to specified rule(s). Measurement ascertains dimensions such as: big or small; much or less; long or short; high or low; of an object, event or other phenomena. According to James M. Bradfield, "Measurement is the process of assigning symbols to dimensions in order to characterize the status of a phenomenon as precisely as possible".

Ans to Q No 2: Measurement is an indispensable part of human life. Almost all aspects of our daily life are touched by measurement. "From birth to death almost every aspect of our daily lives is touched by measurement in its numerous forms". Watches, clocks, balances, thermometers, money systems, and other forms of measurement play prominent roles in the life of every human being. Measurement, in its various forms, controls man's life to a great extent. Day-to-day activities like buying and selling, cooking, stitching, constructing, diagnosing diseases and using medicines, and numerous other such acts require measurement in its different forms.

Ans to Q No 3: Two differences between physical measurement and educational measurement are:

- Physical measurement starts from a true zero point; whereas educational measurement does not have a true zero point. Zero in educational measurement is not absolute but relative.
- Physical measurement is direct, definite, and objective; whereas educational measurement is indirect, indefinite, and subjective.

Ans to Q No 4: Two limitations of educational measurement are:

- educational measurement is indirect, indefinite, and subjective; and
- educational measurement is incomplete, since measurement of achievement in any subject is always based on a sample of student behaviour.

Ans to Q No 5: Measurement in education performs various functions, including the main function of estimating change in student behaviour. Measurement in education helps in selection, classification, comparison, prediction, diagnosis, guidance, besides assessing the effectiveness of teaching and learning.



1.11 MODEL QUESTIONS

A. Short Questions (Answer each question in around 150 words)

Q 1: What do you mean by educational measurement?

Q 2: What are different types of measurement?

Q 3: What do you mean by scales of measurement?

B. Long Questions. (Answer each question in around 300-500 words)

Q 1: What is measurement? Distinguish between physical and psychological measurement.

Q 2: Discuss the nature of educational measurement.

Q 3: Discuss the functions of Educational Measurement

Q 4: Make a comparison between the different scales of measurement.

*** ***** ***

UNIT 2: PSYCHOLOGICAL TEST

UNIT STRUCTURE

2.1 Learning Objectives

2.2 Introduction

2.3 A Brief History of Psychological Testing

2.4 Meaning of Psychological Test

2.4.1 Nature of Psychological Test

2.5 Test, Examination and Evaluation

2.6 Types of Psychological Test

2.7 Functions of Psychological Test

2.8 Uses of Psychological Test

2.9 Meaning of a Good Test

2.9.1 Criteria of a Good Test

2.9.2 Practical Criteria of a Good Test

2.9.3 Technical Criteria of a Good Test

2.10 Let Us Sum Up

2.11 Further Reading

2.12 Answer to Check Your Progress

2.13 Model Questions

2.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- explain the concept of psychological test
- trace the history of psychological Testing
- explain the nature of psychological Test
- distinguish between Test, Examination and Evaluation
- describe the different ways that Psychological Tests can be classified
- describe the various functions of Psychological Tests
- state the various uses of Psychological Tests

2.2 INTRODUCTION

Psychological tests are of immense use in the field of education, above and beyond their use in other fields of human activity. Psychological tests are mainly constructed and used to study intelligence, aptitude, achievement, and various other scholastic and non-scholastic personality traits of individuals and groups in the field of education. These tests help to determine and analyse the nature of individual differences among learners in terms of intellectual development, academic and non-academic personality characteristics associated with age and sex. Further, analysis of the personality characteristics of gifted, retarded, or educationally backward, as well as the personality traits of the neurotics and psychotics are possible through psychological tests.

Educational institutions widely use psychological tests for making various decisions that involve classification, prediction, comparison, and selection of students and also for making administrative policies and management of the institution. Moreover, psychological tests are useful in providing educational and vocational guidance and counselling to students on the basis of the findings made available by test results about the strengths, weaknesses, and other psycho-physical traits of pupils. Psychological tests are also of use in employment, business, industry, and many other areas of work that affect the lives of many people. In this unit, you will be able to trace the history of psychological testing and will also be able to understand the concept, nature, function and uses of psychological test.

2.3 A BRIEF HISTORY OF PSYCHOLOGICAL TESTING

The history of psychological testing is very long and can be traced back to around 2200 B.C. when the Chinese emperors used rudimentary forms of tests to select officials and to examine their fitness for office every third year. Such testing, through modifications over the centuries, ultimately took the form of written exams that were introduced during the Han dynasty

(202B.C.-A.D. 200). However, psychological testing in its modern form was initiated only during the era of brass instruments that flourished in Europe during the late 1800s. During that period, many physicians and psychiatrists developed standardized procedures to study the symptoms of the mentally ill and brain-injured individuals. One of the most renowned among them, was Hubert von Grashey, who developed the precursor to the memory drum to test the visual recognition skill of brain-injured patients in 1885. Francis Galton attempted to measure intelligence by means of tests of sensory and imagery discrimination. Galton also developed his famous 'Galton whistle' for measuring sensitivity to high pitched tones. In 1879, Wilhelm Wundt founded the first psychological laboratory in Leipzig, Germany. Wundt also developed a thought meter, with which he attempted to measure the speed of thought. James Cattell, an American psychologist, made the first reference to mental tests in one of his classic papers published in 1890. French physicians, Esquirol and Seguin, working on the diagnosis and remedy of the mentally retarded, established the necessity for intelligence tests in the late 1800s. However, it was Alfred Binet and his associate Theodore Simon who developed the first proper intelligence test in Paris, France in 1905.The first Binet-Simon intelligence test was a simple 30-item test which was devised mainly to identify schoolchildren who could not profit from normal instruction. The original Binet test of intelligence was revised in 1908, and in 1911.The original French Binet-Simon intelligence test was later adapted in the Stanford University, USA by Lewis Terman which was known as Stanford-Binet intelligence scale. This scale was first introduced in 1916, and was subsequently revised in 1937, 1960, and in 1972. German psychologist William Stern introduced the concept of Intelligence Quotient which was to be obtained by dividing the mental age by the chronological age. Later, Terman suggested multiplying the intelligence quotient by 100 to remove fractions and thus the concept of IQ was born.

2.4 MEANING AND DEFINITION OF PSYCHOLOGICAL TEST

Measurement and evaluation in education is the process of comprehensive and scientific assessment of the intended changes in learner behaviour that are credited to instruction or education. This process of measurement and evaluation provides necessary information about the progress of students towards the achievement of educational goals. Neither the student nor the teacher can progress satisfactorily towards educational goals without such knowledge. Hence, different tools of measurement are used to collect such knowledge and information. A psychological test is one of the most useful tools used in educational measurement and evaluation.

A psychological test is a collection of some questions (items) to be answered or tasks to be performed by the testee (examinee). Such collection of questions or tasks is used to measure intelligence, knowledge, skill, or other traits of personality of an individual or a group. This collection of questions or tasks (test) is required for the reason that intelligence, achievement, or other such trait or characteristic cannot be measured directly unlike the measurement of the objects of the physical world. Measurement of the objects of the physical world is direct and easy because such measurement can be done using universally acceptable standard tools and units. Conversely, psychological measurement that involves the measurement of knowledge, skill, aptitude, achievement, or any other trait of the personality cannot be measured until and unless they are manifest in the behaviour of man. Therefore, it is required to make the abstract trait or quality to be measured explicit in the behaviour of the individual so as to make it measureable, and hence some stimuli in the form of questions or tasks are given to the testees so that their responses could be observed, recorded and measured. The questions to be answered or the tasks to be performed are so selected or framed that the response of the testee to such question/task could make the trait to be measured explicit and measureable. Thus, a psychological test may be understood as a set of

carefully selected stimuli in the form of questions to be answered or tasks to be performed to elicit a sample of behaviour required to measure a particular psychological trait or characteristic. The definition given by psychologist Marshall is noteworthy in this context, he said, "A psychological test may be defined as a pattern of stimuli, selected and organised to elicit responses which reveal certain characteristics in the person who takes them." However, the definition given by Marshall does not include all the features of a psychological test. Hence, some important definitions of psychological test are presented below:

According to **Frank S. Freeman,** "A psychological test is a standardized instrument designed to measure objectively one or more aspects of a total personality by means of samples of verbal or non-verbal responses, or by means of other behaviours."

Psychologist **Anastasi and Urbina** have defined a psychological test as an objective and standardized measure of a sample of behaviour.

A.K.Singh writes, "A psychological or educational test is a standardized procedure to measure quantitatively or qualitatively one or more than one aspect of a trait by means of a sample of verbal or non-verbal behaviour.

In the words of **McIntire and Miller,** "Psychological tests are instruments that require the testee to perform some behaviour. The behaviour performed is used to measure some personal attribute, trait or characteristic - such as intelligence - that is thought to be important in describing and understanding the behaviour."

Analyses of the foregoing definitions of psychological test help us to conclude that a test consists of some items or tasks to be performed by the testee which elicits a sample behaviour of the testee. Such sample of behaviour indirectly reveals the personality trait or characteristic which is being measured.

2.4.1 Nature of Psychological Test

A psychological test is one of the various tools used in psychological

measurement. A psychological test is used as an instrument to elicit response or behaviour of an individual required for the purpose of measuring some trait, or characteristic of the individual who takes the test. The following important characteristics reveal the nature of a psychological test-

- A psychological test may be understood as a set or pattern of stimuli, selected and organised to elicit responses that reveal certain trait, or characteristics in the individual whose behaviour is to be measured.
- Psychological tests are constructed and used to measure psychological variables which are abstract, unobservable, and cannot be measured directly. Hence, psychological tests are tools of indirect measurement, i.e. they measure psychological variables as the variables become manifest in the behaviour of the individual.
- Psychological tests require a testee to perform some observable and measurable behaviour. When a testee takes a test, he has to answer some questions or to perform some task; and in doing so, the testee is revealing some trait or characteristic in the form of some observable and measurable behaviour.
- In a psychological test, the behaviour which is thought to be important in describing or understanding the trait or characteristic is used to measure that trait, or characteristic.
- In psychological testing, scores or categories are assigned to individuals to represent abstract dimensions of behaviour, traits, or performance of individuals. In fact, psychological testing represents performance or dimensions of behaviour in the form of numbers or classifications.
- A psychological test invariably follows standardized procedure. In simple terms, the procedure for administering a psychological test is uniform in all settings and with all examiners. All psychological tests usually accompany an instructional manual which contain detailed directions for administration of the test.

- A psychological test always measures a limited sample of the behaviour which is important in describing or understanding the trait or characteristic to be measured. Since it is not possible to measure the entire domain of the behaviour due to practical constraints, a test always depends on samples of behaviour to make inference about the total domain of the relevant behaviour.
- Psychological tests aim to measure individual difference in terms of performance, traits, or characteristics. Though all people possess the trait or characteristic being measured, they possess it in different amounts. The objective of psychological test is to measure the amount of the trait or quality possessed by an individual, so that individual difference can be estimated with reference to the trait or quality being measured.
- Psychological tests also accompany norms or standards calculated on the basis of average performance of a group on the same test. Such norms help in interpreting the score or performance of an individual by comparing it with the norm for the group to which the individual belongs.
- The most important feature of a psychological test is that it makes inference about the total domain of behaviour on the basis of the measurement of the sampled behaviour. Thus it makes predictions about the non-tested behaviour or performance, or trait of the testee.
- The construction and administration of psychological tests, is usually based on the following assumptions:
  - Psychological tests measure what they intend to measure, and the inferences made about testees on the basis of their responses to thetest are correct. This is known as the validity of a test.
  - An individual's responses to the test and his testscores remain constant over time. This feature of a test is known as reliability.

- The testees understand the meaning of test items as intended by the tester or the constructor of the test.
- It is also assumed that testees can and will report about themselves and their thoughts and feelings accurately and honestly.
- The test score of an individual is the sum of his or her actual ability and some errors that may result from the test itself, faulty administration of the test, subjective influence of the examiner, psycho-physical condition of the examinee, or adverse environmental influences.

However, it is to be borne in mind that these are only assumptions and may not always prove to be true in the case of any test. The quality of a psychological test depends on the extent to which these assumptions actually come out to be true. It is here that the question of accuracy and authenticity of psychological tests arises and the accuracy, authenticity, and exactness of a tool of physical measurement cannot be claimed by a psychological test.

2.5 TEST, EXAMINATION AND EVALUATION

Test, examination, and evaluation are similar terms and are used interchangeably at times. However, these similar terms have different meanings. Test and examinations are tools of evaluation. As tools of evaluation test and examinations have differences in their meaning and scope. As an important tool of evaluation, a test is a set of items or questions to be answered or tasks to be performed on a certain area or topic. The testee is evaluated on the basis of his answers to the questions or on the basis of his performance on the assigned tasks. An examination too, is a set of questions or tasks; however, an examination includes many areas or topics unlike a test which is focussed only on a single area or topic. Thus, the scope of an examination is wider than a test. An examination can be a collection of many interrelated or interconnected tests. An examination may measure various psycho-physical qualities of examinees, but a test

measures a single dimension of behaviour or personality of the testee. In educational arena, a test and an exam both test the knowledge of a student. So, in most cases tests and exams are synonyms.However, examinations are held at the end of the academic session which includes all subjects or disciplines. On the other hand, tests are held during the academic session which is based on particular units or topics of the curriculum. Moreover, tests are generally diagnostic and formative in nature that aim at checking achievements after a lesson or series of lessons, while examinations are summative and they help in making decision about promoting or retaining students at the end of the academic session.

A test and an exam are both tools and instruments of evaluation. A test aims at measuring a particular portion or part of an object, or of a subject. Examination, in contrast, aims at measuring the object or subject in its entirety. For example, a physician makes an examination of a patient to diagnose diseases. Such physical examination involves many tests, such as, blood test, urine and stool test, test of heart beat, and the like.

Evaluation is a concept which is wider than test and examination. Various tests and examinations are used as tools for evaluation. A test or an examination is usually concerned with any one dimension of the cognitive, affective, or psycho-motor behaviour of the individual. Whereas, evaluation is a comprehensive process and it takes into account all the aspects including cognitive, affective, as well as psycho-motor of behaviour of the individual. Besides, evaluation is a continuous process which uses the information obtained from different tests and examinations held from time to time.



CHECK YOUR PROGRESS

Q 1: What is psychological test?

.....

Q 2: Write two features of psychological test.

.....

2.6 TYPES OF PSYCHOLOGICAL TEST

Psychological tests have been classified into different types on the basis of various criteria. Hence, there is no single classification of psychological tests and the different classifications are not exclusive but they often overlap. The important fact to note is that psychological tests can be classified by the type or nature of the behaviour they measure, the way in which they are administered, the method of their scoring, and how they are used.

Classification on the basis of what they measure or the nature of behaviour measured:

Psychological tests have been classified on the basis of what a test measures or predicts. On the basis of this criterion, psychological tests have been broadly categorized into tests of Typical Performance and tests of Maximum Performance.

A test of typical performance requires the testee to express what they would normally do in a particular situation. In other words, this is a test to find out how testee operate or perform in a given situation. In tests of typical performance answers given by testee are not categorized as right or wrong, but are used to find out the choices, preferences, or feelings of the testee. The tests of typical performance usually measure non-cognitive behaviour, such as personality, interests, motivation, values, and attitudes.

Tests of maximal performance usually measure cognitive behaviour of the individual, such as tests of intelligence, aptitude, achievement and the like .The test items in these tests have either correct or incorrect answers and the respondents are asked to answer correctly as many questions or items as they can. Their scores on the test depend on the number of items they get correct. These tests aim to measure the maximum performance of individuals under standard conditions.

The tests of typical performance or the non-cognitive tests can again be classified into **structured or unstructured tests** on the basis of the nature of the stimuli presented before the respondents. A structured test includes structured items, i.e. items to which the respondents can respond

in predefined or limited number of ways. For example, in a test of personality, respondents may be asked to respond in the forms of 'yes/no' or 'true/ false' to questions about various activities or issues of their lives. These are examples of structured questions. On the other hand, unstructured tests use unstructured stimuli or items to which the respondent can respond freely or in ways that are not limited by the tester. For example, the projective tests of personality use unstructured stimuli such as an inkblot or a picture and the testee is asked to describe what they see or imagine the inkblot or picture. The basic idea behind these projective tests is that when unstructured or ambiguous stimuli are presented before an individual, he/ she will project his or her underlying and unconscious motivations and attitudes. However, the scoring of the unstructured test is more complex than the structured tests.

Cognitive tests may again be classified into **ability tests and achievement tests**. Though both of these two types of tests involve learning and involve what the testee has learned and can do, achievement tests usually involve learning from specialized education and training; while, ability tests usually assess learning that has occurred in general and in one's greater environment.

Further classification of Cognitive tests is done on the basis of **speed tests and power tests.** A speed test contains such items that everyone can get all items correct when they are given enough time. Some tests of clerical skills are examples of this type of tests. On the other hand, Power tests contain difficult questions and enough time is given. These tests intend to measure how much the testee knows or can do; instead of testing the speed at which the testee can answer the questions. Hence in a power test sufficient time is given; and it is measured what they can do. However, there are very few tests that either measures speed or power. Most of the psychological tests are a combination of both.

Classification on the basis of Test Administration:

Psychological tests can again be classified on the basis of how they are administered. Psychological tests can either be administered to an individual at a time or be administered to a group at the same time.

Tests which are administered to one individual at a time is called individual test. The individual test may be verbal, non-verbal, or performance type of test. It was Alfred Binet who popularized the individual test. Binet-Simon tests of intelligence, Alexander Pass-along test, Koh's Block Design test, etc are examples of individual tests. Individual tests help in establishing proper rapport with the testee that helps to increase the accuracy of measurement. Moreover, individual tests are more valid and reliable. However, individual are costly, time-consuming, requires specialists to administer, and most of them are verbal tests that cannot be administered to young children or to people who do not know the language used in the test.

In contrast to the individual tests, group tests are administered to a group of people at a time. These tests may also be verbal or non-verbal tests. Examples of group tests include, Army-Alpha test, Army-Beta test, Army General Classification Test (AGCT), among others. Group tests are useful for testing large number of people at a time. Administrations of these tests are simpler than individual tests and hence they don't require experts for the purpose. Thus, these tests are economical in terms of time, money, and effort required. However, group tests are not as much reliable and valid as individual tests are.

Classification on the basis of use of language:

Psychological tests may be classified on the basis of use of language as well. On the basis of use of language by psychological tests, they have been classified into V**erbal, Non-Verbal, and Performance tests.**

Verbal tests are oldest among all tests. These tests use written as well as spoken language. In verbal tests, generally written questions or instructions are given and the answers must be given in spoken or written language. Since testee have to respond using language, these tests are influenced by their knowledge of the language used in the test. Tests of comprehension, memory, vocabulary, word association, etc. are examples of verbal tests. Usually the group tests are verbal tests. During the recent times, some verbal tests are administered using computers and other electronic media. Verbal tests can be very useful in testing many cognitive

qualities of individuals, such as memory, thinking, reasoning, understanding, analysis, synthesis, and many others. However, since these tests demand mastery over language, verbal tests cannot be successfully used with young children, uneducated, or people who use language other than the one used by the test. Moreover, the reliability and validity of verbal tests are also not always beyond doubt.

Contrary to the verbal tests, Non-verbal tests do not use written or spoken language. Even the instructions to testees are given using demonstration, pantomime, gestures, or using some charts. In place of language, test items are framed using pictures, shapes, symbols, or structures. It is important to note that, though non-verbal tests do not use language, they may use paper and pencils. However, use of paper-pencil in non-verbal tests is limited to making non-lingual signs such as, 'X', "", or other such marks to respond to the items. Examples are: Chicago non-verbal test, Army Beta test, Raven's Progressive Matrices Test, etc. Non-verbal tests are useful with young children, uneducated people or with those who use a language other than the one used in the test. However, one limitation of non-verbal test is that these tests may not be useful to assess all mental or psychological qualities of people.

Performance tests, as the name suggests, require the testee to act or perform. In these tests, language is used only to instruct the testee as to what they are required to do or perform. The items in these tests involve manual acts or problem solving activities. They typically involve manipulating objects, tracing mazes, placing pictures in the proper order, and finishing patterns, etc. These tests are used to measure traits that cannot be measured using verbal tests. Performance tests are in general individually administered test. Some famous performance tests are: Bhatia Performance Test, Porteus maze test, Koh's Block Design test, etc. Performance tests are very much useful for testing young children, uneducated person, or people with linguistic difference or problems. These tests help in identifying psychological and neural disorders, besides testing motor abilities and skills of people. However, performance tests are costly and cannot be administered in groups. Another weakness of these tests is that their reliability is not very high.

Classification on the basis of the structure of the items:

Psychological tests can also be classified on the basis of the structure of the items. Accordingly, tests may be objective as well as subjective. Objective type tests include recall or recognition type items of multiple choice, alternate response, or other such type of items. Objective type items are generally very short and simple that demand answers in a single word or sentence. On the other hand, subjective tests include open ended questions. Unlike answers to objective items which are fixed, predetermined, and standard, answers to subjective items are not fixed or predetermined. They require the testee to answer using their subjective feelings or understanding. Objectives test items are valid, reliable, and easy to score but they cannot give much deep information about the testee. On the other hand, though subjective items are difficult to score and sometimes impossible to quantify, they can reveal deep information about the testee.

Classification on the basis of area or content of measurement:

Psychological tests can also be classified according to the area or content of measurement. For example, there are intelligence tests, achievement tests, interest tests, personality tests, etc. The Mental Measurements Yearbook, USA (published every five years) which contain information on English-language psychological test across the world uses 18 content classifications do categorize tests; they are:

- Achievement
- Behaviour Assessment
- Developmental
- Education
- English & Language
- Fine Arts
- Foreign Languages
- Intelligence and General Aptitude
- Mathematics
- Neuropsychological
- Personality

- Reading
- Science
- Sensory-Motor
- Social Studies
- Speech and Hearing
- Vocations
- Psychological Tests and Measurements

Classification on the basis of functions:

Psychological tests have also been categorized on the basis of their function. Some psychological tests are used to diagnose the present psychological conditions of individuals. For example, tests may be used to determine if the testee is suffering from behaviour or character disorder; if the testee is depressed; and the like. Such tests are called **diagnostic tests.** In education, diagnostic tests assess students' understanding of a subject, area or skill. Some other types of tests, called prognostic tests, are used to make predictions. **Prognostic tests** help to identify or categorize people with different risks of specific future conditions. In education such test predicts how well a person is likely to do in a certain school subject or task.

There are other categories of psychological tests such as selection test, placement test, test of classification, test of certification, etc. which are named after the functions they perform.

Classification on the basis of the method of Score interpretation:

Psychological tests have also been classified on the basis of how scores on a test are interpreted. When the score of an individual is compared with the scores of a group taking the same test, then that is called a norm-referenced test. Norm-referenced tests give information about the relative standing of individual in a group. In criterion-referenced tests, scores of each testee is compared to a pre-determined set of criteria or a standard rather to the other members of the test-taking group. Such tests aim to assess if the testee has the attained mastery over a certain skill or skills. These tests help in making decision about "pass" or "fail" and certification. Different board exams such as Law, Medicine, etc are examples of criterion-

referenced tests. If the examinees demonstrate mastery of the skills to practice the profession, they are certified or licensed, otherwise not.

2.7 FUNCTIONS OF PSYCHOLOGICAL TEST

Psychological testing performs many important functions in today's world. The significance of psychological testing extends far beyond the field of education alone. Psychological tests measure the ability of individuals to perform tasks in various fields such as engineering, medical, musical, reading, clerical, spatial, numerical, dextrous, and many others. Psychological tests serve both diagnostic and prognostic functions that help in educational, vocational, as well as personal guidance and counselling that help to build realistic self-image among students, set educational and vocational goals and provide important data and crucial information that help in placement and making reasonable decision about career and occupational choices. Besides, psychological tests measure academic progress of students, help in classifying students on the basis of their ability, help screening students for various courses and career, and help students to develop self-understanding and make personal decisions accordingly. The main functions of psychological testing may be explained in the following ways:

- **Diagnosis:** One of the main functions of psychological tests is diagnosis. Psychological tests diagnose the present psychological conditions of individuals. For example, a diagnostic test may be used to determine if the testee is suffering from behaviour or character disorder; or if the testee is depressed; and the like. In education, diagnostic tests assess students' understanding or skill; their possible weakness or backwardness in any subject or area.Information concerning students' strengths and weaknesses helps in making better plans about what and how to teach students.
- **Prediction:** Psychological tests can identify or categorize people with different risks of specific future conditions. Such tests are called prognostic tests. In education such test predicts how well a person is likely to do in a certain school subject or task.

- **Providing information for formative and summative evaluation:** Psychological tests provide vital information required for conducting formative and summative evaluation. Formative evaluation is very useful in studying if a program is working or not; and if not, then why is it not working, while the program is in progress. Formative evaluation helps the program planners, designers, learners, instructors, and managers to monitor how well the program goals and objectives are being met. It helps to redress the deficiencies in the program as early as possible while the program is in process.Summative evaluation is carried out after the completion of the program to find out the overall success or failure of the program. Psychological tests of different types supply data for both, formative and summative evaluation.
- **Classification:** Psychological tests supply data for classifying individuals on the basis of various traits such as interest, aptitude, attitude, intelligence, or any other trait which is relevant for the purpose. In the educational field, classifying students with various administrative purposes is very important which is aided by data obtained from psychological tests.
- **Selection:** Psychological tests are very useful in making selection of individuals for admission, recruitment, or other activities on the basis of requisite knowledge or skills for the course or career.
- **Comparison:** Psychological testing help in making comparison between individuals on the basis of their achievement, intelligence, personality, temperament, and such other traits. Such comparisons also help to know about individual differences among people.
- **Instructional or Educational function:** Psychological tests are useful in making various instructional or educational decisions. Knowledge gained through the use of psychological tests help in making proper choice of educational and instructional objectives. They also help in making better plans for teaching learning strategies. Moreover, data obtained from tests help to improve curriculum, assess the effectiveness of teaching methods, and evaluate the

extent to which objectives have been attained. Besides, various practical and administrative decisions concerning admission, selection, comparison, prediction, promotion, etc depend on information collected from psychological tests of various kinds.

- **Guidance and counselling:** Psychological testing is indispensable for providing guidance and counselling in the areas of education, vocation, and personal life problems. Tests of interest, ability, intelligence, personality, etc are a must for any success of guidance and counselling program.
- **Research and development:** Psychological tests play important part in research and development. Various psychological tests are required to conduct research on human behaviour. At the same time, tests of various psychological traits may suggest further research leading to new discovery about human behaviour and personality. Such research may help in the development of new theories, principles, or laws to explain and understand human behaviour and personality.



CHECK YOUR PROGRESS

Q 3: Write about any two types of psychological test.

.....

Q 4: What is a diagnostic test?

.....

Q 5: Write two functions of psychological test.

.....

2.8 USES OF PSYCHOLOGICAL TEST

Psychological tests are of immense use in various fields of human activities. The important uses of psychological tests may be mentioned as under:

- Psychological tests are widely used to determine and analyse individual difference in intelligence, aptitudes, achievement, interest, attitudes, vocational fitness, non-cognitive personality characteristics and similar other traits.
- Psychological tests are used in various educational, psychological, cultural, sociological, and employment studies.
- Psychological tests are used in the study of cognitive or mental development, personality characteristics associated with age and developmental stage, sex, race, heredity and environment, intelligence, education and other factors.
- Psychological tests are used to study the nature and characteristics of the normal, the gifted, the retarded, the neurotic, and the psychotic individuals.
- Psychological tests are also used in educational planning, classification, selection, prediction, placement, promotion, evaluation, and in certification.
- Psychological tests are used in educational, vocational, and personal guidance and counselling.
- Psychological tests are used to detect and to measure the special abilities of a person.
- Psychological tests are used in clinical settings such as in mental hospitals to diagnose mental disorders. MMPI, (Minnesota Multi-phasic Personality Inventory), RISB, (Roter Incomplete Sentences Blank), etc. are examples of such clinical tests.
- Psychological tests are used in legal classification of people into different categories such as, normal and abnormal, criminal and innocent, intellectual and mentally retarded, able and disable etc.
- Psychological tests are used to diagnose factors associated with problems of learning, behaviour, attitude, or adjustment.
- Psychological tests are also used in business and industry to select, recruit, and classify personnel for placing them into different branches of job.

- Psychological tests are used in scientific research and investigation.
- Psychological tests have been continuously used in selection of military personnel since the First World War.

The forgoing discussion reveals that psychological tests are very useful in different areas of human activities such as, education, vocation, diagnosis mental disorders, and hence they affect the lives of many people significantly.

2.9 MEANING OF A GOOD TEST

A good test objectively measures what it intends to measure and gives consistent results. In other words, a good test is one that is valid, reliable, objective, practical, and friendly to the test-taker. A good test also fulfils the criteria of standardization. This implies that good tests are administered under uniform condition and scored objectively in order to make the scores obtained by individuals who take the test comparable. Procedures of administration and scoring of good tests are specified in detail in order to avoid any subjectivity. Moreover, a good test isone that has been tried out with large groups of individuals, whose scores provide standard norms or reference points for interpreting the scores obtained by those who took the test. Good tests are administered in a standard manner under uniform procedure. These tests are tested and re-tested and have been proved to produce valid as well as reliable scores. Furthermore, good test provides proper attention to the respondents to make it convenient for them. A good test looks professional but at the same time it is easy to administer.

Thus, a good test may be understood as a test administered under uniform conditions and interpreted in a consistent manner, i.e. all the key aspects of a good test are uniform. The same test is administered to all examinees, the conditions under which the test is administered (time, resources, etc.) are standardized, and the ways in which scores are interpreted are also standardized.

2.9.1 Criteria of a Good Test

Whenever we need to construct a test or to select a test from the available ones, we must see that the test has certain good points or characteristics. In this regard it is important to note that the criteria of a good test may be categorized into two categories: the practical criteria or the usability of the test, and the technical criteria. These two categories of qualities may be presented as below-

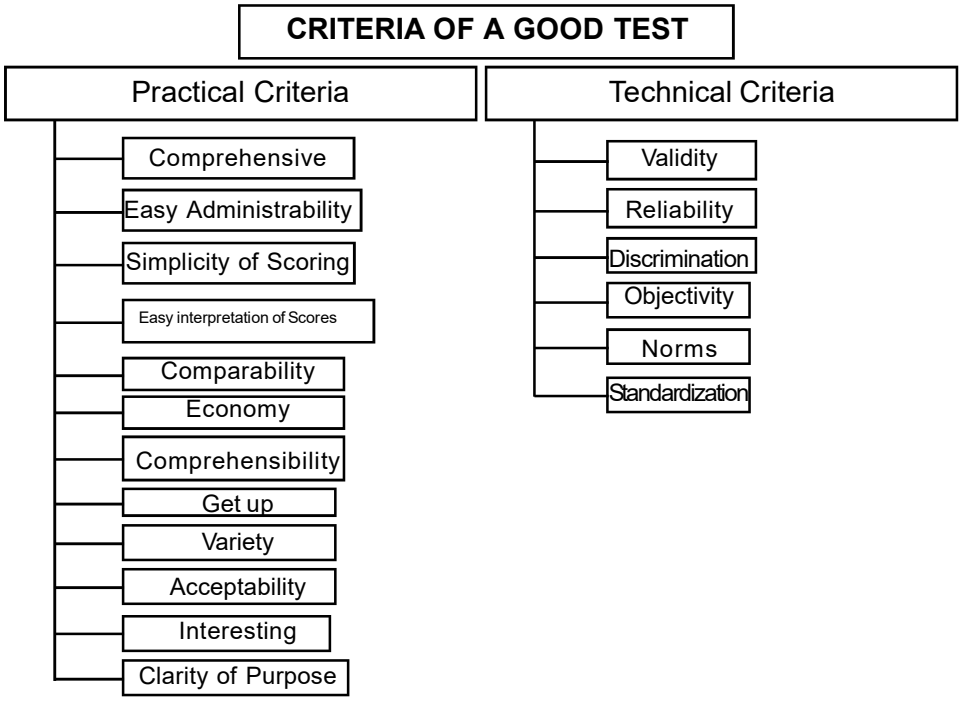


Figure 2.1: Criteria of Good test

It is important to note at this point that the practical as well as the technical criteria mentioned above are very important to make for a good test. Though, the technical criteria of a test are more important than the practical criteria, in the absence of the practical criteria a test may become unusable even if it is of technical quality. Similarly, a test which is highly practicable may not give useful results in the absence of technical qualities. Hence, both practical and technical criteria are important to make a good test.

2.9.2 Practical Criteria of a Good Test

Practical criteria of a test refer to those qualities of a test which make the administration of the test simple, smooth and convenient. The practical criteria of a good test include the following:

- comprehensive
- easily administrable
- simplicity of scoring
- easy interpretation of scores
- comparability
- economy
- comprehensibility
- get-up
- variety
- acceptability
- interesting
- clarity of purpose

A good test is comprehensive: A good test is always comprehensive enough to include items representing each of the abilities and objectives so that it can measure what it is supposed to measure. If a test does not represent all the content or curriculum and objectives of instruction, it can never claim to be a good test. In order to make the test representative and comprehensive, different types of items should find place in the test.

A good test is easily administrable: A test which can be administered smoothly with ease may be called a good test. The directions for administration of a good test are always clear and precise. The test booklet of a good test includes clear description about the necessary preparation for the administration of the test, the method of distribution and collection of test-booklets in simple and clear cut language.

A good test is characterized by simplicity of scoring: The scoring procedure of a good test must be simple. The process

of awarding marks against the responses to the test items should be simple and fast. In order to make the process of scoring simple, fast and easy, good tests should be accompanied by well-constructed 'scoring key'. Scoring keys help in fast and objective scoring of test-booklets.

Easy interpretation of scores: Scores obtained by individuals in a good test can be interpreted easily. Easy interpretation of scores can be ensured by making provision for adequate norms. The scores obtained in a test should not be such that cannot be interpreted or understood easily.

Scores obtained by a good test are comparable: If the scores obtained by different individuals who take the test cannot be subjected to comparison, the worth or value of the test becomes questionable. Good tests are administered under uniform condition and scored objectively which make the scores obtained by individuals or groups in the test comparable.

Good test are economic: A test must be economic in terms of time, money and effort required to administer, score, and interpret the test results. If a test possesses high technical qualities but is very costly, requires much time and effort, then the test cannot claim to be a good test.

Good tests are comprehensible: A good test is comprehensible or intelligible to the testee as well as testers. The items and the instructions should be clear and unambiguous. The processes of administration, scoring and interpretation of good test are easily understandable to the administrator or tester.

Attractive physical get-up: The physical get up or the cover of a good test is attractive and beautiful. The quality of paper used and the print should be of high quality.

Good tests provides for variety: Good tests include a variety of test items: multiple choice items, objective type items and so on. It also includes variety of tasks and activities to be performed by the respondents such as writing, reading, speaking, listening,

re-writing, interpreting, filling blanks, matching, distinguishing, identifying, etc. The tasks and the materials used in the tests should be related to real life situation.

Good tests are acceptable: Acceptability is another feature of a good test. A good test must be acceptable to the testee, the tester and all other concerned parties.

Good tests are interesting: Good tests are interesting to the testees. A test must be interesting in order to motivate individuals to take the test and to retain their interest till the end of the test. Uninteresting and boring tests cannot yield intended results.

Purposes of good tests are clear: Good tests are characterized by clarity of its purpose. This is one of the most important characteristic of a good test. The purpose of the test must be very clear before the testee as well as the tester.

2.9.3 Technical Criteria of a Good Test

The technical criteria of a test are very important for raising the quality of a test to a high level of standard. The technical criteria of a test include:

- validity
- reliability
- discrimination
- objectivity
- norms
- standardization

Validity:

Validity is the most important technical criterion of a good test or other measuring instrument. Cronbach says, "Validity is the extent to which a test measures what it purports to measure." This means that the validity of a test is the extent to which a test measures what it claims to measure. Validity refers to truthfulness of a test. In other words, validity answers the question of whether the test

measures what it intends to measure. The concept of validity is so fundamental that it precedes the aspect of reliability in importance. A test may be reliable but not valid. However, if a test is not valid, then its reliability has no meaning or importance. On the other hand, a valid test is usually reliable. The concept of validity is not absolute and hence validity of a test is always understood or determined with reference to the purpose of the test. Therefore a test may be valid for one purpose but not for others. Any test can be differentially valid, given the various uses to which it is put. This is why F. S. Freeman defines validity as, "An index of validity shows the degree to which a test measures what it purports to measure when compared with accepted criterion." For example, a test constructed to assess the capacity for logical thinking of adolescents may serve the purpose extremely well, but when the same test is given to students of a different age group it may fail to give desired results.

Reliability:

Reliability is another very important technical criterion of a good test. It refers to consistency of results or findings. In the words of Freeman, "The term reliability refers to the extent to which it gives consistent result on testing and retesting." When an investigation is repeated it must yield the same results if it were to be reliable. When more than one investigator or evaluator is making the measurement, all of them must arrive at similar findings in order to claim that the data are reliable. Similarly, when an individual takes the same test twice, if the scores on the two occasions are very similar then the test can be described as reliable. This means, a reliable test must give the same result when given twice to the same person within a short period of time and when different evaluators make the measurement. Anastasi says, "Reliability refers to the consistency of scores obtained by the same individuals when re-examined with the same test on different occasions or with different sets of equivalent items or under variable examining conditions." If a test or other measuring device gives different results on different occasions

or when used by different evaluators on the same individuals, then it should be understood that the test is not reliable.

Reliability of a test refers to the consistency of scores obtained by students on alternate forms or different administration of the same test. However, no two tests with exact content can ever give identical results due to differences in the environmental conditions including psycho-physical state of the test-takers. Even the same test gives different results even when administered to the same group of students at two different occasions. Nevertheless, whenever we administer two different tests with similar material or the same test on different occasions, we wish that the scores obtained by students in the two different administrations are similar. In this situation, the test is reliable to the extent to which the scores are comparable.

Reliability of a test is an indicator of validity as well. If a test cannot give consistent results, it is obvious that the test is not accurately measuring the trait or variable of interest. A test to be valid must be reliable. Therefore, estimation of reliability is considered to be the first-step in the process of validating a test. If the test is not reliable, then there is no need of examining if the test is valid and it can be easily concluded that the test is not valid. Several methods are used to estimate reliability of a test depending on the nature of the measurement.

Discrimination:

A good test must be discriminating. Tests are constructed and administered with the purpose of studying the differences among individuals or groups besides other purposes. Hence, a good test must measure differences in achievement of individuals. Tests which are discriminating can separate the good students from the poor students. In order to enhance the discriminating power of a test, it must have a wide range of scores from the lowest to the highest. Moreover, the test should have items with different levels of difficulty. In more simple words, the test should have items of highest

difficulty level, which will be answered correctly only by the best students along with very easy items that can be answered correctly by all the students. Each item of a good test must be able to discriminate between the students who know more and the students who know less. If an equal number of good as well as poor students answer an item incorrectly, the item fails to discriminate between good and poor students. If an item is answered correctly by poor students and incorrectly by good students then the item is making negative discrimination. Such items must be ambiguous and faulty and hence should be deleted readily or modified to make positive discrimination between good and bad students. Analyses of items after the administration of a test help in determining the relative difficulty and discriminating power of each item. Hence test construction must analyse items carefully and extensively so that items can be made discriminating and of proper difficulty level.

Objectivity:

Objectivity is another important technical criterion of a psychological test. A good test is always objective. Objectivity is like an important precondition to validity and reliability of a test. The scores obtained by an objective test are not influenced by the mental or physical condition of the scorer. Since the items in an objective test have only one correct answer, there is no scope for subjective influence in scoring process. Objectivity of a test has two different aspects: (i) objectivity of test items, and (ii) objectivity of scoring.

Objectivity of test item indicates that the items are very clear, straight forward, and unambiguous and each item has only one correct answer. An objective item should not have more than one correct answer in any condition or situation. Language of the test items should be simple, clear, and easily understandable. There should be clear instructions to the testees as to what they should do or how they should respond to the items. Clear instructions to the testees can ensure consistent score on an answer script even if the answers are checked by different examiners. In order to ensure

objectivity of items it should be kept in mind that ambiguous, essay type and items with unclear instructions may lead to subjectivity.

Objectivity in scoring refers to objective scoring of test booklets or answer scripts. In objective system of scoring, irrespective of the person who makes the scoring, it is possible that the same score is obtained on an answer script. In objective system of scoring, the individual judgement or subjective influence of the individual cannot influence the score. Usually a scoring key is prepared before the administration of the test in order to make the scoring process objective. The scoring key contains all the answers and marking pattern for each correct or incorrect answer to the test items.

Objective items make objective scoring possible. There is a popular allegation against the traditional essay type items of being subjective. This is because of the fact that the scoring of such items is influenced by the mood of the examiner, language and handwriting of the examinee, Halo effect, and the like.

Norms:

A good test is always accompanied by suitable norms for interpreting the scores obtained by individuals on a test and to make comparison between the scores. The crude or raw scores obtained by individuals on a test do not convey much meaning or significance. For example, a raw score of 60 obtained by a student in an examination has little meaning if information about the total marks in the examination, marks obtained by other examinees, the average score of students in the same exam, etc. remain unknown. Moreover, such a score cannot be directly compared with marks obtained by others or by the same individual in a different examination. In order to remove such difficulties and to interpret scores properly and to make the scores comparable, provision of appropriate norms is a must for a good test. A norm or a standard is the average score obtained by a homogeneous group in a test. The score obtained by an individual can be interpreted only by comparing it with the standard or norm set for the group.

Norm is the average or the representative score of a specified population on a particular test. According to Freeman, "A norm is the average or typical score (Mean or Median) on a particular test made by a specified population." For example, Mean score of 10 year old children on a test of intelligence, Mean score of a group of students of 8th standard on a test of mathematics, etc. Reference to such norms helps to interpret and compare the score of an individual besides assigning him a relative position or rank in his group.

There are four different types of norms generally used in tests, they are:

- Age Norm
- Grade Norm
- Percentile Norm
- Standard Score Norm

Age norm for a test is the arithmetic Mean of the raw scores obtained by individuals of a particular age group on the test. For example, the Mean score of a group of 11year old children on an intelligence test is the age norm for 11 year old children for that test. The concept of age norm is very simple and it helps in making comparison between test scores of individuals of the same age group. However, the problem with age norm is that it can be applied only to tests measuring the traits or attributes that tend to grow with the growing age of people.

Grade norm for a test is the arithmetic Mean of the raw scores obtained by a group of students studying in a particular grade on the test. For example, the Mean score of a group of 10th grade students on a test is the grade norm for 10th grade students for that test. The concept of grade norm is similar to the concept of age norm, but for grade norms the average performance is determined on the basis of the grade of the students unlike age norms which is determined on the basis of age. The concept of grade norm too is

simple and it helps to compare the performance of students studying in the same grade. The main limitation of grade norm is that it does not consider the potential individual differences among students studying in the same grade and fixes one single norm for all students of a particular grade.

Percentile norm is a process of converting raw scores of individuals into percentiles. A percentile is a score below which a given percentage of individuals in a group fall. For example, the 35th percentile is the score below which 35% of the individuals lay on the basis of their scores. The advantage of percentile norm is that it is easy to calculate and can be used in any type of test. However, the main limitation of percentile norm is that it can indicate the relative position of the individual in the group only and cannot express the degree or amount of achievement acquired by the individual.

In standard score norm the raw scores of students are converted into standard scores. The conversion is done by calculating the distance of individual student's scores from the Mean score of the group in Standard Deviation units. The most commonly used formula for converting raw scores into standard scores is -

$$z = \frac{x - \mu}{\sigma}$$

Where, Z=Standard Score

X=Raw score of a student

μ =Mean of all the scores

σ =Standard Deviation of all the scores

For example, if a student scores 80 in a test and the Mean and SD of all the students in that test is 60 and 10 respectively, then the Standard score of the student would be:

$Z = {}^{80-60}\!/_{10} = 2\sigma$

This means that the raw score of 80 obtained by the student is equal to 2 or the score is 2 higher than the Mean score. Thus, the raw scores obtained by individuals may be interpreted and compared by converting them into standard scores. However, the

concept of standard score norm is not very simple and requires statistical knowledge to use it.

Standardization:

Another important technical criterion of a good test is that the test must be standardized. Standardization of a test is the establishment of uniform procedures for the administration and scoring of the test. Test standardization is the process of determining and formulating the rules for test administration, scoring and interpretation. Unless a test is standardized, it is only an informal means of measurement that varies from examiner to examiner. A standardized test is one in which all test takers answer the same questions in the same way, the conditions under which (time, resources, etc.) the test is administered are standardized, scored in a uniform manner, and the ways in which scores are interpreted are also standardized which makes it possible to compare the relative performance of individuals who took the test.

A good test must fulfil the criteria of standardization. In order to standardize the test, it is---

- administered under uniform conditions to a very large group of students,
- scored objectively using a scoring key and the scores are arranged for analysis,
- Mean and other statistical values of the scores calculated and norm for the test is determined
- validity and reliability of the test is determined, and
- finally, the test manual or guidelines for administering the test is prepared.



CHECK YOUR PROGRESS

Q 6: What is a good test?

.....

Q 7: What is Validity?

.....

Q 8: Mention two practical criteria of good test.

.....



2.10 LET US SUM UP

After going through the units, it can be summarized as--

- A psychological test is a set of carefully selected stimuli in the form of questions to be answered or tasks to be performed to elicit a sample of behaviour required to measure a particular psychological trait or characteristic.
- The long history of psychological testing can be traced back to the time of the Chinese emperors (around 2200 B.C.) who used rudimentary forms of tests to select officials and to examine their fitness. However, psychological testing in its modern form was initiated only during the late 1800s. During that period, physicians and psychiatrists like Hubert von Grashey, Francis Galton, Wilhelm Wundt, James Cattell, and many others contributed towards the development of modern testing system. It was Alfred Binet and his associate Theodore Simon who finally developed the first proper intelligence test in France in 1905. The original French Binet-Simon intelligence test was later adapted in the Stanford University, USA by Lewis Terman which was known as Stanford-Binet intelligence scale. German psychologist William Stern introduced the concept of Intelligence Quotient (IQ) which was to be obtained by dividing the mental age by the chronological age.
- Psychological tests are tools of indirect measurement to measure psychological variables which are abstract, unobservable, and cannot be measured directly. Moreover, psychological test always depends on samples of behaviour to make inference about the total domain of the relevant behaviour. Any score of an individual on a test is the sum of his actual ability and some errors that may result from the

test itself, faulty administration of the test, subjective influence of the examiner, psycho-physical condition of the examinee, or adverse environmental influences. Hence, the accuracy, authenticity, and exactness of a tool of physical measurement cannot be claimed by a psychological test.

- Although test, examination, and evaluation are similar terms, they have different meanings. Test and examinations are tools of evaluation, which contain a wider, more scientific, and comprehensive process. A test is a set of items or questions to be answered or tasks to be performed on a certain area or topic. An examination too, is a set of questions or tasks; however, an examination includes many areas or topics unlike a test which is focussed only on a single area or topic.
- Psychological tests can be classified by the type or nature of the behaviour they measure, the way in which they are administered, the method of their scoring, and how they are used. Hence, there is no single classification of psychological tests and the different classifications are not exclusive but they often overlap.
- The varied functions of psychological test include: Diagnosis, Prediction, Providing information for formative and summative evaluation, Classification, Selection, Comparison, Instructional or Educational function, Guidance and counselling, Research and development etc.
- Psychological tests are used to study intelligence, aptitude, achievement, and various other scholastic and non-scholastic personality traits of individuals and groups. These tests help to determine and analyse the nature of individual differences among learners in terms of intellectual development, academic and non-academic personality characteristics associated with age and sex, personality characteristics of gifted, retarded, or educationally backward, as well as the personality traits of the neurotics and psychotics.

- A good test objectively measures what it intends to measure and gives consistent results. In other words, a good test is one that is valid, reliable, objective, practical, and friendly to the test-taker. A good test also fulfils the criteria of standardization. Good tests are administered under uniform condition and scored objectively in order to make the scores obtained by individuals who take the test comparable.
- While constructing a test or selecting one from the available tests, it must be seen that the test has certain good points or characteristics. The criteria of a good test may be categorized into two categories: the practical criteria or the usability of the test, and the technical criteria.
- The practical criteria of a good test include: Comprehensiveness or broad in scope, easily administrable, simple to score, scores are easily interpretable, comparability, economic, comprehensibility, attractive get-up, provision of variety, acceptability, interesting, and clarity of purpose.
- The technical criteria of a good test include: Validity, reliability, discrimination, objectivity, norms, and standardization.



2.11 FURTHER READING

1) Anastasi A. (1988): Psychological Testing. McMillan, New York Cronbach,
2) Asaad, Abubakar S. and Hailaya, Wilham M. (2004): Measurement and Evaluation, Concepts and Principles, 1st Edition, Manila, Rex Book Store, Philippines.
3) Cronbach, Lee J. (1960): Essentials of psychological testing, Harper & Brothers, New York
4) Freeman, F.S. (1965): Theory and Practice of Psychological Testing, 3rd Edition, Oxford and IBH Publishing Co. Pvt. Ltd., New Delhi
5) L. J. (1990): Essentials of psychological testing,5th Edition, Harper Collins Publishers: New York

6) Michael, A. Leonard (2016): Characteristics of a Good Test, Department of Library and Information Science, Delta State University, Abraka
7) Sandfeord, Peter (1939): The Educational Psychology on Objective Study, Longmans Edition
8) Singh, A.K. (2006). Tests, Measurements and Research Methods in Behavioural Sciences, Patna: Bharati Bhavan.
9) Stanley & Hopkins (1966): Education and Psychological Measurement and Evaluation, Prentice Hall Pvt. Ltd. Eaglewood Cliff, New Jersey

2.12 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: A psychological test is a standardized instrument that uses a set of stimuli to elicit a sample of behaviour which reveal certain characteristics or traits in the individual who takes the test. Such sample of behaviour indirectly reveals the personality trait or characteristic which is being measured.

Ans to Q No 2: Two important features of psychological test: i) Psychological tests are tools of indirect measurement; they measure psychological variables as the variables become manifest in the behaviour of the individual. (ii) Psychological tests depend on samples of behaviour to assess, the behaviour to be measured.

Ans to Q No 3: Two types of psychological tests are individual tests and group tests. Tests which are administered to one individual at a time is called individual test. Group tests are those that are administered to a group of people at a time.

Ans to Q No 4: Some psychological tests are used to diagnose the present psychological conditions of individuals. For example, tests may be used to determine if the testee is suffering from behaviour or character disorder; if the testee is depressed; and the like. Such tests are called diagnostic tests.

Ans to Q No 5: Two important functions of psychological test are: i) classification: psychological tests supply data for classifying individuals on the basis of various traits such as interest, aptitude, attitude, intelligence, or any other trait which is relevant for the purpose; and

(ii) selection: psychological tests help in making selection of individuals for admission, recruitment, or other activities on the basis of required knowledge or skills for the course or career.

Ans to Q No 6: A good test is one that is valid, reliable, objective, practical, and friendly to the test-taker. Good tests are administered under uniform condition and scored objectively in order to make the scores obtained by individuals who take the test comparable.

Ans to Q No 7: Validity is the most important technical criteria of a good test or other measuring instrument. Validity refers to truthfulness of a test. In simple language, validity of a test may be defined as the extent to which a test measures what it claims to measure.

Ans to Q No 8: Any two practical criteria of a good test are: (i) Clarity of purpose of the test, and (ii) Comprehensiveness or broad in scope.

2.13 MODEL QUESTIONS

A. Short Questions (Answer each question in around 150 words)

Q 1: Define psychological test.

Q 2: Write briefly about the uses of psychological test in education.

Q 3: Distinguish between test, examination, and evaluation.

Q 4: What do you mean by a standardized test?

Q 5: What do you mean by usability of a test?

Q 6: What is the meaning of objectivity?

B. Long Questions. (Answer each question in around 300-500 words)

Q 1: Trace the history of psychological testing.

Q 2: Discuss the nature of psychological test.

Q 3: Make a classification of psychological tests.

Q 4: Discuss the functions of psychological test

Q 5: Write about the different uses of psychological test.

Q 6: Discuss the practical criteria of a good test.

Q 7: Define validity. What are the different types of validity?

Q 8: Discuss the different methods of estimating reliability of a test.

Q 9: What is a norm? Write about the different types of norm.

*** ***** ***

UNIT 3: EVALUATION IN EDUCATION

UNIT STRUCTURE

3.1 Learning Objectives

3.2 Introduction

3.3 Meaning and Definition of Evaluation

3.3.1 Measurement, Assessment and Evaluation

3.4 Nature of Educational Evaluation

3.5 Relation between Measurement and Evaluation

3.6 Principles of Evaluation

3.7 Steps in Evaluation

3.8 Importance of Evaluation in Education

3.9 Tools and Techniques of Evaluation

3.10 Types of Evaluation

3.11 Let Us Sum Up

3.12 Further Reading

3.13 Answer to Check Your Progress

3.14 Model Questions

3.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- understand the meaning and types of Evaluation
- distinguish between Measurement and Evaluation
- understand the nature and principles of Educational Evaluation
- explain the steps and tools of Evaluation
- explain the significance of Evaluation in Education

3.2 INTRODUCTION

Evaluation plays a very important role in all areas of life. It helps people to assess and determine what is good and what is bad. As such Evaluation forms an integral and indispensable part of the process of education. It is seen that Education, as a process, is considered complete

only when it is accompanied by evaluation. Evaluation is an appraisal of changes in student behaviour as an outcome of learning. Such information is vital in education, because neither students nor teachers can progress effectively without adequate information about their progress towards goals.It is a continuous and comprehensive process of assessing student's progress towards the goals and objectives. The previous practice of holding examination at the end of academic year has been proved to be ineffective as well as defective. It has been widely realized and recognized that students' assessment should continue with the process of learning, which is a continuous process. Moreover, education being considered as a process of all round development; assessment of such development should include cognitive, affective, as well as psychomotor growth or change in student's behaviour. The modern concept and means of evaluation is also a part of the current trend in education to reform the faulty examination system. Evaluation involves collecting information, interpreting the information and making judgments about what is to be done in order to improve a child's behaviour. In this unit, you will be able to achieve a clear picture about the concept, importance, principles and types of evaluation in education.

3.3 MEANING AND DEFINITION OF EVALUATION

Evaluation is the process of supplying information for making decisions. Decision making or the act of selecting one out of many alternatives is based on making judgment or assigning values to alternatives. Such act of value judgment, again, is based on certain measurements, rules, objectives, or other criteria. The act of evaluation thus includes the tasks of collecting, organizing, analyzing, and processing information so that value judgment could be formed for making decisions.

The concept of evaluation in education is relatively new. Evaluation in education is a process of making judgement about the worth or value of the educative process and its outcome with reference to the objectives. It involves making judgments about the educational plans including goals and objectives; curriculum, methods, and personnel; decision about terminating,

continuing, or modifying various programs.Evaluation is also a process of making critical judgment about the extent to which change in behaviour of students has affected in the light of the set goals. However, it is not just an assessment or measurement of knowledge in a subject or subjects. It is, instead, a comprehensive assessment of students' motivation, aptitude, attitude, interest, values, skills, ideals, goals, and other traits of personality. So it can be said that evaluation is the assessment of physical, mental, characteristic, temperamental, behavioural, or the dynamic changes of the total personality of the individual.

The National Council of Educational Research and Training (NCERT), in its booklet,"The concept of evaluation in education" (1963), describes evaluation as the process of determining three important aspects:-

- The extent to which an objective is being attained,
- The effectiveness of the learning experiences provided in the classroom, and
- How well the goals of education have been accomplished.

Thus, three important elements are closely interwoven in the process of evaluation; they are: aims and objectives of education, learning experiences provided to attain the goals and objectives, and change in student behaviour as a consequence of the learning experiences provided. Evaluation is a process of analyzing the interrelation betweenthese three elements which can be diagrammatically presented as under:

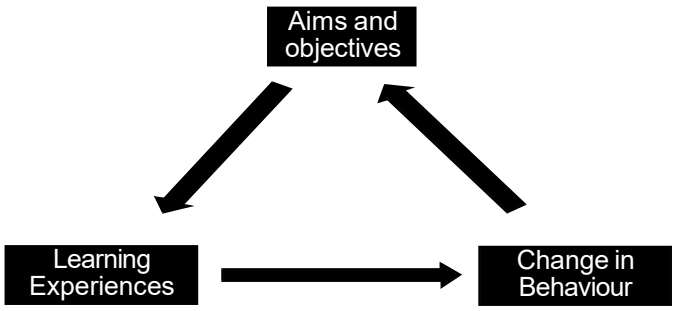


Figure 3. 1: The process of evaluation

The diagram above shows clearly that evaluation is always carried out with reference to the aims and objectives. Hence, educational evaluation or the process of making judgment about students' progress must refer to

the predetermined goals of education. Highlighting this fact, Ralph W. Tyler defined evaluation as, "the process of determining to what extent educational objectives are actually being realized by the programme of curriculum and instruction." In order to make the concept of evaluation more clearly, some important definitions have been presented below.

According to Bradfield and Murdock, "Evaluation is the assignment of symbols to phenomenon in order to characterize the worth or value of a phenomenon usually with reference to some social, cultural or scientific standard."

Cronbach defines evaluation as, "the collection and use of information to make decisions about an educational program."

In the views of James M. Lee, "Evaluation is the appraisal of pupil's progress in attaining the educational goals set by the school, the class and himself. The chief purpose of evaluation is to guide and further the students learning."

Patton explained evaluation as the "systematic collection of information about the activities, characteristics, and results of programs to make judgments about the program, improve or further develop program effectiveness, inform decisions about future programming, and/or increase understanding."

Thorpe (1980) described educational evaluation as "Evaluation is the collection, analysis and interpretation of information about any aspect of a programme of education, as a part of a recognized process of judging its effectiveness, its efficiency and any other outcomes it may have".

According to Bloom, "Evaluation is concerned with securing evidence on the attainment of specific objectives of instruction."

The definitions quoted above reveal that evaluation is a very comprehensive process. It helps to build an educational program, assess its effectiveness, and helps to improve it. Evaluation is a much wider concept than traditional examination which takes into account the cognitive, affective, and psycho-motor change in behaviour or learning outcomes. It is important to note at this point that there are two different situations in the context of educational evaluation; they are:-i) Evaluation In an Educational Program (EIEP), and (ii) Evaluation Of an Educational Program (EOEP).



CHECK YOUR PROGRESS

Q 1: Define evaluation.

.....

3.3.1 Measurement, Assessment and Evaluation

It is important to distinguish Evaluation from other similar concepts such as Measurement and Assessment. Though these terms are used interchangeably at times, they convey different meanings.

Measurement is the act of assigning a quantitative meaning to a phenomenon or object or criterion) i.e. the length of the table, marks in a subject, etc. It is a process of assigning numbers to objects, or events to give quantitative meaning to a person or object. In educational scenario, quantitative measures help to determine a child's performance e. g. a student's score in Science is 40. This measure can provide no further interpretation or it cannot be said if he has passed or failed. Measurement stops at assigning the quantitative score or number but does not make any value judgement on the score or performance.

Assessment is a fact finding process aimed at describing the conditions that exists at a particular time. Assessment uses measurement to collect information. Simultaneously it organises the measurement data into interpretable forms. In education, assessment describes the progress of students towards an educational goal at a point in time. However, assessment does not provide any explanation or the underlying reasons for such progress. It neither offers any recommendations for further action or improvement. In education, assessment is the process of describing the nature and the extent of pupils' learning in reference to the objectives of instruction.

Evaluation means adding value judgement to assessment. It makes judgement on the effectiveness, utility or desirability of a product, process or progress with reference to pre-defined and agreed upon objectives or values. Evaluation also includes recommendations for further improvement of the process or product. Thus, evaluation is a qualitative measure of a phenomenon. It judges the effectiveness, suitability, or goodness of the process or program being evaluated in order to reach meaningful decisions about that process or program.



LET US KNOW

Assessment is a fact finding process aimed at describing the conditions that exists at a particular time. Assessment uses measurement to collect information. Again, evaluation means adding value judgement to assessment.

3.4 NATURE OF EDUCATIONAL EVALUATION

Evaluation is the systematic process of making judgement about the adequacy, effectiveness, value, or worth of a phenomenon according to some criteria and purposes. Such judgment is based on the collection of relevant data and their comparison with the criteria and/or their effectiveness in terms of fulfilling the purposes. Valid or successful evaluationis characterized by exact and specific definition of what is to be evaluated, the goals of evaluation, specific standards or criteria against which evaluation is to be made, correct observations and measurements, and valid and logical conclusions. Following characteristics reveal the nature of evaluation.

- The concept of evaluation, which is relatively new in the field of education, is more scientific and comprehensive than examination or measurement.
- Evaluation is quantitative as well as qualitative measure of a phenomenon. It is the act of adding value judgement to an assessment or measurement.

- Although evaluation depends on measurement, evaluation and measurement are not the same; measurement is only a part of evaluation. Measurement assigns quantitative score and evaluation makes value judgement on that score.
- Evaluation is a process of making judgement about the value or worth of something, and hence it is subjective.
- Evaluation is based on some social, cultural, or scientific standards and/or goals.
- Educational evaluation is closely associated with the goals and objectives of education, i.e. educational evaluation is based on educational objectives.
- Evaluation is a continuous and systematic process which takes into account all the three domains of students' behaviour: cognitive, affective, and conative or psycho-motor.
- Evaluation is not just a process of assessing knowledge in certain subjects; rather it is an elaborate assessment of students' motivation, aptitude, attitude, interest, values, skills, ideals, goals, and other psycho-physical traits of personality. In fact, evaluation is the assessment of physical, mental, characteristic, temperamental, behavioural, or the dynamic changes of the total personality of the individual.
- Evaluation is a continuous process unlike tests or examinations that are held periodically or at the end of academic session. Since evaluation is continuous, no value or measure should be considered final or absolute.
- Three vital elements are closely interwoven in the process of educational evaluation; they are: aims and objectives of education, learning experiences provided to attain the goals and objectives, and change in student behaviour as a consequence of the learning experiences provided.
- The basic nature of evaluation is philosophical and normative.

3.5 RELATION BETWEEN MEASUREMENT AND EVALUATION

The concepts of Measurement and Evaluation are often confused to be the same and hence the terms are used interchangeably. Measurement and Evaluation, though similar, have different meanings. The concept of evaluation is much wider and more scientific than measurement. Measurement is only a part of evaluation. However, evaluation is based on measurement. Measurement is a numerical score of an object, subject, or some other characteristic, but evaluation is the act of adding subjective judgment or value to that score or measurement. For example, determining the length and breadth of a piece of cloth is an act of measurement; whereas, assessing the usefulness or adequacy of that piece of cloth to stitch a dress of definite size is an act of evaluation. Similarly, if in a test of Mathematics a student scores 45, this is a measurement of his knowledge in Mathematics. Now, if his score is compared with other students appearing the same test and he is given a rank or grade, then this is evaluation.

The definitions of measurement and evaluation as given by Bradfield and Murdock help in making the distinction between measurement and evaluation clear. According to them, "Measurement is the process of assigning symbols to dimensions in order to characterize the status of a phenomenon as precisely as possible" whereas, "Evaluation is the assignment of symbols to phenomenon in order to characterize the worth or value of a phenomenon usually with reference to some social, cultural or scientific standard."

Similarly, Munroe emphasizing another aspect of difference between measurement and evaluation wrote, "In measurement the emphasis is upon single aspect of subject-matter, achievement of specific skills and abilities, whereas in evaluation the emphasis is upon broad personality changes and major objectives of educational programme."

A comparative analysis of the features of measurement and evaluation may help us understand the difference between them:

TABLE 3.1: DIFFERENCE BETWEEN MEASUREMENT AND EVALUATION

Measurement	Evaluation
Measurement is the act of assigning numerical value to things and ordering them.	Evaluation is the act of making subjective value judgement about things on the basis of information obtained by measurement.
Measurement is objective, definite, and impersonal.	Evaluation is subjective, relatively indefinite, and personally meaningful.
Measurement is the quantification of objects or their characteristics.	Evaluation is a qualitative value judgement of the quantity.
Measurement determines the presence, absence, or intensity of some variable being measured.	Evaluation gives normative value and interpretation to measures besides determining the presence or absence of a variable.
Measurement is quantitative and numerical.	Evaluation is qualitative and normative value.
Measurement supplies information required for making evaluation.	Evaluation uses information supplied by measurement.
Measurement is a discrete process and its scope is limited.	Evaluation is an integrated process and an integral part of education. Its scope is comprehensive.
Measurement is concerned with immediate objectives and hence it is present oriented.	Evaluation is associated with long term goals of education; hence it is a continuous process.
Measurement is scientific in nature.	Evaluation is philosophic in nature.

The analysis presented above help us to understand the similarity-dissimilarity, and the relation between measurement and evaluation. Measurement and evaluation are two closely related concepts that are interrelated and interdependent.

3.6 PRINCIPLES OF EVALUATION

The process of evaluation must be based on certain important principles. These are:

- The first and foremost principle of evaluation is the determination or consideration of aims and objectives of evaluation. The evaluator must be very clear about the objectives of evaluation before conducting evaluation.
- The tools for conducting evaluation must be selected in accordance with the aims and objectives of evaluation. Because all tools may not be effective with the particular objectives of evaluation.
- Different tools and methods must be used for evaluation at its different stages since any single tool cannot make comprehensive evaluation of an individual or object.
- The evaluator must be well versed about the strengths and weaknesses of the various tools used in evaluation. Scrutinizing the validity and reliability of the tools is important in this regard.
- Since evaluation is a continuous process, no single result of evaluation should be considered to be final or absolute.
- Instead of conducting evaluation only for the sake of evaluation, it should be considered as a means of acquiring higher goals.
- The evaluator must be aware about the moral and ethical issues involved in the process of evaluation.



CHECK YOUR PROGRESS

Q 2: State two principles of evaluation.

.....

3.7 STEPS IN EVALUATION

A systematic programme of evaluation must follow the following steps:

- **Determining the objectives of evaluation:** Deciding what to evaluate is the first important step in the evaluation process. Evaluation starts with formulating and stating the objectives of evaluation clearly.

- **Defining the objectives in behavioural terms:** The second important step in educational evaluation is defining the objectives in behavioural terms. This is important because broad or vaguely stated objectives cannot be measured and evaluated. For example, if we want to evaluate "study skills" then it needs to be defined by listing what the students are expected to acquire. Without such behavioural definition of objectives evaluation cannot proceed successfully.
- **Selecting and fixing criteria or standards for evaluation:** Evaluation is always done by comparing observations or data with predetermined criteria or standards. Hence, it is important at this stage to select and fix such criteria or standard for making evaluation.
- **Selection and development of tools and techniques for data collection:** Selecting appropriate tools and techniques for collecting relevant data is the next important step in evaluation. The evaluator may have to develop his own tool and technique if available tools do not fit into his purpose of evaluation.
- **Collecting and recording data:** In this step collection of data using selected tools and their recording in appropriate form is done.
- **Analysis and interpretation of data:** After collection of data and their recording, those are analysed using various statistical and other proceduresso that the data could be interpreted and valid conclusions can be arrived at. The collected data are then compared with the previously determined standards or criteria for evaluation.
- **Evaluation and decision making:** In the final step of evaluation, decision about the value of the product or process being evaluated is taken and judgment is made about their adequacy or effectiveness on the basis of analyses, interpretation and comparison with the standards, criterion, and/or goals.

3.8 IMPORTANCE OF EVALUATION IN EDUCATION

Evaluation is an integral part of teaching-learning process. It is as important as objectives and curriculum of teaching and learning. Evaluation

provides with the most important feedback concerning suitability of objectives and curricula, extent to which objectives have been obtained, effectiveness of teaching and the methods, and other important elements of the process of teaching and learning. The report of the Kothari Commission (1964-66) realizing the importance of evaluation observed, "The new approach to evaluation will attempt to improve the written examination so that it becomes valid and reliable measure of educational achievement and to devise techniques for measuring those important aspects of the student's growth that cannot be measured by written examinations."

Some important points about the importance of evaluation may be stated as under:

- Evaluation appraises the success and failure of education. It assists in the improvement of education system. Evaluation helps in knowing the extent of learner's progress towards goals and the success and failure of the methods and learning experiences in achieving the goals.
- Evaluation analyses and reviews the goals and objectives of education. Evaluation starts with clarifying the goals and objectives of education and their statement.
- Traditional examination system is defective and fails to make proper evaluation of students. Evaluation makes continuous and comprehensive assessment of learners which is more scientific.
- Evaluation is important in view of its role in discovering individual differences among students. Such discovery of differences among students helps to shape objectives, curricula, and methods to suit different categories of students.
- Evaluation helps in promotion, placement in various courses and career, and grading of students. It helps modifying learning experiences or curriculum in order to attain the objectives of education.
- Evaluation makes comprehensive appraisal of desired changes in student behaviour and provides valuable insights about the needs,

possibilities, strengths, and weaknesses of students. Such information about students makes instruction effective and useful.

- Learning experiences or the curriculum is the means to attain the objectives of education. Through the study of different subjects and activities learner's behaviour takes shape along desirable lines. If the learning experiences fail to bring desirable changes in behaviour then it should be understood that the selection of subjects and activities have been defective. Thus evaluation helps in making necessary modification in the learning experiences or the curriculum.
- Evaluation provides a basis for guidance. Evaluation provides a total picture of the individual which helps to guide him in educational and vocational field. Thus, success of educational, vocational and personal guidance depends on proper evaluation of students' strengths, weakness and other personality characteristics.



CHECK YOUR PROGRESS

Q 3: Why evaluation is important in education?

.....

3.9 TOOLS AND TECHNIQUES OF EVALUATION

Successful evaluation is based on accurate and complete data about individuals and their characteristics. Data or information may be collected from a variety of direct or indirect sources. However, it is of vital importance to adopt systematic techniques to collect required data that are usable, valid and reliable. Various instruments, devices, or other means are used to collect different types of data for evaluation. Such instruments used for collecting data are called tools. Different tools are required for collecting different types of data for various purposes. The evaluator may need to use more than one tool in combination to meet his specific purpose. A technique is a practical method or procedure applied to some particular

task.Techniques of evaluation consist of tools used to collect data concerning particular outcomes of learning. Examples of evaluation techniques include anecdotal report, interview, case study, observation etc. whereas tools of evaluation include tests, question papers, observation schedules, questionnaire, rating scale, checklists, cumulative record etc.

Some important tools and techniques of evaluation may be presented as under:

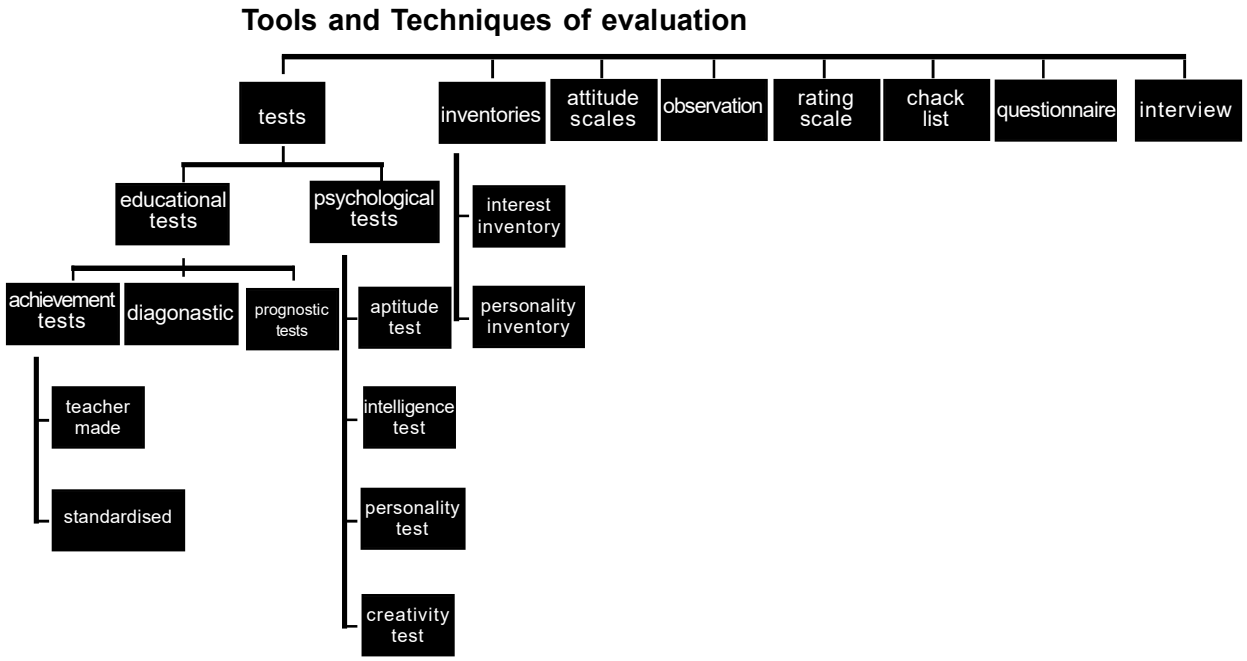


Figure3. 2: Tools and technique of evaluation

Tests: Tests are widely used tools in educational and psychological evaluation. A test is a collection of some questions to be answered or tasks to be performed by the testee to measure educational achievement or psychological traits of an individual or a group. Different tests can be administered individually or in groups. Educational tests again are classified as achievement, diagnostic, and prognostic tests.

An achievement test measures the knowledge or skills in an area or subject attained by individuals after a specific period of instruction or training. Achievement tests may be oral, written, and performance based. These tests may again be constructed by the classroom teacher himself, for his immediate and local needs (teacher-made test); or may be

constructed by test-specialists or experts following standard procedure (standardized test) .(For more details about teacher made and standardized test, refer UNIT-viii)

Tests that are used to diagnose the strengths and weaknesses of students in particular subjects are known as diagnostic tests. For example, tests may be used to determine if a learner is facing difficulty in Maths, or in History, and if 'yes' then why? Such tests are called diagnostic tests. Diagnostic tests assess students' understanding of a subject, area or skill. Prognostic tests are used to make predictions. Prognostic tests help to identify or categorize people with special talents or qualities in the future. In education such test predicts how well a person is likely to do in a certain school subject or task.

Psychological tests are used to measure various psychological characteristics or traits. Tests for measuring aptitude, intelligence, personality, creativity etc. are important among psychological tests. A test of aptitude attempts to predict the capacities or the degree of achievement that may be expected from an individual in a particular activity. Tests of creativity are used to measures creative potentialities of people. In addition to the tests of intelligence and personality, psychological tests help to measure the ability of individuals to perform tasks in various fields such as engineering, medical, musical, reading, clerical, spatial, numerical, dextrous, and many others.

Inventories: Inventories are instruments that attempt to assess one or more aspects of an individual's behavior. Inventories contain certain questions or statements concerning some personal characteristics, liking, disliking, interest, hobby etc which are written in the first person. Testee require to respond to such questions or statements in terms of 'true' or 'false', 'agree' or 'disagree'. Inventories are self-reporting tools. Depending upon the objectives, inventories are classified as interest, adjustment, and personality inventories. Examples are: Strong interest inventory, Bell Adjustment inventory, Minnesota Multiphasic Personality inventory etc. Inventories are economic, simple, and objective tools. However, one has to be careful in using inventory data as data obtained may be biased or

superficial. Another problem with inventory is that it is very difficult to validate.

Attitude scales: An attitude scale is a tool used to measure attitude of people towards some issue, object or subject. An attitude is a learned emotional response set for or against some object, individual, issue or other phenomenon. It is a mental or neural readiness of an individual to respond or to act either in positive or in negative way towards something. Attitude scale provides a basis for assigning a numerical value to a person's attitude. It is a tool designed to produce scores indicating the intensity and direction (positive or negative) of a person's feelings about an object or event. Attitude scales consist of statements and the respondent is asked to respond to the statements and on the basis of his responses a score is allotted for each item which indicates his position. The most popular attitude scales are the Thurston's scale, Likert's scale, and Guttman scale.

A **Thurston's scale** (also known as Thurstone technique of scaled values) consists of items in the form of statements which the respondent has to agree or disagree. Items with which the respondent agrees are scored. Each item has a predetermined value and the respondent's score on the scale forms the median score of the items with which he agrees. The item scores are determined by asking a number of judges to rank each item on the scale using an eleven-point scale reflecting the attitude that is being measured. The final score of each item is the median of the judges' individual scores.

Likert scales (also known as the method of summated ratings) consist of statements followed by five or seven responses with numerical scale values. Respondents have to indicate the extent of their agreement with each item by choosing any one alternative. The sum of the scores for each item is their score on the scale.

The **Guttman attitude scale** consists of a set of hierarchical statements relating to the attitude being measured. These statements reflect an increasing intensity of attitude. The point at which the respondent disagrees with a statement reflects the respondent's scale position.

Another popular way of attitude scaling is the semantic differential method, developed by Osgood, Suci, and Tannenbaum. In this scale

respondents rate the concepts on a bipolar scale (usually seven-point). The two ends of the scale are defined by pairs of adjectives with supposedly opposite meanings (e.g. good/bad, wright/wrong, etc.)

Observation: Observation is a very popular and useful technique of data collection for evaluation. It is the act of watching the behaviour of an individual or a group, events, or noting down physical characteristics in natural or actual setting. The results of observations are recorded using different means and are preserved. Observation helps to measure and evaluate external or overt behaviour of individual. Observation can be natural or controlled; observation is natural when it is conducted in natural situations without manipulating the environment and without the knowledge of the subjects being observed, and observation is controlled when the environment is controlled by the observer and the subjects are informed that they are being observed. Natural observation is advantageous because of the fact that people behave naturally if they do not know that they are being observed. However, natural observation is not always possible or feasible and hence controlled observations must also be used.

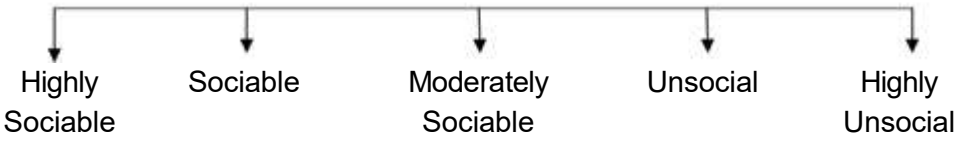
Observations can also be participant or non-participant. In participant observation, the observer becomes one among the people whose behaviour is observed and he takes part in their activities while observing them. In non-participant observation, the observation is conducted from a distance without interfering in the natural setting and activities of the individuals being observed.

Observation must be systematic and carefully planned. The main advantage of observation technique is that it allows to collect data where and when an event or activity is occurring. However, observation is susceptible to observer bias. Moreover, people try to perform better when they know they are being observed.

Rating scale: Rating scales consist of some statements expressing opinion or judgment concerning some situation, object, or character on a scale of values. It consists of some points indicating the observable trait in its different degrees. The rater is asked to rate his choice on a 3, 5, or 7 point scale.

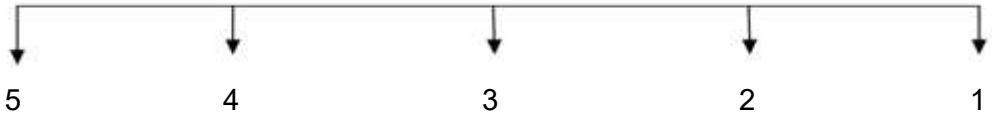
Though there are different forms or types of rating scales, the Numeric and Graphic rating scales are most commonly used. An example of an item in aGraphic rating scale is given below:

Q. How sociable is Madhu?



The same item in a Numeric rating scale would be as under:

Q. How sociable is Madhu?



Though rating scales are very popular and widely used tools of evaluation, they have several limitations. These limitations include: Halo effect, or the tendency to rate on the basis of previous performance or general impression about the person to be rated;Generosity error or the tendency of raters'to be too generous while rating; Leniency error or the tendency to assign higher rate to people known to the rater; on the contrary sometimes awareness about this error makes raters rate lower than they should actually do; Central tendency error or the tendency to rate around the average; and other such errors.

Check list: Check list is a simple tool of evaluation. It contains a list of behaviours or questions and the respondent needs to check 'yes' or 'no' against the question or behaviour. It is a very useful tool for identifying the presence or absence of knowledge, skills, or behaviours.

Questionnaire: A questionnaire is a set of questions concerning some social, psychological or educational topic or issue. A questionnaire may either be administered face to face or mailed. Collecting information about the object or subject to be evaluated is the purpose of using questionnaire. Questionnaire may be administered individually or in groups.

Questionnaires may be classified into two types: closed and open. A questionnaire requiring the respondent to answer in short or to

select one from the given alternatives is known as a closed type of questionnaire. Closed type of questionnaires are very useful for collecting certain special types of data. Closed questionnaires are relatively objective, scoring is easy and fast, and answering to such questionnaire is easy for the respondents.Open questionnaires require long and open answers from the respondents in their own language. Such questionnaires can delve deep into the issue and provide freedom to the respondents to answer in any length and way. However, analysing and interpreting the open answers is a difficult task.

Chief advantages of questionnaire include their economy, ability to cover wide sample, freedom to respondents, etc. However, questionnaire is very time consuming tool and there is no check if respondents misunderstand or omit some questions. Moreover, the rate of return of mailed questionnaires is very low.

Interview: Interview is a means of face to face and direct communication for collecting required data. Advantage of interview over questionnaire is that the problem of misunderstanding or omitting any question by the respondent can be avoided. Once the interviewer establishes rapport with interviewee and creates conduciveenvironment for interview it becomes easy to collect all important data from the interviewee. Another big advantage of interview is that body language of the respondent supplies additional information about the subject.

Interviews may be classified into structured and unstructured types. A structured interview follows predetermined and standardized steps. Same questions are asked to all the interviewees in the same predetermined sequence and the interviewer cannot change, add, or delete any question. Unstructured interviews are flexible and the questions or their sequence may be changed by the interviewer according to the requirement. He may add or delete questions if needed. Such flexibility helps in establishing proper rapport with the interviewee which helps in collecting required information. Unstructured interview is variously known as 'focussed', 'depth', or 'non-directive' interview.

Interview is a very important tool characterized by flexibility and high response rate which can elicit deep and sensitive information. However, interview is costly, time consuming, invades privacy of the respondents, and suffers from inhibitions due to face to face presence of interviewer. Moreover, information obtained from interview is difficult to interpret.

Educational tests: Most important tools of educational evaluation are educational tests. Educational tests may be classified as under:

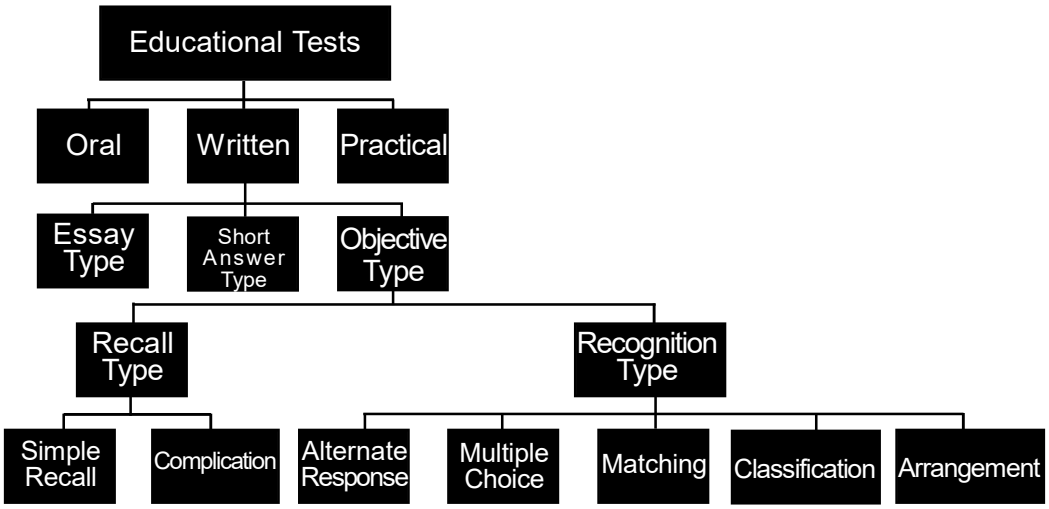


Figure 3.3: Classification of Educational Test



CHECK YOUR PROGRESS

Q 4: What is likert scale?

.....

Q 5: What is a questionnaire?

.....

3.10 TYPES OF EVALUATION

The process of evaluation has been classified into different types on the basis of its goals, functions, and on the basis of its reference points. A brief discussion about the important types has been presented below:

Formative evaluation: This type of evaluation proceeds along the process of teaching and learning. Formative evaluation helps in providing necessary feedback to students and teachers while teaching learning continues. Such feedback about the ongoing process helps to adjust teaching and learning to improve achievement. Formative evaluation provides teachers and students with continuous feedback. Examples of formative evaluation are: unit tests, weekly or monthly tests, etc.

Summative evaluation: Summative evaluation is terminal in nature. This type of evaluation is done at the end of academic session or year to assess the extent of progress made towards the objectives of instruction. This evaluation helps in making decisions about promoting or retaining students, grading and/or certifying them.

- A brief comparison between formative and summative evaluation:
- Formative evaluation considers evaluation as a process; summative evaluation considers it as a product.
- Formative evaluation is done simultaneously with the process of teaching and learning; summative evaluation is done at the end of teaching and learning.
- Formative evaluation gives immediate feedback to teacher and students; summative evaluation does not give immediate feedback.
- Formative evaluation helps in necessary revision of objectives, methods or curriculum while the process of teaching and learning continues; summative evaluation assesses the adequacy of objectives and effectiveness of learning at the end of teaching learning process.
- Formative evaluation is generally criterion based and continuous; summative evaluation is norm based and terminal.

Evaluation may also be classified on the basis of their determinant (with which the scores or results of evaluation are compared). When an individual is evaluated by comparing his performance with the performance of the group to which he belongs, then that is called **norm-referenced evaluation**. Norm-referenced evaluation gives information about the relative standing of individual in a group. In criterion-referenced evaluation,

performance of each individual is compared to a pre-determined set of criteria or a standard rather to the other members of the group. Criterion-referenced evaluation assesses if an individual has attained mastery over a certain skill or skills. This evaluation helps in making decision about "pass" or "fail" and certification.

Evaluation may also be classified on the basis of two different levels: **Programme level evaluation** i.e. evaluation of a programmewhether the implementation of the programme has been successful or not; and **Student Level evaluation** i.e. evaluation of a student about his performance in a programme of study.

Evaluation has also been classified on the basis of their function. When evaluation is used to diagnose specific area of weakness of students or their understanding of a subject, area or skill, such evaluation is known as diagnostic evaluation.Similarly when students are evaluated with the purpose of placing them in the proper course or career, evaluation is known as placement evaluation.



3.11 LET US SUM UP

After going through the units, we learnt the following:

- The relatively new concept of evaluation in education is a process of making judgement about the worth or value of the educative process and its outcome with reference to the objectives.
- Measurement and evaluation, though similar, have different meanings. Measurement is a numerical score, but evaluation is the act of adding subjective judgment or value to that score or measurement.
- Evaluation process must follow certain principles and systematic steps.
- For successful evaluation various data collection tools must be used. These tools include: tests, inventories, attitude scale, observation, rating scale, checklists, questionnaire, interview etc.

- Evaluation could be of different types including formative, summative, diagnostic, placement, norm-referenced or criterion referenced, etc.



3.12 FURTHER READING

1) Bloom, B.S. (1956): Taxonomy of Educational Objectives. Handbook I, Cognitive Domain, David McKay Co. Inc,New York
2) Gupta, S.P. (2009): Modern Measurement and Evaluations with Statistics, Sharda Pustak Bhavan, Allahabad
3) NCERT (1963): The Concept of Evaluations in Education, New Delhi.
4) Stanley & Hopkins (1966): Education and Psychological Measurement and Evaluation, Prentice Hall Pvt. Ltd. Eaglewood Cliff, New Jersey
5) UGC (1972): Examination Reforms: A Plan of Actions, New Delhi



3.13 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Evaluation is a qualitative measure of a phenomenon. It judges the effectiveness, suitability, or goodness of the process or program being evaluated in order to reach meaningful decisions about that process or program.

Ans to Q No 2:

- The first and foremost principle of evaluation is the determination or consideration of aims and objectives of evaluation. The evaluator must be very clear about the objectives of evaluation before conducting evaluation.
- The tools for conducting evaluation must be selected in accordance with the aims and objectives of evaluation. Because all tools may not be effective with the particular objectives of evaluation.

Ans to Q No 3: Evaluation forms an integral and indispensable part of education. Evaluation provides important feedback concerning suitability of objectives and curricula, extent to which objectives have been obtained, effectiveness of teaching and the methods, and other important elements of teaching-learning process. Such information is vital in education, because neither students nor teachers can progress

effectively without adequate information about their progress towards goals.

Ans to Q No 4: Likert scale is used to measure attitude. It consists of statements followed by five or seven responses with numerical scale values. Respondents have to indicate the extent of their agreement with each item by choosing any one alternative. The sum of the scores for each item is their score on the scale.

Ans to Q No 5: A questionnaire is a set of questions given to respondents in order to collect data concerning some social, psychological or educational topic or issue.



3.14 MODEL QUESTIONS

A. Short Questions (Answer each question in around 150 words)

Q 1: What is the difference between measurement and evaluation?

Q 2: What is halo effect?

Q 3: Distinguish questionnaire and inventory.

B. Long Questions. (Answer each question in around 300-500 words)

Q 1: Discuss the nature of educational evaluation.

Q 2: Write briefly about the principles of evaluation.

Q 3: Discuss the steps involved in evaluation.

Q 4: Write about the different types of evaluation.

*** ***** ***

UNIT 4: TEST CONSTRUCTION AND STANDARDIZATION

UNIT STRUCTURE

4.1 Learning Objectives

4.2 Introduction

4.3 Meaning of Psychological Test

4.4 Meaning and characteristics of a Standardized Test

4.5 Principles of Test Construction

4.6 Methods of Test construction and Standardization

4.7 Let Us Sum Up

4.8 Further Reading

4.9 Answer to Check Your Progress

4.10 Model Questions

4.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- define Psychological Test
- define Standardized Test
- highlight the principles of Test Construction
- explain the steps involved in Test Construction.

4.2 INTRODUCTION

Psychological Test is one of the most important tools used in measurement and evaluation. Psychological tests are very useful in the field of education and in various other fields of human activity. These tests are used to study intelligence, aptitude, achievement, and various other scholastic and non-scholastic traits of individual's personality. These tests help to determine the nature of individual differences among learners. Psychological tests are very useful in studying intellectual development as well as academic and non-academic personality characteristics that are associated with age and sex, personality characteristics of gifted, retarded,

or educationally backward, personality traits of the neurotics and psychotics. Educational institutions widely use psychological tests for the dual purpose of making various decisions and administrative policies and management of the institution. Psychological tests are also useful in providing educational and vocational guidance and counselling to students besides their use in employment, business, industry, and many other areas of work that affect the lives of many people. Since we have already discussed the history and concept of psychological test in detail in the previous unit, here we are going to focus on the meaning of standardized test, principle of methods of test construction.

4.3 MEANING OF PSYCHOLOGICAL TEST

A psychological test is a collection of some questions to be answered or tasks to be performed by the psychologist or teachers. Such collection of questions or tasks is used to measure intelligence, knowledge, skill, or other traits of personality of an individual or a group. This collection of questions or tasks (test) is required for the reason that intelligence, achievement, or other such trait or characteristic cannot be measured directly unlike the measurement of the objects of the physical world. Measurement of the objects of the physical world is direct and easy because such measurement can be done using universally acceptable standard tools and units. Conversely, psychological measurement that involves the measurement of knowledge, skill, aptitude, achievement, or any other trait of the personality cannot be measured until and unless they are manifest in the behaviour of man. Therefore, it is required to make the abstract trait or quality to be measured explicit in the behaviour of the individual so as to make it measureable, and hence some stimuli in the form of questions or tasks are given to the subjects or the person whose behaviour is being measured so that their responses could be observed, recorded and measured. The questions to be answered or the tasks to be performed are so selected or framed that the response of the testee to such question/task could make the trait to be measured explicit and measureable. Thus, a

psychological test may be understood as a set of carefully selected stimuli in the form of questions to be answered or tasks to be performed to elicit a sample of behaviour required to measure a particular psychological trait or characteristic. The definition given by psychologist Marshall is noteworthy in this context: "A psychological test may be defined as a pattern of stimuli, selected and organised to elicit responses which reveal certain characteristics in the person who takes them." However, this definition does not include all the features of a psychological test. Hence, some important definitions of psychological test are quoted below:

According to Frank S. Freeman, "A psychological test is a standardized instrument designed to measure objectively one or more aspects of a total personality by means of samples of verbal or non-verbal responses, or by means of other behaviours."

Psychologist Anastasi and Urbina have defined a psychological test as an objective and standardized measure of a sample of behaviour.

A.K.Singh stated that "A psychological or educational test is a standardized procedure to measure quantitatively or qualitatively one or more than one aspect of a trait by means of a sample of verbal or non-verbal behaviour.

In the words of McIntire and Miller, "Psychological tests are instruments that require the testee to perform some behaviour. The behaviour performed is used to measure some personal attribute, trait or characteristic - such as intelligence - that is thought to be important in describing and understanding the behaviour."

Analyses of the foregoing definitions of psychological test help us to conclude that a test consist of some items or tasks to be performed by the testee which elicit a sample behaviour of the testee. Such sample of behaviour indirectly reveals the personality trait or characteristic which is being measured.



CHECK YOUR PROGRESS

Q 1: Define psychological test.

.....

4.4 MEANING AND CHARACTERISTICS OF A STANDARDIZED TEST

A standardized test is a test that is administered, scored, and interpreted in the same way for all test-takers. A standardized test establishes uniformity of procedures for administration and scoring of the test. This uniformity of procedure is important to make the scores obtained by different persons comparable. According to L. J. Cronbach, a standardized test is one in which procedure, apparatus, and scoring has been fixed so that precisely the same test can be given at different times and places.

Following characteristics of a standardized test make its meaning more clear:

- A standardized test is administered under identical conditions, it does not matter where, when, by whom or to whom the test is administered, it is done in a similar way.
- A standardized test is always scored objectively and in a standard or consistent manner; the procedures for scoring a standardized test are specified so that different scorers arrive at the same scores for the same set of responses.
- A standardized test requires all test takers to answer the same questions in the same way.
- Scores obtained by different individuals or groups in standardized test can be compared in order to assess the relative performance of individual students or groups of students.
- Standardized tests are usually designed to measure relative performance of individuals or groups. Hence establishment of norms

for comparing the performance of individuals or groups forms an important part of a standardized test construction.

- The formulation of directions for a standardized test is a major part of its construction. Directions for standardized tests specify the materials to be used, time limits to be given, oral instructions and demonstrations to respondents, and all other details of the testing situation.
- Standardized tests are highly reliable and valid.

4.5 PRINCIPLES OF TEST CONSTRUCTION

Construction of psychological test and its standardization must be well planned and methodical. There are some fundamental principles that should be observed and guide the test-makers when constructing tests. Some of these important principles are as follows:

- Construction of psychological test must start with Proper Planning for the Test. While planning the test, the intended function of the test, the area, the specific objectives, type of test items and their effectiveness, cost and practicality must be considered.
- Another important principle of test construction involves the task of defining objectives in behavioural terms and in precise terms that will serve as a guide to testing.
- The content to be covered by the test should be clearly specified.
- A blueprint for test construction must be prepared. The blueprint should include the number of items to be included in the test under each topic of the content.
- The instruction to the candidate should be clear, unambiguous and precise.
- Tests must be constructed incorporating items that help in measuring all the instructional objectives.
- Use of test items must be appropriate for particular objectives. It must be seen that items can test higher mental abilities rather than simple recall of information.

- An important principle of test construction is to make the test reliable, valid and objective. A test is reliable when it gives consistent result, valid when it measures what it intends to measure, and objective when the score is free of influence of the person making the measurement.
- Items having answers with doubts, unimportant, meaningless and ambiguity must be avoided.
- Language of the items should be clear and straight forward. At the same time, the rules of rhetoric, grammar and punctuation should be strictly followed.
- Any item without clear cut answer must be avoided.
- Tricky questions or items should be avoided.
- Items which furnish the answers to other items must be avoided.



CHECK YOUR PROGRESS

Q 2: Write any two principles of test construction

.....

4.6 METHODS OF TEST CONSTRUCTION AND STANDARDIZATION

Test construction is a very complex and an extensive process. Various important factors must be considered when one wants to construct a psychological test. There are numerous points which are common to all types of tests which must be taken care of while constructing a test. Important among these factors or points are the area in which the test is to be constructed, the age group and the grade for which the test is to be made. Moreover, the general or the specific purpose to be served by the test must be considered. In addition to these points, certain important steps are to be followed while constructing a psychological test. Following are the important steps to be followed:

1. **Planning the Test:** The first important step in the test construction is planning of the test. In this step, the test constructor has to determine the objectives of testing besides considering other important things. An analysis of the curriculum or content relating to the topic on which the test is to be made, consideration of the target population for which the test is to be made, the limitations under which the test is to be constructed, the length and the nature of test items to be used, method of scoring and interpretation, etc are some important considerations that seek attention while planning for a test. Therefore, in the stage of planning the test maker has to follow the steps mentioned below:

- **Determination of the objectives of the test:** Psychological tests may be constructed with different purposes. The objectives of a psychological test may vary from the measurement of entry behaviour of the pupils, the effect of teaching on their performance during the course of their learning, to the measurement of their terminal behaviour. Moreover, psychological test may be made to measure different psychological attributes including intelligence, aptitude, interest, personality and the like. Therefore it is very important to decide with what purpose or objective the test is to be made at the very beginning of test construction. Only after determining the objectives clearly the next steps in test construction can proceed.
- **Preparation of the Blue-print:** The second important step to be taken by the test maker while planning his test is to prepare a blue print or the test-specifications. A test blueprint is a document that reflects the content of a test. It contains the instructional objectives, the questions or tasks to match the instructional objectives, and the learning domains and levels therein. Thus, a test blueprint is the plan that one creates and uses when building a test. The preparation of a blue print requires the following:
- **List all the topics** or units selected for testing and the time allotted for each topic or unit as shown in the following example:

Serial No	Topic/Unit	Period of instruction
1	The solar system	16
2	The water world	10
3	Food and nutrition	8
4	Means of communication	8
5	Our environment	12
	Total hours	54

- **Determine the weight-age or marks to different areas of content** by calculating the percentage to adjust marks for each topic, in the following manner:

Hour of instruction per topic ×100

Total hours

Serial No	Topic/Unit	Period of instruction	Calculation	Percentage
1	The solar system	16	16/54X100	29.6=30
2	The water world	10	10/54X100	18.5=18
3	Food and nutrition	8	8/54X100	14.8=15
4	Means of communication	8	8/54X100	14.8=15
5	Our environment	12	12/54X100	22.2=22
	Total hours	54		100

- **Determine the weight-age to different objectives** and the different types of items to be included, as shown in the following example:

Instructional objectives	Type of test items			Total
	Objective type	Short answer type	Essay type	
Knowledge	3			3
Understanding	10	7		17
Application	7	14	15	36
Analysis		7	15	22
Synthesis		7	15	22
Evaluation				
Total	20	35	45	100

- Preparation of the table of specification or **the blue print:**

Instructional Objectives / Topics	Knowledge			Understanding			Application			Total
	Very Short	Short	Essay	Very Short	Short	Essay	Very Short	Short	Essay	
The solar system	1	2		1	4			7	15	30
The water world	1			2					15	18
Food and nutrition	1			1	3		4	6		15
Means of communication	2	6		2	5					15
Our environment	1	2		3		15	1			22
Total	6	10		9	12	15	5	13	30	100

2. **Preparing the first draft of the test:** After planning for the test, preparing the first draft of the test on the basis of the blueprint is the second important step in test construction. In this step, item for the test instructions to the testee and the scoring key is prepared. Preparation of items for the test is the most important task that needs utmost care on the part of the test-maker. Following guidelines help in preparing items for a test:
   - Items must relate to the objectives of the test, i.e. items must measure the attribute or performance described in the specific learning outcomes or objectives.
   - Items must represent the whole content or area and all the instructional objectives.
   - Items should be clearly written and should not be ambiguous. Testee should easily understand the item.
   - The difficulty level of items should be of appropriate level.
   - Items must be grammatically correct and should not give any cue to testees.
   - Items without clear cut answer must not be included in the test.
   - Items containing answers to other items must be avoided.
   - Test items must be technically correct, i.e. reliable, valid and objective.

Another important aspect of test construction which goes hand in hand with the construction of test items is the preparation of instructions for the test. Technical qualities of a test like validity and reliability depend on the clarity of instructions to a great extent. Hence instructions for the test must clarify the purpose of the test, mention the time allotted for completing the test, provide clear directions to testees about how to answer the items, state how the answers to items would be recorded, and how to deal with the effect of guessing. At the same time the test maker has to prepare a scoring key and explain the method how to score the items.

Some test experts suggest that the preliminary draft of the test should be submitted to experts for their opinion and should be

administered to a small sample from the population and analyse their responses. This is called the small group try out of the test. The preliminary draft of the test may call for further modification in the light of experts' opinion and small group try out. Thus the preliminary draft is prepared for try out on a very large sample in the next step.

3. **Trying out the test:** Once the preliminary draft of the test is ready, it should be administered to a large and random representative sample of the population for which the test is constructed. The attempt to administer the test helps to determine the validity, reliability, and usability of the test. Moreover it helps to improve defective items and delete unnecessary ones. Trying out the test also helps in determining the difficulty and discriminating power of the items in the test. The sample on which the test is administered must be inclusive of poor, good and brilliant students. The physical and psychological condition should be normal. It must be seen that the testing condition is not affected by any unwanted factor. There should be proper invigilation and arrangement for sitting, lighting and other required conditions.

   After the administration of the test, the scoring of the test booklets should be done with the help of the scoring key and procedure prepared beforehand.

4. **Item analysis:** After the administration and scoring of the preliminary draft of the test, the task of analysing items starts. The main purpose of analysing items is to improve the quality of test items for future administration. The simple meaning of item analysis is determination of relative difficulty level and discriminating power of test items. Item analysis helps in discarding the items that are too easy or too difficult. Generally, an item is considered to be too easy when the item is attempted correctly by most the respondents and an item is too difficult when it is answered correctly by a few respondents only. Hence, only items with average or moderate difficulty level are retained for the final draft of the test. In addition to the task of determining the difficulty level of items, determining the

discriminatory power of the test items is another important task done by item analysis. Discriminating power of test items distinguish between good (students who know more) and bad (students who know less) students. If any item is answered correctly by an equal number of good as well as bad students then it should be understood that the item is not making any discrimination between good and bad students. On the other hand, if an item is answered correctly by more number of good students than bad students, then the item is discriminating between good and bad students.

The act of item analysis also includes the task of distracter analysis in case of multiple choice type items. A multiple choice type item presents some incorrect alternatives along with one correct alternative. The incorrect alternatives are called distracters. Distracter analysis is the process of determining if the alternatives could serve their purpose of distracting the respondents.



CHECK YOUR PROGRESS

Q 3: What is a Blueprint?

.....

Q 4: What is the main purpose of item analysis?

.....

5. **Preparing the final draft of the test:** Items selected on the basis of item analysis are put into the final draft of the test. Items with proper difficulty level and discriminating power are included in the final draft of the test. Clear instructions are given to the respondents for answering each item. On the basis of the administration of the preliminary draft of the test, time required for completing the test is determined and recorded.
6. **Standardization:** The last of test construction is standardization. The main purpose of this step is to determine the qualitative standard

and norms for the test. The main tasks involved in this step are as follows:

- Administering the final draft of test on a very large group of students,
- Scoring the test booklets and arranging the scores for analyses,
- Calculating Mean and other statistical values of the scores obtained by the testees,
- Determining the norm for the test on the basis of the statistical values of the scores,
- Determining the reliability of the test items,
- Determining the validity of the test items,
- Considering the various aspects of usability of the test,
- Preparing test manual or guidelines for administering the test.

Thus, the construction of a psychological test goes through the steps discussed above. (Further details about some of important steps like **item analysis** have been discussed in subsequent chapters.)



ACTIVITY 4.1

Prepare a blueprint taking any topic from any school subject.

.....



4.7 LET US SUM UP

After going through the unit, it can be summarized as-

- Psychological test is one of the most important tools used in measurement and evaluation.
- A psychological test is a collection of some questions to be answered or tasks to be performed by the testee. The questions to be answered or the tasks to be performed are so selected or framed that the response of the testee to such question/task could make the trait to be measured explicit and measureable. Thus, a psychological test may be understood as a set of carefully selected stimuli in the form of questions to be answered or tasks to be performed to elicit a

sample of behaviour required to measure a particular psychological trait or characteristic.

- A standardized psychological test is a test that is administered, scored, and interpreted in the same way for all test-takers.
- A standardized test is characterized by uniformity of procedures for administration and scoring of the test. Owing to its uniformity of procedure, the scores obtained by different persons on standardized test are comparable.
- Construction of psychological test and its standardization must be well planned and methodical. Some fundamental principles should invariably govern the construction of psychological tests. Every effort must be made to make the test valid, reliable, objective, and usable.
- Some important steps must be followed while constructing a psychological test. Important among these steps are:
  - Planning for the test
  - Preparing the first draft of the test
  - Trying out the test
  - Item analysis
  - Preparing the final draft of the test:
  - Standardization



4.8 FURTHER READING

1) Anastasi, Anne (1982): Psychological testing, Edition 5, illustrated. Macmillan,The University of Michigan.
2) Bloom, B.S. (1956): Taxonomy of Educational Objectives. Handbook I, Cognitive Domain, David McKay Co. Inc, New York.
3) Ebel, R.L. and Frisbel, D.A. (1979): Essential of Educational Measurement, 3rd Edition, Prentice Hall of India, New Delhi.
4) Gupta, S.P.(2009): Modern Measurement and Evaluations with Statistics, Sharda Pustak Bhavan, Allahabad
5) Sax, G. (1970): Principle of Educational Measurement and Evaluation, Wordsworth Pub. California.

6) Stanley & Hopkins (1966): Education and Psychological Measurement and Evaluation, Prentice Hall Pvt. Ltd. Eaglewood Cliff, New Jersey

4.9 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: A psychological test is a collection of some questions to be answered or tasks to be performed by the testee. The questions to be answered or the tasks to be performed are so selected or framed that the response of the testee to such question/task could make the trait to be measured explicit and measureable. Thus, a psychological test may be understood as a set of carefully selected stimuli in the form of questions to be answered or tasks to be performed to elicit a sample of behaviour required to measure a particular psychological trait or characteristic.

Ans to Q No 2: Two important principles of test construction are: (i) Construction of psychological test must start with Proper Planning for the Test. While planning the test, the intended function of the test, the area, the specific objectives, type of test items and their effectiveness, cost and practicality must be considered. (ii) Another important principle of test construction is to make the test reliable, valid and objective. A test is reliable when it gives consistent result, valid when it measures what it intends to measure and objective when the score is free of influence of the person making the measurement.

Ans to Q No 3: A test blueprint is a document that reflects the content of a test. It contains the instructional objectives, the questions or tasks to match the instructional objectives, and the learning domains and levels therein. Thus, a test blueprint is the plan that one creates and uses when building a test.

Ans to Q No 4: The main purpose of item analysis is to improve the quality of test items for future administration. The process of item analysis involves determination of relative difficulty level and discriminating power of test items.

4.10 MODEL QUESTIONS

A. Short Questions (Answer each question in around 150 words)

Q 1: What do you mean by a standardized test?

Q 2: What is small group try out?

Q 3: What do you mean by item analysis?

B. Long Questions. (Answer each question in around 300-500 words)

Q 1: Enumerate the main principles of test construction.

Q 2: Explain the process of preparing a blue print for construction of a psychological test.

Q 3: Discuss the steps involved in test construction.

Q 4: Write about the tasks involved in standardizing a test.

*** ***** ***

UNIT 5: TEST ITEMS

UNIT STRUCTURE

5.1 Learning Objectives

5.2 Introduction

5.3 Meaning of Item Analysis

5.3.1 Objectives or purposes of Item Analysis

5.4 Analysis and selection of Test Items

5.4.1 Determination of difficulty level

5.4.2 Determination of discriminating power

5.4.3 Distracter analysis

5.5 Preparation of different Types of Tests Items

5.5.1 Guidelines for construction of different types of Test Items

5.6 Let Us Sum Up

5.7 Further Reading

5.8 Answer to Check Your Progress

5.9 Model Questions

5.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- understand the meaning and purpose of item analysis
- understand the analysis and selection process of test items
- explain the process of determining difficulty and discrimination of test items
- explain the procedure of distracter analysis
- understand the process of preparing different types of test items.

5.2 INTRODUCTION

A psychological test is basically a collection of a set of test items. An item is a statement in the form of a question. E. Lindquist defined an item as, "A scoring unit in an exercise." According to K. L. Bean, an item is "a single task or question that usually cannot be broken down into any smaller

unit". Nunnally described an item as the "lowest common denominator of a test". Hence, the quality of a test depends on the quality of test items to a large extent. The vital importance of item analysis can be understood from the fact that most of the important criteria including validity, reliability, and objectivity of a test depend on the quality of its items. Item analysis can ensure high reliability, validity, and objectivity by evaluating items and suggesting necessary improvement in them. Improving a test, in fact, means selecting appropriate items and revising improper items through the process of item analysis. Item analysis suggests deletion of unnecessary items from a test to make it shorter and at the same time to increase its validity and reliability. This is why item analysis is one of the most important aspects of test construction. However, item analysis is a broad term for a set of different methods used to evaluate test items. In this unit, a detail description regarding item analysis is presented so that learners can prepare different test items on their own.

5.3 MEANING OF ITEM ANALYSIS

Item analysis is a very important process of standardized test construction. Item analysis evaluates the quality of test items. Quality of a test depends on the quality of items. Hence, one cannot hope to construct a good test without analysing items constructed for the test. Item analysis in its simplest meaning is the process of determining relative difficulty level and discriminating power of the test items. In other words, the task of item analysis is to determine how difficult or how easy an item was, and whether the item could differentiate between good and poor students. The validity and reliability of a test also depend on the validity and reliability of the items in the test.

According to **Abodunrin**, item analysis is "the process involved in examining or analysing testee's responses to each item on a test with a basic intent of judging the quality of item specifically, the difficulty and discriminating ability of the item as well as effectiveness of each alternative."

In the views of **Guilford**, "The major goal of item analysis are the

improvement of total score reliability or total score validity and the achievement of better item sequences and types of score distributions. Both the validity and reliability of a test depend on characteristics of its items. Item analysis ensures high reliability and validity into a test in advance."

According to **Arthur S. Reber**, "In a narrower sense, the term (item analysis) will be used specifically for a measurement of how effectively each individual item contributes to the overall validity of the test."

According to **Baron-Cohen, Wheelwright, et** al, "Item analysis is a process in which responses to test items are examined in order to test the quality of those items. Statistical methods are routinely used to identify any test items that do not belong on the test. Items are removed when they do not show the same kind of patterns of association with other items that is observed among retained test items. In addition, an item analysis will reveal if an item is too easy, too difficult, and/or will show a difference between different types of test-takers."

While constructing a test, one has to construct more items than the number of items actually required for the test. These initial items are administered on a large sample and the answers are carefully analysed. After analysing the items for difficulty, discrimination, validity, reliability, objectivity, etc., selected items are retained for the final draft of the test and the defective items are discarded. This process of analysing items can ensure high quality of a test. Hence, item analysis is an indispensable part of the process of constructing a good test.

5.3.1 Objectives or Purpose of Item Analysis

Item analysis allows the test constructor to exercise quality control over the test which is being constructed. Well defined objectives of learning and well-prepared items may help in good beginning in the process of test construction, but item analysis can give feedback on how successful the test actually was. The main objectives of item analysis may be listed as below:

- The first and foremost purpose of item analysis is to improve the qualitative standard of test items for using it in the future.

- One important purpose of item analysis is to help the process of construction of appropriate, effective, high quality test. Item analysis helps in deleting and/or revising the defective or ineffective items.
- Another objective of item analysis is to determine the relative difficulty level of each of the items on the test. Items only with proper difficulty level are retained for the final draft of the test.
- Determining how well did the test distinguish between students or determining the discriminating power of each item is another objective of item analysis.
- Multiple choice type items use distracters with the correct answer to the item. Item analysis helps to know if all the distracters are functioning properly.
- Item analysis also helps to find out why some items did not work properly, and thus suggests ways to improve them
- Another important purpose of item analysis is studying the strengths and weaknesses in educational achievement of the test-takers.
- Item analysis throws light on the nature of students' achievement, effectiveness of teaching, and the characteristics of test items.
- Item analysis gives insight into the act constructing better items by way of studying the incorrect responses of the test-takers.
- The main purpose of item analysis is to evaluate the qualitative standard of a test. The final draft of a test is prepared on the basis of the findings about each of the items acquired through item analysis.

Thus, the purpose of item analysis is to assess how well a test is functioning. Besides, each item on a test is intended to sample performance of test-takers on a particular learning outcome. The test as a whole estimates the performance on all the learning outcomes. Hence, analysis of each item is a must if we want to know the performance of individuals on particular learning outcomes.



CHECK YOUR PROGRESS

Q 1: What is item analysis?

.....

Q 2: Mention any two purposes of item analysis.

.....

5.4 ANALYSIS AND SELECTION OF TEST ITEMS

The process of item analysis mainly involves the tasks of determining difficulty level and discriminating power of each item besides analysis of distracters. Psychometricians and test experts believe that there is no need for analysing the responses of all the respondents for determining the relative difficulty level and discriminating power of test items; analysing the responses of a section of respondents may serve the purpose well. For this, two groups from among the respondents of the items should be formed and named the 'High group' and the 'Low group'. In order to form these groups, all the respondents should be ranked on the basis of their scores in the test. Then, for convenience of statistical calculations, any percentage between highest 25% and 35% of respondents should be included in the High group and any percentage between lowest 25% and 35% of respondents should be included in the Low group.

It is important to note at this point that various formulas are used to determine difficulty level and discriminating power of test items. In the following section the process of determining difficulty level and discriminating power using simple and easily understandable formulas have been presented.

5.4.1 Determination of Difficulty Level

The difficulty level of item is simply the percentage of students who answer an item correctly. The item difficulty index ranges from 0.00

to 1.00; the higher the value, the easier the item. If an item is answered correctly by everyone, or everyone gets it wrong, then it must be understood that the item is too easy or too difficult to be included in the test. Such item will not discriminate between good and poor students. Hence items only with average difficulty level should be included in a test.

In order to determine the difficulty level of an item statistically, the total number of students answering an item correctly in both high and low group should be divided by the total number of students in the two groups. This calculation can be done by using the following formula:

$$D_f = \frac{C_H + C_L}{N_H + N_L}$$

Where,

D_f= Difficulty Index

C_H= Number of respondents who attempted the item correctly in the High group

C_L= Number of respondents who attempted the item correctly in the Low group

N_H = Total number of respondents in the High group

N_L = Total number of respondents in the Low group

For example, if an item is attempted correctly by 12 out of 20 students in the high group and 8 out of 20 students in the low group, then the index of difficulty level for the item would be:

$$D_f = \frac{C_H + C_L}{N_H + N_L} = \frac{12 + 8}{20 + 20} = \frac{20}{40} = 0.50$$

It is important to remind here that item difficulty index may range from 0.00 to 1.00; the higher the value, the easier the item. According to famous statistician **H.E. Garrett**, any index of item difficulty ranging from 0.40 to 0.60 is acceptable.

The following table shows the different indices of difficulty levels and their interpretations:

Table 5.1: Indices of item difficulty level

Difficulty Index	Difficulty Level
0.00-0.20	Very difficult
0.21-0.40	Difficult
0.41-0.60	Moderate/most acceptable
0.61-0.80	Easy
0.81-1.00	Very easy

Source: Asaad &Hailaya (2004)

The procedure and formula discussed above may help in analyzing the item difficulty level for a test. However, it has been observed that an item with proper difficulty level may not be able to distinguish between good and poor learners sometimes. For example, if an item is attempted correctly by 10 out of 20 students in the high group and 10 out of 20 students in the low group, then the index of difficulty level for the item would be 0.50;

$$D_f = \frac{C_H + C_L}{N_H + N_L} = \frac{10+10}{20+20} = \frac{20}{40} = 0.50$$

Though the difficulty level for the item mentioned above is acceptable, the item could not make any discrimination between good and poor learners; because, the item was answered correctly by an equal number of learners from both high and the low group. On the other hand if an item is attempted correctly by all 20 out of 20 students in the high group and none out of 20 students in the low group, then the index of difficulty level for the item would again be 0.50.

$$D_f = \frac{C_H + C_L}{N_H + N_L} = \frac{20+0}{20+20} = \frac{20}{40} = 0.50$$

But, the discriminating power of this item would be highest; because this item is answered correctly by all in the high group and none in the low group. This example shows that only acceptable index of difficulty level cannot make for good items. This is why determination of discriminating power of an item is an indispensable part of item analysis.

5.4.2 Determination of Discriminating Power

A measuring tool must be able to discriminate between good and poor students. According to **Linn and Gronlund**, the discrimination index of an item is defined as "the degree to which it discriminates between students of high and low achievement". In other words, the item must be able to pick out and choose the good students from the poor ones; and more good students than poor students should be able to answer the item. There are various methods for determining the discriminating power of test item; one such method is to subtract the number of students giving correct answers in the low group from the number of students giving correct answers in the high group and dividing that by the number of students either in the high or in the low group. This calculation can be done by using the following formula:

$$D_S = \frac{C_H - C_L}{N}$$

Where,

D_S= Index of Discriminating power

C_H= Number of students attempting the item correctly in the High group

C_L= Number of respondents attempting the item correctly in the Low group

N = Total number of respondents in either the High or the Low group

For example, if an item is answered correctly by all the 20 students of the high group and by none in the low group, then the index of discriminating power of the item would be:

$$D_S = \frac{C_H - C_L}{N} = \frac{20 - 0}{20} = \frac{20}{20} = 1.00$$

In the same way, if an item is answered correctly by 10 out 20 in the high group and 10 out of 20 in the low group, then the index of discriminating power of the item would be:

$$D_S = \frac{C_H - C_L}{N} = \frac{10-10}{20} = \frac{0}{20} = 0.00$$

Important to note here is that, index of discriminating power of an item may range between +1.00 and -1.00. The highest level of discrimination of an item is indicated by +1.00 and this happens when an item is answered correctly by all the students of the high group and by none in the low group. When an equal number of students from both high and low group answer an item correctly, then the index of discriminating power would be 0.00. And a negative index of discrimination is obtained when more students from the low group than the high group answers the item correctly. Such item with negative index of discrimination should be readily deleted from the test.

The following table shows the different indices of discriminating power and their interpretations:

Table 5.2: Indices of discriminating power of test items

Index of discrimination	Interpretation
0.10 or less	Poor/ Questionable item
0.11-0.20	Not discriminating
0.21-0.30	Moderately discriminating
0.31-0.40	Discriminating
0.41-1.00	Very discriminating

Source: Asaad &Hailaya (2004)

5.4.3 Distracter Analysis

Distracter analysis is an important aspect of item analysis. In a Multiple Choice type item students are asked to select one of the alternatives from a list of possible answers. The correct alternative is called the key while the incorrect alternatives are called the distracter. The function of a distracter in a Multiple Choice type item is to distract those test takers from the correct answer who actually

do not know the answer. All the distracters in a multiple Choice Type item must work properly to make the item effective. Distracter analysis is the process of determining how the distracters were able to function effectively by drawing the test takers away from the correct answer. The process of distracter analysis involves counting the number of times each distracter is selected in order to determine the effectiveness of the distracter. An effective distracter should be selected by a significant number of respondents. However, an acceptable value in this regard depends to a large extent on the difficulty index of the item. For example, an item was attempted by a total of 100 students and the difficulty index of the item was 0.55, and it is expected that the remaining 0.45 should be evenly distributed among the distracters (item difficulty is equal to the percentage of students that correctly answered the item). Thus, if the item difficulty is 0.55, then 55 students answered the item correctly. Now if rest of all the 45 students selected option C, the distracters B and D are useless in their role as distracters. Similarly, if 25 students selected D and another 20 selected B, then C is not an effective distracter and should be removed. The ideal condition would be for each of the three distracters to be selected by 15 respondents. Therefore, for an item which has an item difficulty of 0.55, the ideal effectiveness of each distracter can be quantified as 15/100 or 0.15.

Distracter analysis of test items can also be done in another way. In this way, the formula used for determining item discrimination can be used in distracter analysis. However, the analysis and interpretation would differ from the regular item discrimination. Here, instead of a positive value, **a negative value is acceptable** since more students from the lower group should select distracters that are effective. In this method, the responses of the High and the Low group to an item is tabulated and the discrimination index calculated as shown below:

Table 5.3: Distractor analysis

Distractor	No of high group respondents who selected	No of low group respondents who selected	Index of Discrimination
A*	42	18	(42-18)/60=0.40
B	16	16	(16-16)/60=0.00
C	12	3	(12-3)/60=0.15
D	8	32	(8-32)/60=0.40

*** The Key or the correct answer.**

In the table above, Alternative A is the key and hence a positive value with a discrimination index of 0.40 is acceptable. The value for distracter B is 0 and this means the distracter could not discriminate between the good students in the high group and the poor students in the low group. Hence, effectiveness of this distracter is questionable and must be revised or removed. The positive value (0.15) obtained by distracter C shows that more of the good students selected this distracter. Hence, this distracter again is not functioning effectively. However, the final distracter, i.e. distracter D, has functioned effectively by attracting more students in the low group than in the high group. A distracter is effective when it seems to be the correct answer to poor students.

It is important to note here that if any distracter is not selected by any of the respondents from any of the groups, then the distracter is useless. When a distracter is not selected by anyone then it results in the reduction of one incorrect option and thereby helping respondents to guess the correct answer from the remaining options.



CHECK YOUR PROGRESS

Q 3: What do you mean by item difficulty level?

.....

Q 4: What is discriminating power of test items?

.....

Q 5: What is distractor analysis?

.....

5.5 PREPARING DIFFERENT TYPES OF TEST ITEMS

Results or scores on a good test must reflect the actual ability of the test-taker that the test supposes to measure. The type of test items to be constructed depends on the purpose of the test or the nature of the ability that the test intends to measure. Therefore, the test maker must have a good ground over the types of items with their relative merits and demerits. From a broad perspective, items can be categorized into two types: **Objective type test items, and Subjective type test items**

Objective type test items: This type of test items consist of factual questions that require very short answers that can be scored quickly and unambiguously by anyone with a scoring key. These items require short answers which usually consist of one word, a phrase or a sentence.

Subjective type test items: These types of test items involve subjective judgement of the scorer or evaluator. These items are difficult and expensive to prepare, administer and evaluate correctly, even though these items may be more valid.

Different types of objective type items include:

- Alternate response (true-false, yes-no) items
- Matching items
- Multiple choice items
- Completion items

Alternate response type test items: In this type of items, a factual statement is made and the learner is required to respond with any of the alternatives given (usually yes-no, or true-false) according to his judgement about the correctness of the statement. **The chief advantages** of this type of items include: these items are easy to prepare, can be scored objectively and they can cover a wide range of topics. Their **disadvantages**

include: not easy to construct the items that are definitely true or false or can simply be answered as yes or no, and they are prone to guessing.

Matching type items: This type of items require linking contents of one list to contents in another list. The test items are presented in the form of two columns of interrelated contents and ask the testee to link or connect the contents. The main **advantages** of this type of test items are: These items can measure associations and relationships as well as sequence of events very well; they can be used to ask questions beginning with who, when, where and what; these items are relatively easy to construct; and they are easy to score. The **disadvantages** of this type of items are: they cannot usually be constructed to measure higher order thinking and guessing is easy in this type of items.

Multiple choice test items: A multiple choice item is usually a statement followed by four or five alternative responses from which only the best or correct one must be selected. The statement or question is called the 'stem' and the alternatives or choices are the 'options' and the correct alternative is called the 'key. Incorrect options or alternatives are called 'distracters'. **Advantages** of multiple choice items are: they can measure various levels of learning; they are easy to score; very popular among all concerned; and proven to be an effective measurement tool. **Disadvantages** include: effective multiple choice items are difficult to construct; very well constructed items only can measure higher order of thinking; and contain a number of plausible distracters making it possible for the test-taker to guess the correct answer.

Completion type items: In completion type items, a sentence is presented from which an important word, figure, or phrase left out and learners are required to supply the words or figures which have been left out. These items are presented in the form of questions or phrases in which a learner is required to respond with a word or statement. **Advantages** of this type of item are: relatively easy to construct; can cover a wide range of content; and can reduce guessing. The **disadvantages** of this type of items are that these are primarily used for lower levels of thinking; items are sometimes ambiguous; construction is not simple; and scoring is dependent on the judgment of the evaluator.

Different types of Subjective type test items: In general there are two types of test which are: Short answer type or restricted response items, and Essay type items or extended response items.

A. **Short answer type items:** These are fairly structured items that require brief answers, usually within few words or sentences. These items can represent wide area of topic. Their **advantages** include: more valid and reliable than essay type tests; reasonably free of subjective influences since the boundary of answers are stipulated; students cannot write irrelevant answers; preparation, administration, and scoring is easy; less time is required to answer the questions; these items are a combination of essay type and objective type items. **Limitations** of short answer type items are: restricts the scope of the topic to be discussed; and indicate the nature of the desired response and thereby limits originality.

B. **Essay type items:** These items are traditional and popular that require lengthy answers and demand higher level thinking like analysis, synthesis, explanation, comparison, presentation etc. These items are very useful for measuring educational achievement and testing the depth of knowledge among the students. However, essay type answers are alleged to be less valid, less reliable, and often subjective. **Advantages** of Essay type answers: they can measure higher cognitive abilities such as analysis, synthesis and evaluation; ascertain individual difference in terms of attitudes, values and creative thinking; they can test abilities of written communication, representation, expression, and understanding; construction and administration of essay items is easy and economic; they can develop the habit of deeper study among students. The **limitations** of essay type answers are: they cannot cover wide area of content since only a few questions can be asked; students write irrelevant answers since the scope of answer is not specified; validity, reliability, and objectivity of essay type items is questionable; scoring of essay type answers is not free of subjectivity;

time consuming; measures memory rather than learning and understanding; reduces the speed of thinking to the speed of writing.

5.5.1 Guidelines for Construction of Different Types of Test Items

Following are some of the guidelines for construction of different types of test items-

A. Guidelines for construction Objective type test items:

- Objective type item should be constructed in such ways that demand answers that are either right or wrong and each question should have only one possible answer. Words such as, 'all', 'generally', 'never', 'absolutely' or 'none' etc. should not be used because these words are ambiguous in many situations.
- Statements in true-false items must always be either true or false, and must be unambiguous or clear. Negative statements should not be used.
- In matching type of items an unequal number of premises and responses may be provided in order to reduce guessing. The items should be kept short. Headings for each column should accurately describe its content and the basis for matching specified.
- The completion type items must be specific and unambiguous. For example: Education is a _____ process; this item is ambiguous because various words such as "dynamic", "social", "bi-polar", "tri-polar" or many other can be correct answer. Similarly, another item: Gandhiji was born in _____ is not clear if it was asking his year of birth or the place or even the country of his birth. If the answer is numerical or a quantity, then the unit must be indicated in the question. Sometimes using 'an' before the blank gives clue to the students that the answer must begin with a vowel.

- In a multiple choice items: a table of specification should be prepared to include items from topics and the skills to be tested. Multiple choice items should be constructed to test all the objectives of instruction. The statement of the item should state the problem clearly without unnecessary information. Options should be carefully selected and must include the best answer or key. Distracters should be relevant and not far-fetched. No item should provide clues or the answers to another question in the same test. All items should be independent. Tricky questions should be avoided and the instructions to test-takers should be very clear.

B. Guidelines for construction of subjective type test items:

- Test items should be prepared keeping the level of mental development of the test-takers in view.
- Items should not be too comprehensive, too general, and ambiguous.
- A model answer for every item should be prepared for guiding the act of scoring.
- Action verbs as solve, interpret, compare and contrast, discuss or explain should be used to direct the examinee to focus on the desired behaviour.
- Appropriate directions should be provided indicating expected length of the answer, time limits and features of the response that count in scoring.
- A scoring rubric should be created that would list the desired features of the response with a number of points awarded for each specific feature.
- In order to raise the reliability of items and to guard subjective scoring more than one evaluator may be employed to score the answers.
- The language used in writing test items should be clear and unambiguous.

- Useful words such as 'how', 'why', 'contrast', 'describe' and 'discuss' should be used in writing items for a test.
- The questions should clearly define the scope of the answer required.
- The time provided for the learner to respond to the questions should be sufficient.
- The items should match the goals or objective of testing.
- The length of the answer required should be specified.



5.6 LET US SUM UP

The unit can be summarized as-

- Item analysis is a very important process of standardized test construction. Item analysis evaluates the quality of test items. In its simplest meaning, item analysis is the process of determining relative difficulty level and discriminating power of the test items.
- The purpose of item analysis is to assess how well a test is functioning. It allows the test constructor to exercise quality control over the test. Item analysis serves the purpose of improving the qualitative standard of test items for using it in the future. It helps the process of construction of appropriate, effective, high quality test. Determining how well did the test distinguish between students or determining the discriminating power of each item is an objective of item analysis. The main purpose of item analysis is to evaluate the qualitative standard of a test. The final draft of a test is prepared on the basis of the findings about each of the items acquired through item analysis.
- The process of item analysis involves the tasks of determining difficulty level and discriminating power of each item besides analysis of distracters. The difficulty level of item is simply the percentage of students who answer an item correctly. In order to determine the difficulty level of an item statistically, the total number of students answering an item correctly in both high and low group should be

divided by the total number of students in the two groups.

- The discrimination index of an item is defined as the degree to which it discriminates between students of high and low achievement. An item must be able to pick out and choose the good students from the poor ones; and more good students than poor students should be able to answer the item. There are various methods for determining the discriminating power of test item; one such method is to subtract the number of students giving correct answers in the low group from the number of students giving correct answers in the high group and dividing that by the number of students either in the high or in the low group.
- Distracter analysis is an important aspect of item analysis. In a Multiple Choice type item students are asked to select one of the alternatives from a list of possible answers. The correct alternative is called the key while the incorrect alternatives are called the distracter. The function of a distracter in a Multiple Choice type item is to distract those test takers from the correct answer who actually do not know the answer. Distracter analysis is the process of determining how the distracters were able to function effectively by drawing the test takers away from the correct answer.
- From a broad perspective, items can be categorized into two types: Objective type test items, and Subjective type test items
  - Objective type test items consist of factual questions that require very short answers that can be scored quickly and unambiguously by anyone with a scoring key. These items require short answers which usually consist of one word, a phrase or a sentence.
  - Subjective type test items involve subjective judgment of the scorer or evaluator. These items are difficult and expensive to prepare, administer and evaluate correctly, even though these items may be more valid.
- Utmost precautions should be taken while constructing different types

of test items following proper guidelines so that items can be made valid, reliable, and objective.

5.7 FURTHER READING

1) Asaad, A.S. & Hailaya, W.M. (2004): Measurement and Evaluation; concepts and principles, Rex Book Store. Inc. Manila
2) Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001): The Autism-spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians, Journal of Autism and Developmental Disorders, 31, 5-17.
3) Bean, K.L.: Construction of Educational and Personal Tests, McGraw Hill Book, Inc. New York
4) Bloom, B.S. (1956): Taxonomy of Educational Objectives. Handbook I, Cognitive Domain, David McKay Co. Inc,New York
5) Gupta, S.P. (2009): Modern Measurement and Evaluations with Statistics, Sharda Pustak Bhavan, Allahabad
6) Linn, R.L. and Gronlund, N.E. (1995): Measurement and Assessment in Teaching, 7th Edition. Prentice-Hall, New Jersey
7) Nunnally, J. C. (1967): Psychometric Theory, McGraw-Hill, New York
8) Stanley & Hopkins (1966): Education and Psychological Measurement and Evaluation, Prentice Hall Pvt. Ltd. Eaglewood Cliff, New Jersey
9) UGC, (1972): Examination Reforms: A Plan of Actions, New Delhi

5.8 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Item analysis is a very important process of standardized test construction. Item analysis evaluates the quality of test items. Quality of a test depends on the quality of items. Hence, one cannot hope to construct a good test without analysing items constructed for the test. Item analysis in its simplest meaning is the process of determining relative difficulty level and discriminating power of the test items. In other words, the task of item analysis is to determine how

difficult or how easy an item was, and whether the item could differentiate between good and poor students. The validity and reliability of a test also depend on the validity and reliability of the items in the test.

Ans to Q No 2: Two purpose of item analysis are: (i) one important purpose of item analysis is to help the process of construction of appropriate, effective, high quality test. Item analysis helps in deleting and/or revising the defective or ineffective items, and (ii) another objective of item analysis is to determine the relative difficulty level of each of the items on the test. Items only with proper difficulty level are retained for the final draft of the test.

Ans to Q No 3: The difficulty level of item is simply the percentage of students who answer an item correctly. The item difficulty index ranges from 0.00 to 1.00; the higher the value, the easier the item. If an item is answered correctly by everyone, or everyone gets it wrong, then it must be understood that the item is too easy or too difficult to be included in the test. Such item will not discriminate between good and poor students. Hence items only with average difficulty level should be included in a test. In order to determine the difficulty level of an item statistically, the total number of students answering an item correctly in both high and low group should be divided by the total number of students in the two groups.

Ans to Q No 4: The discrimination index of an item is defined as the degree to which it discriminates between students of high and low achievement. In other words the item must be able to pick out and choose the good students from the poor ones; and more good students than poor students should be able to answer the item. Out of the various methods for determining discriminating power of test item one method is to subtract the number of students giving correct answers in the low group from the number of students giving correct answers in the high group and dividing that by the number of students either in the high or in the low group.

Ans to Q No 5: Distracter analysis is an important aspect of item analysis. In a Multiple Choice type item students are asked to select one of the

alternatives from a list of possible answers. The correct alternative is called the key while the incorrect alternatives are called the distracter. The function of a distracter in a Multiple Choice type item is to distract those test takers from the correct answer who actually do not know the answer. Distracter analysis is the process of determining how the distracters were able to function effectively by drawing the test takers away from the correct answer. The process of distracter analysis involves counting the number of times each distracter is selected in order to determine the effectiveness of the distracter.



5.9 MODEL QUESTIONS

A. Short Questions (Answer each question in around 150 words)

Q 1: What is the importance of item analysis?

Q 2: What are the merits and demerits of objective type test items?

Q 3: Write about the different types of test items.

B. Long Questions. (Answer each question in around 300-500 words)

Q 1: Explain the process of determination of difficulty level of test items.

Q 2: Explain how discriminating power of test items are calculated.

Q 3: Explain the process of distracter analysis.

Q 4: Suggest some measures to improve the quality of essay type test items.

*** ***** ***

UNIT 6: VALIDITY

UNIT STRUCTURE

6.1 Learning Objectives

6.2 Introduction

6.3 Meaning and Definitions of Validity

6.4 Factors affecting Validity of a test

6.5 Types of Validity of a test

6.6 Methods of estimating Validity of a test

6.7 Let Us Sum Up

6.8 Further Reading

6.9 Answer to Check Your Progress

6.10 Model Questions

6.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- define Validity of a test
- understand the factors that affect Validity of a test
- explain the different types of Validity of a test
- explain the methods of estimating Validity of a test

6.2 INTRODUCTION

Psychological tests and other tools of measurement serve various purposes in education. Various decisions in educational theory and practice depend on the results or data collected using psychological tests and other tools of measurement. However, if the data collecting instruments give erroneous results, then it would lead to wrong decisions and would fail to serve its objective. Therefore, an instrument of data collection must be valid to give correct results about the object of measurement. An attempt has been made in this unit to give a clear understanding about the concept, factors, types and methods of validity of a test.

6.3 MEANING AND DEFINITIONS OF VALIDITY

Validity is the most important technical criterion of a good test and other types of measuring instrument. Validity of a test may be defined as the extent to which a test measures what it claims to measure. In other words, validity answers the question of whether the test measures what it intends to measure. The concept of Validity is so fundamental that it precedes the aspect of Reliability in importance. A test may be reliable but not valid. However, if a test is not valid, then its reliability has no meaning or importance. On the other hand, a valid test is usually reliable. The concept of validity is not absolute and hence validity of a test is always understood or determined with reference to the purpose of the test. Therefore a test may be valid for one purpose but not for others. Any test can be differentially valid, given the various uses to which it is put. For example, a test constructed to assess the capacity for logical thinking of adolescents may serve the stated purpose extremely well, but when the same test is given to students of a different age group, it may fail to give the desired results. Different psychologists have given various definitions of validity. A few of the important definitions are quoted below:

According to **L. J. Cronbach**, "Validity is the extent to which a test measures what it purports to measure."

F. S. Freeman defines validity as, "An index" that "shows the degree to which a test measures what it purports to measure when compared with accepted criterion."

Henry Garrett stated that, "The validity of a test or of any measuring instrument depends upon the fidelity with which it measures what it purports to measure."

Anastasi says, "The question of test validity concerns what the test measures and how well it does so."

In order to ascertain if a test actually measures what it aims to measure, two sets of measures of the variable aimed to measure are needed. One set of measure is obtained by the test itself and the other set is obtained by some well-established and acceptable instrument that

measures the variable. The latter is known as the Criterion. If the scores or measures obtained by the test correspond well with the criterion then the test may be accepted as valid. Thus, the validity of a test is the correspondence or correlation between the variable measured by the test and the variable objectively defined and measured. The level of correspondence is shown by the coefficient of correlation obtained from measures of the test and criterion which is also known as validity Coefficient. Hence, Monroe states that, "under the head of validity we inquire into the degree of constancy of the functional relation existing between the scores yielded by the test and the abilities specified as being measured in the statement of that function." Thus, Validity is capable of a statistical interpretation by means of a coefficient of correlation. This coefficient of correlation is estimated between the two series of scores: those of the tests and of the accepted criterion.

6.4 FACTORS AFFECTING VALIDITY OF A TEST

The validity of a test is affected by various factors. The most important factors that affect the validity of a test are as follows:

- **Factors inherent in the test itself:** The validity of a test largely depends on the items included in the test. In fact, validity of the test depends on how well each item included in the test could achieve the objectives of the test. If the items in a test fail to serve the purpose for which they were constructed, then the test will lack validity. Moreover, there are other factors related to the test that affect the validity the test. These factors include:
  - **Clarity of instruction:** Clarity of instruction about how to answer and record the items, if guessing is allowed etc. always raises the validity coefficient of a test. On the other hand, unclear instruction to the testee may reduce validity of a test.
  - **Length of the test:** The length of a test too affects its validity. A test is a collection of many items that represent the content or cover the objectives of the test. But if the test does not include

enough items, it may fail to represent the whole content or objectives affecting its validity adversely.

- **Language used by test items:** The language of test items should always be clear, simple and unambiguous. Complicated vocabulary, complex structure of sentence, ambiguous words and phrases make it difficult for the takers of the test to understand the items which lowers the validity of the test.
- **Difficulty level of test item:** Difficulty level of items included in the test may also affect the validity of the test. Generally it has been observed that when the difficulty level of the items such in a test is homogeneous, validity of the test is low; on the other hand, when difficulty level in a test is heterogeneous, validity of the test is high.
- **Item discrimination power:** Like the level of difficulty, the power of discrimination of test items also affects the validity of a test. Tests with items having higher discrimination tend to possess higher validity and tests with items having low discrimination are of lower validity.
- **Defective construction of items:** Defective item construction lowers the validity of the test. Items that provide clues to answer, or items that carry answers to some other question, or such other defects in item construction lower the validity of the test.
- **Inappropriate items:** Sometimes items are constructed to measure one trait or variable, and in practice they measure some other trait or variable. Such inappropriate items lower the validity of the test.
- **Faulty arrangement of items:** Items in a test must be arranged logically and in an ascending order of difficulty. If items are not arranged maintaining a logical sequence, that may affect the validity of the test adversely.
- **Guessing:** If the items and the test is constructed and arranged in such ways that it allows the students to guess the correct answers then the validity of the test will surely be doubtful.

- **The Criterion of validation:** In order to estimate validity of a test the scores obtained in the test have to be compared with some acceptable measure or criterion. Various indices of validity may be obtained when different criterion is used to validate the test.
- **Method of estimating validity:** There are various methods of estimating the validity of a test. Different estimates of validity are obtained when different methods are used to estimate validity of the test.
- **Nature of the group:** Validity of a test also depends on the nature of the sampling group on which the test is administered for estimating validity. Validity is always specific to a particular group of individuals.
- **Factors related to test-takers' response:** The psycho-physical condition of the test-takers, their emotional and mental conditions besides cultural factors including socio-economic condition, social class, language, religion, educational background, sex etc. influence the scores of individuals and thereby affect the validity of a test.
- **Factors related to administration and scoring of the test:** Sometimes administration and scoring procedure of a test may interfere with its validity. Factors such as unclear direction, inappropriate time, cheating, and unfair help to candidates, unreliable and subjective scoring and adverse psycho-physical conditions affect the scores of individuals on the test ultimately affecting its validity.

The factors discussed above and many other similar factors may affect the validity of a test.

6.5 TYPES OF VALIDITY OF A TEST

Validity of a test is closely associated with the purpose of the test. It is for this reason that the validity of a test is always estimated with reference to the purpose of the test. Every test has its own distinctive purpose. On the basis of the different purposes of different tests validity is estimated

using various methods. Thus each test has its own different type of validity. Different test-experts and Psychometricians have classified validity into different types and named them somewhat differently. Out of the many different types, the following are the most popular and widely used types of validity.

- Face validity
- Content validity
- Criterion validity
- Construct validity
- Factorial validity

Face validity: Face validity is the simplest type of validity of a test. However, face validity is not validity in actual psychometric meaning. It is a superficial or extra quality of a test and it refers only to what the test 'appears to measure' and not to what it actually measures. Basically, face validity is not one of validity in the technical sense but from a practical point of view it is important for establishing rapport and to increase acceptability of the test among stakeholders. The items of the test should not appear to be inappropriate and irrelevant. When the test items appear irrelevant or inappropriate, the result is always poor regardless of the actual validity of the test. Thus, if the test consists of items that look to measure the variable that the test intends to measure, then the test has face validity. For example, if items in a test constructed for selecting candidates for recruitment of teachers use words and phrases that relate to classroom and teaching, then it may be said that the test has face validity. But at the same time it cannot be said that the items in the test could actually measure the skills required for teaching. It is noteworthy that even though a test possesses face validity, it may not be truly valid. Hence, while assessing the validity of a test, only face validity is not sufficient. It is generally considered only at the preliminary stage of validating a test. Moreover, when there is an urgent need to estimate the validity of a test within very limited time, then face validity may be used. Face validity, in a sense, ensures relevance to the context by using appropriate words and phrases in the items.

Content validity: Content validity is also known as curricular validity, Rational or Logical validity, internal validity etc. Content validity is particularly applicable in the tests of educational achievement. In the field of education, learners are guided through a curriculum in order to achieve the goals and objectives of education. At the end of education, if the items included in the test to measure the learner's progress can represent the curriculum correctly and can assess the extent of achievement of the objectives, then the test is said to possess content or curricular validity.

Content validity is the extent to which the items of a test match the instructional objectives. In other words content validity is the process of determining whether the sample used for the test is representative of the area or aspect to be measured. It is non-statistical type of validity. Thus, content validity may be defined as the extent to which a test measures a representative sample of the subject matter or content and the behavioural changes being measured. However, it goes without saying that the content validity of a test must always be examined with reference to a particular objective to be measured. The measure of content validity is presented subjectively after careful examination of the content of the test with the objectives of the course of instruction.

Content validity is mostly relevant for tests used to evaluate achievement, education, occupation, and the like. At the same time, content validity is not appropriate for test of aptitude, personality, intelligence etc.

Criterion Validity: Criterion Validity is also known as Empirical Validity or Statistical Validity. In criterion validity or criterion-related validity, the scores obtained by a test are compared and correlated with some criterion. A criterion is a standard with which scores on a test are compared or evaluated. Thus, criterion validity is the extent to which the test scores correspond with the criterion. It is a quantitative or statistical procedure of calculating the coefficient of correlation between scores on the test and the criterion. The validity of the test is also determined by its ability to predict performance on the criterion. There are different types of criterion related validity based on different criteria that are used as the standard for judgment.

They are as follows:

Predictive Validity: Predictive Validity refers to the extent to which a test can predict something that it should theoretically be able to predict. The predictive validity of a test is determined on the basis of an established criterion. A test has predictive validity, if scores on it predict future performance of test-takers in a particular field. In order to calculate predictive validity the test is administered and scores are obtained. In the next stage, the scores of the same individuals signifying their actual performance in the particular area (criterion) are compared with their previous scores on the test. If the two sets of scores correlate closely then the test is said to possess predictive validity. For example, scores obtained by students in an entrance examination are compared and correlated with their actual scores in that course or curriculum (for which they were selected by the entrance examination) in order to estimate predictive validity of the test. Predictive validity is substantial for entrance exams, vocational tests, and aptitude tests.

Concurrent Validity: Concurrent Validity is the extent to which a test correlates with a previously validated test for the same construct. In order to estimate concurrent validity of a newly constructed test, the scores obtained by students in the (newly constructed) test are compared and correlated with their scores on a previously validated and standardized test. For example, scores on a newly constructed test are compared with Stanford-Binet in order to estimate validity of the new test. When a new test is validated against some established and standardized test, the latter is known as the criterion of validation. When a test constructor makes a new test and wants to make it sure that the test accurately measures the variable for which the test was constructed, he has to follow the procedure. In order to do this he has to find out established tests for measuring the same variable, administer both the tests, and compare the results of the tests with each other. If the results closely correlate, then the new test is having concurrent validity. Criterion validity is generally used for validation of intelligence, multiple aptitude and tests of personality.

Construct Validity: In psychology, a construct is a concept developed in the mind to explain some phenomenon. Psychological tests are constructed to measure different constructs. While making a test for measuring a particular construct, the test maker must clearly define the construct in clear terms what he wants to measure. Construct validity is the extent to which a test measures the theoretical construct as defined by the test maker. According to Anastasi, construct validity is the extent to which the test may be said to measure a theoretical construct or trait or psychological variable.

Factorial Validity: A test is usually constructed to measure different mental construct or factors. For measuring different such factors or constructs, different units of items are included in a test. Factorial validity refers to the coefficient of correlation between the scores on units of items constituting a factor and the scores on the total test. Tests are generally composed of different factors. The correlation of the test with each factor constituting the test is called factorial validity of the test.



CHECK YOUR PROGRESS

Q 1: What is validity?

.....

Q 2: What is face validity?

.....

Q 3: Write any two factors that affects the validity of a test.

.....

Q 4: What are the two different types of criterion-related validity?

.....

Q 5: What is Construct Validity?

.....

6.6 METHODS OF ESTIMATING VALIDITY OF A TEST

Validity of a newly constructed test may be estimated using various methods. Following are some of the common methods used to estimate validity of a test:

Correlation Coefficient method: Test constructors generally validate new tests by correlating the scores of the new test with the results obtained from the best accepted test, which is known as the criterion of validation. The coefficient of correlation gives the extent of validity index of the test. A high coefficient of correlation between the results of the two tests indicates high validity for the new test. When coefficient of correlation is used to describe validity, they are called validity coefficients. According to Safrit and Wood, in general validity coefficients of 0.90 or higher are desirable but values above 0.80 are acceptable. They also say that, for tests used to make predictions lower validity co-efficient (0.50 to 0.60) are acceptable.

For the purpose of validating new tests using the correlation coefficient method, Pearson's method of correlation is most widely and popularly used. However, the technique of correlation depends on the nature of data obtained on the test as well as on criterion. According to Berg and Latin (2008), following methods of correlation are commonly used for validity of newly constructed tests:

Validity type	Statistics (method of correlation)
Face validity	None
Content validity	None
Concurrent and predictive validity	✓ Pearson Correlation ✓ Spearmen Correlation ✓ Multiple Correlation
Construct validity	Independent t

The table presented above clarifies one important thing about the Correlation Coefficient method that this method can be applied only in the cases of concurrent and predictive validity.

Cross-Validation Method: Cross Validation is the process of checking the validity of a test for a second time on a second sample. Whatever method is used for validating the test, cross validation is very important. It is a process in which a test is administered to a different sample of individuals drawn from the same population for the purpose of verifying the results obtained from the original process of validation. In other words, cross validation means trying out the previously developed and refined test on a completely new group for a second examination of its validity.

The common types of cross validation methods include:

Holdout validation: Though a method of cross validation, in this method the data never are crossed over. Observations or scores are chosen randomly from the preliminary administration to form a sample of the validation data, and the remaining observations are retained as the training data. Normally, less than one third of the initial sample is used for validation data.

K-fold cross-validation: In this method, the original sample is divided into K subsamples. Out of the K subsamples, one subsample is used as the validation data, and the remaining sub-samples are used as training data. The cross-validation process is then repeated K times (the folds), with each of the K subsamples used once as the validation data. The results from all the folds are then averaged or combined to produce a single estimation.

Leave-one-out cross-validation: This method uses a single observation from the original sample as the validation data, and the other observations as the training data. This process is repeated and each observation in the sample is used once as the validation data.

Expectancy Table Method: An expectancy table is a two-way table showing the relationship between two tests. According to Helmstadter, an expectancy table provides the probabilities indicating that students achieving a given test score will behave in a certain way in some second situation. This method can be used to help in predicting the performance level on another test; in other words, an expectancy table can be used to estimate predictive validity.

An expectancy table helps in estimating the chances of an individual's success in some future activity from his present performances on a test. It is a table of numbers from which the likelihood or probability that a particular individual or group of individuals will attain a specified level of success can be predicted.

An expectancy table is a simple and straightforward method of estimating predictive validity of a test. This table may help in answering questions like: what are the probabilities of a student securing 1st Division in Secondary School Leaving Examination completing all of the six semesters of Degree Course? Or what are the probabilities of a student passing in the 3rd Division of successfully completing the Degree Course? An expectancy table may help in answering such questions. The following illustration may show how an expectancy table works:

Table: Expectancy table

Divisions in S.L.E	Percentage of students Completing Semesters					
	1	2	3	4	5	6
1st	99	95	91	89	89	85
2nd	93	90	85	83	78	75
3rd	80	75	71	67	61	57

The table above shows that the probability of a student securing 1st Division in Secondary School Leaving Examination of completing all the six semesters of Degree Course is 85%. Similarly probabilities of a student passing in the 3rd Division of successfully completing the Degree Course are 57% only.

It is important to note that the format of an expectancy table is not uniform. The format and the arrangement of data depend on what the user wants to accomplish or predict. However, all expectancy tables provide estimation of the probabilities that a particular level of performance may be predicted if the previous test score is known. Thus an expectancy table estimates the validity of the test scores in terms of their predictive efficiency in the form of probabilities.

Method of Factor Analysis: Factor analysis is a statistical method used to study inter-correlation of items in a test to establish its validity. Factor analysis is a technique that is used to reduce a large number of items in a test into fewer numbers of factors. The correlation of the test with each factor constituting the test is calculated to find out the contribution of each of the factors to the total performance of the test and also to find out if all the factors are internally consistent and assessing the same variables as intended by the test.

Besides the methods discussed above, other methods such as item analysis, expert's judgment and opinion etc. are also used to validate newly constructed tests.



6.7 LET US SUM UP

After going through the unit, it can be summarized as---

- Validity refers to the truthfulness of a test. Validity of a test is defined as the extent to which a test measures what it claims to measure. The concept of validity is so fundamental that it precedes the aspect of reliability in importance. A test may be reliable but not valid. However, if a test is not valid, then its reliability has no meaning or importance. On the other hand, a valid test is usually reliable. The concept of validity is not absolute and hence validity of a test is always understood or determined with reference to the purpose of the test. Therefore a test may be valid for one purpose but not for others.
- Validity of a test is affected by various factors such as: factors inherent in the test itself including difficulty and discrimination level of items, length of the test, quality of items, clarity of instruction, guessing etc.; the criterion of validation; methods of estimating validity; nature of the group; factors related to test-takers' response; factors related to administration and scoring of the test and the like.
- Different test-experts and Psychometricians have classified validity into different types; the most popular and widely used types of validity

are: Face validity, Content validity, Criterion validity, Construct validity and Factorial validity.

- There are various methods of estimating the validity of a test. The important methods include: Correlation Coefficient method, Cross-validation method, Expectancy table method, Factor Analysis and the like.



6.8 FURTHER READING

1) Anastasi, Anne (1982): Psychological testing, Edition 5, illustrated. Macmillan, The University of Michigan
2) Berg, Kris E. and Latin, Richard Wayne (2008): Essentials of Research Methods in Health, Physical Education, Exercise Science, and Recreation, Lippincott Williams & Wilkins
3) Ebel, R.L. and Frisbel, D.A. (1979): Essential of Educational Measurement, 3rd Edition, Prentice Hall of India, New Delhi
4) Gupta, S.P. (2009): Modern Measurement and Evaluations with Statistics, Sharda Pustak Bhavan, Allahabad
5) Kelley, T. L. (1927): Interpretation of educational measurements, Macmillan, New York
6) Newton, Paul & Shaw, Stuart (2014): Validity in Educational and Psychological Assessment, Sage Publications, London.
7) Safrit, Margaret J. And Wood, Terry M. (1995): Introduction to Measurement in Physical Education and Exercise Science, Mosby, Missouri
8) Stanley & Hopkins (1966): Education and Psychological Measurement and Evaluation, Prentice Hall Pvt. Ltd. Eaglewood Cliff, New Jersey



6.9 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Validity is the most important technical criteria of a good test or other measuring instrument. Validity refers to the truthfulness of a test. Every test is constructed with a particular purpose of measuring a mental trait or criterion. Validity of a test is concerned with the purpose

of the test and it is the extent to which the test measures what it purports to measure.

Ans to Q No 2: Face validity is the simplest type of validity of a test. Face validity is not validity in actual psychometric meaning. It is a superficial or extra quality of a test and it refers only to what the test 'appears to measure' and not to what it actually measures.

Ans to Q No 3: Two factors that affect the validity of a test are: (i) the validity of a test largely depends on the items included in the test. In fact, validity of the test depends on how well each item included in the test could achieve the objectives of the test; (ii) Validity of a test also depends on the nature of the sampling group on which the test is administered for estimating validity. Validity is always specific to a particular group of individuals.

Ans to Q No 4: The two different types of criterion related validity are: (i) predictive validity, and (ii) Concurrent validity.

Ans to Q No 5: Construct validity is the extent to which a test measures the theoretical construct as defined by the test maker. According to Anastasi, construct validity is the extent to which the test may be said to measure a theoretical construct or trait or psychological variable.



6.10 MODEL QUESTIONS

A. Short Questions (Answer each question in around 150 words)

Q 1: What is the importance of validity in test construction?

Q 2: Define content validity of a test.

Q 3: What is construct validity of a test?

B. Long Questions. (Answer each question in around 300-500 words)

Q 1: Discuss the factors that affect the validity of a test.

Q 2: What do you mean by criterion validity? Discuss the different types of criterion related validity.

Q 3: Explain the different methods of estimating validity of a test.

*** ***** ***

UNIT 7: RELIABILITY

UNIT STRUCTURE

7.1 Learning Objectives

7.2 Introduction

7.3 Meaning and definition of Reliability

7.4 Factors affecting Reliability of a test

7.5 Methods of estimating Reliability of a test

7.6 Relation between Reliability and Validity

7.7 Let Us Sum Up

7.8 Further Reading

7.9 Answer to Check Your Progress

7.10 Model Questions

7.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- define Reliability of a test
- know about the different factors that affect Reliability of a test
- explain the process of estimating Reliability of a test
- understand the relation between Reliability and Validity.

7.2 INTRODUCTION

Reliability is one of the most important technical criteria of a good test. Any measuring tool used to collect data for measurement and evaluation in education or in any other field, must be reliable besides fulfilling other criteria such as validity, objectivity and the like. Reliability, like validity is a very important criterion of any measuring instrument. A good test or other measuring device must be reliable. If the tools used for data collection are erroneous, the data used will also be erroneous. In order for any data to be of value and of use, they must be reliable as well as valid. Here in this unit we will discuss the meaning and other aspects of reliability in a detailed way.

7.3 MEANING AND DEFINITION OF RELIABILITY

Reliability refers to consistency of results or findings. When an investigation is repeated it must yield same results if it were to be reliable. When more than one investigator or evaluator is making the measurement, all of them must arrive at similar findings in order to claim that the data are reliable. Similarly, when an individual takes the same test twice, if the scores on the two occasions are very similar, then the test can be described as reliable. This means, a reliable test must give the same result when given twice to the same person within a short period of time and when different evaluators make the measurement. On the contrary if a test or other measuring device gives different results on different occasions or when used by different evaluators on the same individuals, then it should be understood that the test is not reliable.

Reliability of a test refers to the consistency of scores obtained by students on alternate forms or different administration of the same test. However, no two tests with exact content can ever give identical results due to differences in the environmental conditions including psycho-physical state of the test-takers. Even the same test gives different results when administered to the same group of students on two different occasions. Nevertheless, whenever we administer two different tests with similar material or the same test on different occasions, we wish that the scores obtained by students in the two different administrations are similar. In this situation, the test is reliable to the extent to which the scores are comparable.

Reliability of a test is an indicator of validity of the test as well. If a test cannot give consistent results, it is obvious that the test is not accurately measuring the trait or variable of interest. A test to be valid must be reliable. Therefore, estimation of reliability is considered to be the first-step in the process of validating a test. If the test is not reliable, then there is no need of examining the validity and it can be easily concluded that the test is not valid. Several methods are used to estimate reliability of a test depending on the nature of the measurement.

In the views of Anastasi, "Reliability refers to the consistency of

scores obtained by the same individuals when re-examined with the same test on different occasions or with different sets of equivalent items or under variable examining conditions."

According to Freeman, "The term reliability refers to the extent to which it gives consistent result on testing and retesting."

Henry E. Garrett says, "The reliability of a test or of any measuring instrument depends upon the consistency with which it gauges the ability to whom it is applied."

According to Stodola & Stordahl, "Reliability is the correlation between two or more sets of scores on equivalent tests from the group of individuals."

In the views of Mehrens and Lehmann, "Reliability can be defined as the degree of consistency between two measures of the same thing."

After an analysis of the definitions mentioned above, it can said that reliability refers to the consistency of test scores. It represents the extent of accuracy or precision of a measuring tool. If a test or other tool gives consistent results on testing and retesting in variable conditions then the tool is accepted to be reliable. Reliability is a precondition to the validity of a test as well. However, reliability is the characteristic of the scores only and it is not the property of the test itself.

7.4 FACTORS AFFECTING RELIABILITY OF A TEST

While constructing or using a test, it is very important to understand that reliability of a test may be affected by various factors. Some of the important factors that affect reliability of a test may be discussed as under:

- **Length of the test:** Test reliability depends on the length of the test. According to Psychometricians, increasing the length of a test by adding items of acceptable quality and representing the content increases the reliability of the test. On the contrary if the length of test is reduced by deleting items then the reliability of the test will also be reduced.

- **Methods used to estimate reliability:** There are different methods that may be used to estimate reliability of tests. Different indices of reliability of the same test may be obtained when different methods to estimate reliability are used.
- **Nature of the group:** Reliability of a test also depends on the nature of the group upon which the test is administered to estimate reliability. The test is more reliable if the group of students on which the test is administered is more heterogeneous in ability; conversely, if the test is administered to a homogeneous group then the reliability of the test is low.
- **Difficulty level of items:** Difficulty level of test items may also affect reliability of a test. Too difficult or too easy items in a test reduce reliability of the test.
- **Objectivity of scoring:** reliability of a test is high when the scoring procedure is objective. Subjective scoring leads to variation in scores in the test resulting in lower consistency of scores and thus affect the reliability index of the test adversely.
- **Clarity of instruction:** When the instructions to candidates who take a test are clear and they understand what to do, the scores seem to be consistent on repeated testing. If the instructions are unclear then reliability of the test goes down.
- **Interval between testing and retesting:** In order to estimate reliability coefficient of a test or other measuring device, the instrument has to be administered twice to get two sets of scores. The interval of time between the two administrations of the test affects the reliability index of the test. Normally, when the interval between the two administrations of the test is short then the reliability index of the test is high and when the interval between the two administrations of the test is long then the reliability index of the test is low.
- **Environment of test administration:** the physical condition under which the test is administered for estimating reliability, affects the test. Factors such as noise, temperature, light, seating arrangement

etc. influence the test scores which ultimately affect the reliability of the test.

- **Psycho-physical condition of the test-takers:** The psychological and physical conditions of the respondents while appearing the test affect their test scores. Factors such as physical wellbeing, emotional stability, motivation, anxiety, etc. affect test scores and thereby affect the reliability of a test.
- **Guessing:** When the items provide scope for guessing while answering the items then that leads to inconsistent scores by same individuals on the same test. Thus, the factor of guessing may significantly affect the reliability of the test.

These are some of the important factors affecting the reliability of a test.



CHECK YOUR PROGRESS

Q 1: What is reliability?

.....

Q 2: Mention two factors that affect reliability of a test.

.....

Q 3: What are the two types of reliability?

.....

7.5 METHODS OF ESTIMATING RELIABILITY OF A TEST

There are various methods of estimating reliability of test. However, the different methods of estimating reliability may be categorized under two different types. They are: (i) methods of estimating relative reliability, and (ii) methods of estimating absolute reliability. Relative reliability is generally stated in terms of a coefficient or correlation, which is known as reliability coefficient. In this method the correlation between the scores obtained from

two different administration of the test is estimated and labelled as reliability coefficient. The second type of methods estimating absolute reliability is known as Standard Error of Measurement. Standard error of measurement is actually an estimate of the deviation of a set of obtained scores from their true scores.

Several methods are used to estimate relative reliability or reliability coefficient of a test. The most popular methods are:

- The test-retest method of reliability
- The equivalent or parallel form method of reliability
- The split-half or the odd-even method of reliability
- The method of rational equivalence or the Kuder-Richardson method of reliability

The test-retest method: This is a very simple method of estimating reliability of a test. In this method, the same test or the same form of the test is administered twice on the same group of individuals with a reasonable interval of time between the two administrations. Thus, the two sets of scores are obtained from the two administrations of the test. Now, in order to estimate the reliability of the test, coefficient of correlation is calculated between the two sets of scores. This coefficient of correlation is the test-retest reliability of the test which is sometimes called "temporal stability" or "coefficient of stability" as well. This reliability coefficient indicates the extent of consistency of scores in the administration of the test. The greater the index of coefficient of correlation, the more reliable is the test.

The advantages of the test-retest method of reliability are:

- it is a simple method,
- it is an economic method, since only one administration of the test is required to estimate this reliability,
- this method does not require the construction of a second test or form,
- this method is applicable to almost all types of tests,
- this method does not involve complex statistical procedures, simple coefficient of correlation helps.

The disadvantages of the test-retest method of reliability are:

- it is difficult to decide the interval between the two administrations of the test;
- if the interval between the two administration is short, immediate memory, habit, familiarity with the test items may raise the reliability coefficient to a very high level and on the other hand if the interval between the two administration is long, changes in behaviour and maturity of students may make the reliability coefficient low,
- if the conditions under which the two different administrations of the test held are different, then that results in low reliability of the test,
- this method is time consuming.

The equivalent or parallel form method: This method is also known as alternative form method of reliability. In order to estimate reliability using this method, two equivalent or parallel form of the test is prepared. These two forms of the test should be similar in terms of objectives, content, difficulty level of items, form and type of items, length of the test, and other such characteristics. Efforts should be made to make the two forms of the test as much similar as possible but not the same. Freeman holds that both forms (of the test) should meet all of the test's specifications as follows:

The number of items should be the same.

- The kinds of items in both should be uniform in respect to content, operations or traits involved, levels and range of difficulty, and adequacy of sampling.
- The items should be similarly distributed as to difficulty.
- Both test forms should have the same degree of item homogeneity in the operations or traits being measured. The degree of homogeneity may be shown by inter-correlations of each item with subtest scores, or with total-test scores.
- The means and standard deviations of both forms should correspond closely.

- The mechanics of administering and scoring should be uniform.

Though these ideal conditions are desirable, achieving all these is very difficult. After preparing the two equivalent forms of the test, they are administered to the same group of individuals one after the other. The two sets of scores obtained from administering the two forms of the test are correlated to calculate the coefficient of correlation. This coefficient of correlation is the reliability coefficient of the test. According to Guilford, the alternate form method indicates both equivalence of content and stability of performance.

The advantages of the equivalent form method are as follows:

- In this method there is no need to administer the test twice,
- There is no problem of time interval between two administrations,
- The problem of familiarity with test or immediate memory affecting reliability is absent here,
- This method is an advancement over the test-retest method.

The limitations of this method are:

- It is very difficult to prepare two identical forms of a single test,
- This method requires the construction and standardization of a second form of the test which is time consuming and difficult,
- Since items in the two forms are similar in content, difficulty and form, the practice and familiarity effect cannot be controlled completely,
- There are many tests that cannot be divided into two forms.

The split-half or the odd-even method: The split half method is the most commonly used method of estimating reliability of a test. This method divides the test into two equal halves and the scores on one half of the items are correlated with the scores on the other half. However, the test can be split into two halves by using different methods. Commonly, the odd numbered items are pulled together to form one half and the even number items are pulled together to form another half of the test. That is why this method is also known as odd-even method. This method is applied by administering the test only once to the group and then scores on the two halves is drawn separately. Then the scores on the two halves are correlated

using the Spearman-Brown formula to find out the reliability coefficient of the test. Thus, this method measures the internal reliability of the test. If the results show that the two halves do not correlate highly it implies that the two halves are not measuring the same variable. When preparing parallel forms of a test are not advisable and re-testing is not possible this method of estimating reliability is used. Usually this method is used to estimate reliability of attitude scale, personality, inventories, questionnaire, intelligence, performance test, etc. This method of estimating reliability is most popular since only one administration of the test is required. This method thereby saves time, money and effort required to construct and administer a second test.

The major advantages of this method are:

- The problem of preparing equivalent halves of same test is eliminated,
- Issues of memory, familiarity, practice effects, fatigue, maturity etc are removed,
- The variations in testing situations are eliminated since only one administration is required.

The main disadvantages of this method are:

- A test can be divided into two halves in a number of ways; and hence the correlation between the scores on the two halves may not have a unique value,
- It is very difficult to divide a test into two halves in such a way that both halves are equivalent,
- Splitting a test into two equivalent forms is difficult also because items of a test measure same trait or ability.
- This method cannot be used in power tests and heterogeneous tests.

The method of rational equivalence: The rational equivalence method is another method of estimating reliability based on a single administration of the test. This method uses the formula developed by Kuder-Richardson, and hence this method is also known as Kuder-Richardson method. This method can be appropriately used when all the

items in the test are psychologically homogeneous. This means that all the items in the test must measure the same variable or variables in the same proportion. This is known as 'inter-item consistency'. Inter-item consistency means that every item in the test is highly correlated with all other items. If the items in the test are not highly homogeneous, then this method would give a very low index of reliability coefficient.

Kuder-Richardson developed different formulas for computing reliability of a test. Following is the most popularly used formula:

$$r_{11} = \left(\frac{n}{n-1}\right)\left(\frac{SD_t^z - \sum pq}{SD_t^z}\right)$$

Where,

r_{11}= reliability coefficient of the whole test

n = total number of items in the test

SD = Standard Deviation of the whole test

SD^2= the variance of the total test

P = the percent of individuals passing each item

Q = the percent of individuals not passing each item

$\sum pq$ = the sum of the products of pq

The chief advantages of using this method are:

The difficulties of the other methods can be overcome by using this method,

- This method requires single administration of the test,
- The method is simple and economic,
- This method gives accurate results when items are homogeneous,
- This method is scientific.

The limitations of this method are:

- This method cannot be used successfully when the items of the test are not homogeneous,
- Different formulas given by Kuder-Richardson give different estimates of reliability of the same test
- This method does not work with power test or tests with heterogeneous items.

Method of estimating absolute reliability or Standard Error of Measurement:

There is hardly any measurement tool that is perfectly reliable. All instruments have some errors and individuals who measure also make mistakes. Hence inconsistency of tools of measurement and those who use them is not very uncommon to find. Hence, scores obtained by individuals in a test (X) can be thought of as a function of two components, i.e. a true score (T) and an error component (E). Thus, an obtained score(X) = T ± E

The variation between the true score and the observed score is called the measurement error. In statistics, 'error' is any source of variability that cannot be explained. A true score is free from errors of all kinds and it remains the same. But the obtained score may vary from trial to trial because obtained score is influenced by errors. Theoretically, true score is represented by the average (Mean) of the obtained scores on an infinite number of measurements. Each obtained score is either bigger or smaller than the true score. This variation of obtained score from true score is called the errors of measurement. When the error of measurement is smaller, the reliability coefficient will be higher. The errors of measurement or the variation of obtained score from true score are normally distributed and the standard deviation of these variations or errors of measurement is called the standard error of measurement (SEM). The standard error of measurement is a reflection of the reliability of the test.

The formula for determining SEM is written as below:

$$SE_M == SD_x \sqrt{1 - r_{xx}}$$

Where,

SE_M = Standard Error of Measurement

SD_x = Standard deviation of the obtained scores

r_{xx} = Reliability Coefficient of the test

Since all measurement contains some error, it is highly unlikely that any test will yield the same scores for a given person each time they are retested. The standard error of measurement (SE_M) estimates how repeated

measures of an individual on the same test are distributed around his "true" score. The true score is always unknown because no test can give a perfect reflection of the true score. The SE_M is a measure of absolute reliability of a test, the smaller the SEM, the greater the reliability and the larger the SE_M, the lower the reliability of the test.



CHECK YOUR PROGRESS

Q 4: Mention any two limitations of the test-retest method of estimating reliability.

.....

Q 5: What is inter-item consistency?

.....

7.6 RELATION BETWEEN RELIABILITY AND VALIDITY

The relation between validity and reliability is very close. Both are indispensable criteria for a good test or measuring instrument. Reliability and validity are the two different but interrelated dimensions of the same thing called test efficiency. Explaining the relation between reliability and validity, F.S. Freeman wrote, "The first necessary condition of a valid test is that it has an adequate degree of reliability. If the reliability coefficient of a test is 'zero', it cannot correlate with anything. A test that correlates poorly even with itself cannot correlate well with a measure of another variable."

The relationship between validity and reliability may be seen through the following points:

- Validity means truthfulness of a test; it indicates how well the purpose of the test is being served by the test. It is a measure of the extent to which a test measures what it purports to measure. On the other hand, reliability refers to the dependability or consistency or stability of the test scores. It is not directly concerned about the purpose of the test.

- While reliability is concerned with the consistency of the test scores or self-correlation of the test, validity is the measure of the correlation of the test with some external criterion.
- A valid test is always reliable, but a reliable test is not necessarily valid. A test with high correlation with itself (reliability) may not have any correlation with external criterion (validity).
- A test which is not reliable cannot be valid. If a test is not reliable, it implies that, it is not measuring the same thing every time it is used; if it is not measuring the same thing then it cannot be valid. From this perspective, reliability is the precursor of validity. However, reliability alone cannot ensure validity of a test. For example, a clock may consistently (reliable) give wrong time (invalid). But a valid test is always reliable. A test to be valid must consistently measure what it intends to measure.
- A test has maximum reliability when items in the test are homogeneous; on the other hand, a test is more valid when items are heterogeneous.
- Reliability of a test requires inter item consistency with uniform difficulty level; on the other hand validity requires items with varied difficulty level.

Though there are some differences between validity and reliability, significant relation between the two exists and both are equally important for a good test.



7.7 LET US SUM UP

After going depth into this unit, it can be summarized as-

- Any measuring tool used to collect data for measurement and evaluation in education must be reliable. Reliability refers to the consistency of test scores; it represents the extent of accuracy or precision of a measuring tool. If a test or other tool gives consistent results on testing and retesting in variable conditions then the tool is accepted to be reliable. Reliability is a precondition to the validity of

a test as well. However, reliability is the characteristic of the scores only and it is not the property of the test itself.

- Reliability of a test may be affected by various factors. Some of the important factors that affect reliability of a test include: length of the test, methods used to estimate reliability, nature of the group, too difficult or too easy items, lack of objectivity in scoring, lack of clarity of instruction to candidates, interval between testing and retesting, environmental conditions under which the test is administered, psycho-physical condition of the test-takers at the time of taking the test, guessing of correct answer by the respondents, etc.
- There are various methods of estimating reliability of test. The different methods of estimating reliability may be categorized under two different types. They are: (i) methods of estimating relative reliability, and (ii) methods of estimating absolute reliability. Relative reliability is a coefficient or correlation, which is known as reliability coefficient. Several methods are used to estimate relative reliability or reliability coefficient of a test. The most popular methods are:
  - the test-retest method of reliability
  - the equivalent or parallel form method of reliability
  - the split-half or the odd-even method of reliability
  - the method of rational equivalence or the Kuder-Richardson method of reliability
- The second type of methods estimating absolute reliability is known as Standard Error of Measurement. Standard error of measurement is actually an estimate of the deviation of a set of obtained scores from their true scores.
- Validity and reliability are two indispensable criteria for a good test or measuring instrument. The relation between validity and reliability is very close. They are the two different but interrelated dimensions of the same thing called test efficiency.

7.8 FURTHER READING

1) Anastasi, Anne (1982): Psychological testing, Edition 5, illustrated. Macmillan, The University of Michigan.
2) Ebel, R.L. and Frisbel, D.A. (1979): Essential of Educational Measurement, 3rd Edition, Prentice Hall of India, New Delhi.
3) Freeman, F.S. (1962): Theory and Practice of Psychological Testing, Holt, Rirehant & Winston, New York
4) Gupta, S.P. (2009): Modern Measurement and Evaluations with Statistics, Sharda Pustak Bhavan, Allahabad
5) Kelley, T. L. (1927): Interpretation of educational measurements, Macmillan, New York
6) Mehrens, W.A. and Lehmann, I.J. (1962): Standardized tests in Education, Holt, Rinehart and Winston, New York
7) Newton, Paul & Shaw, Stuart (2014): Validity in Educational and Psychological Assessment, Sage Publications, London.
8) Pradhan Hemant (1996): Essentials of Educational Measurement, Evaluation and Statistics, Saha Publishers and Distributers, Sambalpur
9) Stanley & Hopkins (1966): Education and Psychological Measurement and Evaluation, Prentice Hall Pvt. Ltd. Eaglewood Cliff, New Jersey

7.9 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Reliability refers to the consistency of test scores; it represents the extent of accuracy or precision of a measuring tool. If a test or other tool gives consistent results on testing and retesting in variable testing conditions then the tool is accepted to be reliable.

Ans to Q No 2: Two factors affecting the reliability of a test are: (i) environmental conditions under which the test is administered and (ii) psycho-physical condition of the test-takers at the time of taking the test.

Ans to Q No 3: The two types of reliability are: (i) relative reliability, and (ii) absolute reliability. Relative reliability is generally stated in terms of a coefficient or correlation, which is known as reliability coefficient. In this

method the correlation between the scores obtained from two different administration of the test is estimated and labelled as reliability coefficient. The second type of methods estimating absolute reliability is known as Standard Error of Measurement. Standard error of measurement is actually an estimate of the deviation of a set of obtained scores from their true scores.

Ans to Q No 4: The two limitations of the test-retest method of estimating reliability are: (i) test retest method requires two administration of the same test; if the interval between the two administration is short, immediate memory, habit, familiarity with the test items may raise the reliability coefficient to a very high level and on the other hand if the interval between the two administration is long, changes in behaviour and maturity of students may make the reliability coefficient low, (ii) if the conditions under which the two different administrations of the test held are different then that results in low reliability of the test.

Ans to Q No 5: Inter-item consistency means that every item in the test is highly correlated with all other items. This means that all the items in the test must measure the same variable or variables in the same proportion. This inter-item consistency is very important for calculating reliability of a test using the kuder-Richardson formula.



7.10 MODEL QUESTIONS

A. Short Questions (Answer each question in around 150 words)

Q 1: Define reliability.

Q 2: What is standard error of measurement?

Q 3: Explain the relation between reliability and validity.

B. Long Questions. (Answer each question in around 300-500 words)

Q 1: Discuss the factors that affect reliability of a test.

Q 2: Explain any two methods of estimating relative reliability of a test.

Q 3: Discuss the relative advantages and disadvantages of the different methods of estimating relative reliability of a test.

Q 4: Explain the process of estimating absolute reliability of a test.

*** ***** ***

UNIT 8: EDUCATIONAL ACHIEVEMENT AND ITS MEASUREMENT

UNIT STRUCTURE

8.1 Learning Objectives

8.2 Introduction

8.3 Meaning of educational achievement

8.4 Measurement of educational achievement

8.5 Objectives of Achievement Tests

8.6 Types of Achievement Tests

- 8.6.1 Standardized achievement tests
- 8.6.2 Teacher made achievement tests
- 8.6.3 Essay type, Objective type, and short answer type tests
- 8.6.4 Difference between Standardized tests and teacher-made tests

8.7 Let Us Sum Up

8.8 Further Reading

8.9 Answer to Check Your Progress

8.10 Model Questions

8.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- understand the meaning and objectives of achievement test
- know about the different types of standardized achievement test
- understand the meaning of teacher made achievement tests
- explain the procedure for teacher made test construction
- know about essay type, short answer type and objective type tests

8.2 INTRODUCTION

Educational achievement is an important area of concern for teachers, parents, policy planners, and educational administrators. In educational institutions, students are provided with various knowledge,

experience and skills based on the the goals and objectives of that specific education. Examinations are held at regular intervals to assess how much of these knowledge and experience have been acquired by the students or to assess if the students have been successful in achieving the goals of education. Hence the tests of educational achievement help to measure the level of achievement of students in the field of education. In this unit we are going to discuss about different types of achievement test and its applicability in the field of education.

8.3 MEANING OF EDUCATIONAL ACHIEVEMENT

Educational achievement is the degree to which a learner has achieved the educational goals and objectives. It refers to the achievement or performance of students in academic areas such as reading, writing, arithmetic, language arts, science and other disciplines as measured by achievement tests. According to Simpson and Weiner, achievement is defined as measurable behaviour in a standardised series of tests.

From a simpler and more practical point of view, educational achievement refers to the knowledge, attainment, understanding, skill, performance, etc. acquired by an individual during the process of learning. This means, all desirable changes in student behaviour as a result of learning may be called educational achievement. In order to assess the qualitative standard of scholastic and non-scholastic attainment of students, effectiveness of teaching and teaching methods, and overall standard of educational institution, regular evaluation is conducted. This evaluation process is closely associated with the objectives of education and the learning experiences in the form of a curriculum provided to the students to help them to achieve the objectives. Teachers impart knowledge on the basis of a curriculum to help students achieve the objectives of education and instruction. The extent of success of the teachers and students in this task of shaping behaviour according to the goals of education is their educational achievement.

At this point it is important to make a distinction between

achievement and other similar concepts like aptitude and intelligence. An aptitude is the ability to learn or to develop proficiency in an area when provided with appropriate education or training; it is a natural skill, talent, or capacity of an individual to learn. Aptitude is oriented towards future. While aptitude is the potential to learn, achievement is learning itself. It is the amount or extent of the knowledge or skills that an individual has already learned. Its orientation is towards the past. Aptitude is an ability to learn specific things with training but intelligence is a general mental ability to reason, solve problems, think abstractly, comprehend ideas and language, and learn; which is not a product of training or learning. The orientation of intelligence is towards the present.

8.4 MEASUREMENT OF EDUCATIONAL ACHIEVEMENT

Although the measurement of achievement is complex due to the intangibility and lack of agreement about the concept of achievement itself, and also due to the relative imprecision of many of the techniques of educational measurement, psychologists believe that educational achievements can be measured. Educational achievement in the true sense usually leads to difference in behaviour of individuals. The person who has achieved more behaves differently from the person who has achieved less. Measurement of achievement is the process of finding out such differences. Many important educational achievements can be measured satisfactorily through examinations and tests of various types that claim to measure knowledge, skills, facts and other qualities that education aims to impart and develop.

According to **Bruce and Neville**, educational achievement is measured by standardised achievement test developed for school subjects. What this means is that academic achievement is measured in relation to what is attained at the end of a course, since it is the accomplishment of medium or long term objective of education. What is important is that the test should be a standardised test to meet national norms. For a test to be standardised, it must be valid for over a period of time.

According to **N.M. Downie**, "Any test that measures the attainments and accomplishments of an individual after a period of training or learning".

Thorndike and Hagen define an achievement test as, "The type of ability test that describes what a person has learned to do".

Groulund stated that achievement is "a systematic procedure for determining the amount a student has learned through instructions"

Simpson and Weiner asserted that achievement test intends to measure systematic education and training in school occupation towards a conventionally accepted pattern of skills or knowledge.

8.5 OBJECTIVES OF ACHIEVEMENT TESTS

The main objective of achievement test is to determine the extent of success of students in achieving the knowledge, skills or the experiences that were provided to them. Achievement tests aim to measure intellectual competence of individuals in different areas. It measures what and how much an individual has learned or attained skills. Schools regularly use achievement tests to measure the attainment of students. Some of the important objectives of achievement test may be mentioned as under:

- One of the main objectives of achievement test is to determine the extent to which educational objectives have been attained.
- Achievement test aims to measure the effects of education and training on the behaviour of the students in terms of developing conventionally accepted pattern of skills or knowledge.
- Achievement test aims to assess how much an individual know about certain topics or how well they can perform certain skills. Moreover, achievement test help in selecting individuals for entry into Universities, for certification, and the like.
- One important objective of achievement test is to provide basis for promotion or retention of students and grading them.
- Achievement test aims to find out the relative position of students with regard to various academic areas.

- Helping the selection of students for their proper placement in particular course or career is another objective of achievement test.
- Motivating the students is an important objective of achievement test.
- Diagnosing the strengths and weaknesses of students in the field of learning and providing them necessary remedial guidance and support is another objective of achievement test.
- Another objective of achievement test is to inform the student and his parents about his academic attainments.
- Providing teachers and school administrators with information about necessary reform of the curriculum is another objective of achievement test.
- Another objective of achievement test is to evaluate teaching effectiveness besides measuring the progress of students in the course.
- One important objective of achievement test is to assist summative as well as formative evaluation of students.



CHECK YOUR PROGRESS

Q 1: What is achievement test?

.....

Q 2: What is the purpose of achievement testing?

.....

8.6 TYPES OF ACHIEVEMENT TESTS

The field of psychological testing is full of numerous types of achievement tests. Depending on the purpose, standardization, administration, functions and other such criteria, achievement tests have been classified differently. However, all achievement tests can be broadly categorized into standardized tests and teacher-made tests. A simple classification of these tests may be presented as below:

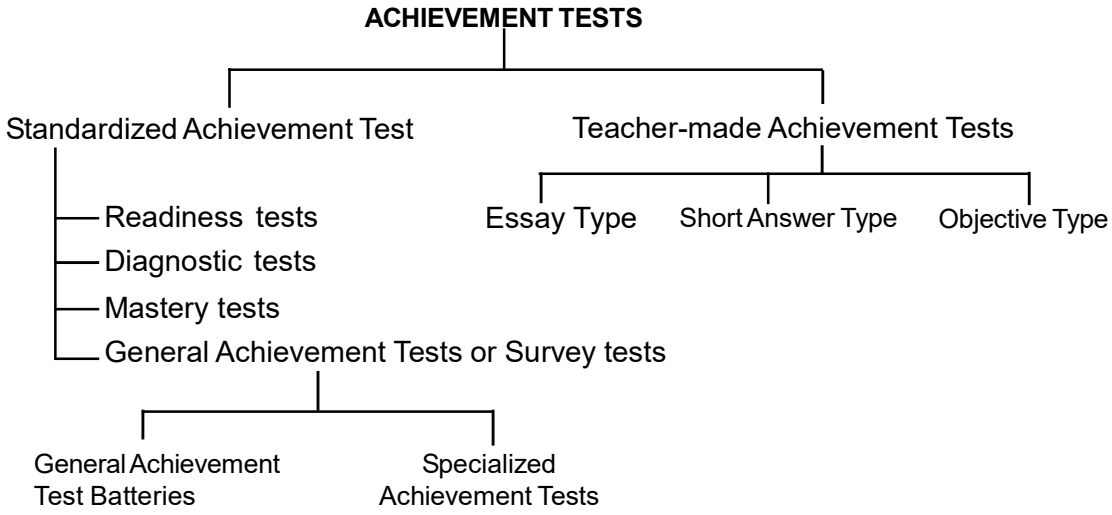


Figure 8.1: Classification of achievement Test

8.6.1 Standardized Achievement Test

Although the purpose of the two types of achievement tests, viz-standardized tests and teacher-made tests is the same, they have differences in terms of their procedures of construction and use. Standardized achievement tests are constructed by experts. A standardized achievement test is administered, scored, and interpreted in the same way for all students and in all conditions of testing. A standardized test establishes uniformity of procedures for administration and scoring of the test. This uniformity of procedure makes the scores obtained by different persons comparable.

According to **L. J. Cronbach**, a standardized test is one in which procedure, apparatus, and scoring has been fixed so that precisely the same test can be given at different times and places. According to **Newkirk and Greene**, "A test is standardized (i) if it is composed of exercises that have been selected in the light of usual teaching practice and evaluated as to innate difficulty and (ii) if it is accompanied by norms or standards permitting the interpretation of results in levels of accomplishment".

The main characteristics of a standardized test are as follows:

- A standardized test is administered under identical conditions, it does not matter where, when, by whom or to whom the test is administered, it is done in a similar way.
- A standardized test is always scored objectively and in a standard or consistent manner; the procedures for scoring a standardized test are specified so that different scorers arrive at the same scores for the same set of responses,
- A standardized test requires all test takers to answer the same questions in the same way,
- Scoring of standardized achievement test is objective, standard and based on well prepared scoring key.
- Scores obtained by different individuals or groups in standardized test can be compared in order to assess the relative performance of individual students or groups of students.
- Standardized tests are usually designed to measure relative performance of individuals or groups. Hence establishment of norms for comparing the performance of individuals or groups form an important part of a standardized test construction.
- The construction of standardized achievement test is systematic and methodical which passes through definite steps and follows certain principles.
- Item construction for standardized achievement test passed though the stages of item analysis.
- The formulation of directions for a standardized test is a major part of its construction. Directions for standardized tests specify the materials to be used, time limits to be given, oral instructions and demonstrations to testees, and all other details of the testing situation.
- Standardized tests are highly reliable and valid.

(The principles and method of standardized test construction has been discussed in UNIT IV)

Here, we are going to discuss about the types of standardized achievement test in detail-

Readiness tests: A readiness test, a type of prognostic test, is used to predict a student's future success in a particular learning activity. The test measures the student's physical, mental, or emotional readiness or maturity to handle formal instruction. It is constructed and used to predict how ready or likely an individual is to benefit from instruction in a particular field, i.e. reading or mathematics. These tests are constructed to study the developmental conditions of an individual to determine whether or not, or to what extent, the individual could learn particular skills or acquire particular knowledge and experience. In educational situations, readiness tests measure cognitive, affective, and psycho-motor and such other factors that are involved in learning, in order to determine if a student has the necessary background or readiness to gain from a particular course of instruction. Wrightstone's Prognostic Test of Mechanical Abilities, Hoke's Prognosis Test of Stenographic Ability etc are examples of some famous readiness test.

Diagnostic tests: Diagnostic tests assess students' understanding of a subject, area or skill. Usually diagnostic tests are administered before giving instructions. The purpose of these tests is to determine what students already know about the content that would be covered by instruction. These tests help to identify learning difficulties of students in different subjects and discover their preferred learning styles and their strengths and weaknesses. Diagnostic tests help in assessing a student's understanding of different subjects. The findings of diagnostic tests can be used to provide remedial instruction. Thus diagnostic tests help formative assessment of students and provide directions to teachers about the direction they need to take in their instruction. Examples of diagnostic tests include: Schonnel's Diagnostic Test in Arithmetic, Compass Diagnostic Test in Arithmetic, Durrell Analysis of Reading Difficulty, Gates Basic Reading Test etc. It is important to note that the diagnostic tests can diagnose or identify the weaknesses but cannot suggest remedial measures for their removal.

Mastery tests: Another important type of achievement test is mastery test or proficiency test. These tests compare the mastery or proficiency of students in a particular area with an established standard of mastery. Mastery tests measure the minimum level of mastery required to achieve success in a particular field. Unlike other tests that use a 101 point scale (ranging from 0 to 100) to assess students' performance or achievement in a particular area, a mastery test assesses the students' performance as all or nothing: either the individual has mastered or has not mastered the tested area at a specified level of performance. For example, a mastery test on arithmetic measures whether or not an individual has mastery, based on an established set of criteria. If a testee meets those criteria, he is passed, otherwise he is failed. Missouri Mastery and Achievement Tests is an example of a mastery test by which students' mastery over school subjects is measured.

General Achievement Tests or Survey Tests: General Achievement Tests are most widely used test in the field of achievement testing. A general achievement or a survey test measures the performance of an individual in a single score and does not measure or discover specific strengths or weaknesses. This general achievement test or survey test is a group of subject-matter tests designed for particular grade levels. The major purpose of administering such a test is to determine an individual's general standing in a group rather than his specific strengths and weakness. These tests are generally based on educational objectives and are very wide in scope. These tests are designed to survey or measure achievement in different areas such as language, mathematics, spelling, social science, etc. This is the most comprehensive way of assessing achievement. Such tests give an overall picture of an examinees standing in various subjects.

The general achievement or a survey tests may be categorized into two types: **(i) General Achievement Test Batteries, and (ii) Specialized Achievement Tests.**

(i)General Achievement Test Batteries: A test battery refers to a set of tests grouped together and administered to individuals. A general achievement test battery is a group of subject-matter tests designed for particular grade levels. The major purpose of administering a battery of test is to determine an individual's general standing in a group rather than his specific strengths and weakness. General achievement test batteries are normally used to determine general standing of individuals in group. The battery is a group of subject area tests, usually containing a limited number of questions with in each subject area. A few examples of such test batteries are as follows:

Metropolitan Achievement Tests: This battery of tests was published in 1956. The test is spread from grade-1 to grade-9. The time required for administration of the battery ranges from 2hours to 41/2 hours. There are a total of 15 individual tests in the battery meant for different grades. They are:

- Word Knowledge test: applicable to all the grades
- Word Discrimination test: for 1st to 4th grades
- Reading test: applicable to all the grades
- Test of Arithmetic concept and skills: for 1st and 2nd grade
- Test of problem solving ability: for 3rd to 9th grade
- Test of computation: for 3rd to 9th grade
- Test of spelling: applicable to all the grades
- Language use test: for 7th to 9th grade
- Written language test: for 3rd to 9th grade
- Test of using different kinds of sentences: for 7th to 9th grade
- Grammar test: for 5th to 9th grade
- Language study skill: applicable to all the grades
- Social science knowledge test: for 5th to 9th grade
- Test on social science skills: for 5th to 9th grade
- Test for science: for 5th to 9th grade

Sequential Test of Educational Progress (STEP): Educational Testing Service (USA), the world's largest private

nonprofit educational testing and assessment organization developed the Sequential Test of Educational Progress (STEP). These test batteries were developed for four different stages-

Stage-1 - grade 4 to 6

Stage-2 - grade 7 to 9

Stage-3 - grade 10 to 12

Stage-4 - grade 13 to 14

This test battery consists of seven tests, namely:

1. Reading ability
2. Writing ability
3. Mathematical achievement
4. Science
5. Social studies
6. Listening
7. Essay writing.

Excluding the last test all the other tests are objective type tests. The total time required to administer the test is 7hours and 35 minutes. But these are administered separately and the scores obtained in the different tests are converted into a standard score to make a comprehensive evaluation of students.

Stanford Achievement Test: The Stanford Achievement Test is similar to the Metropolitan Achievement Test and was published for the first in 1923 and was revised in 1953. This test consists of four batteries. They are-

1. Primary - a total of five tests for 2nd and 3rd grade
2. Elementary - a total of six tests for 3rd and 4th grade
3. Intermediate - a total of nine tests for 5th to 7th grade
4. Advanced-a total of nine tests for 7th to 9th grade

Time for administration of the tests for different grades is 85 minutes, 145 minutes, 232 minutes and 227 minutes respectively. The various tests include the following fields:

Area	Tests
Language	Reading, Meaning of Paragraph, Word –Meaning, Spelling and Language
Arithmetic	Arithmetic, Counting, and Logic
Social Studies	Social Studies
Science	Science
Study Skills	Study Skills

(ii) Specialized Achievement Tests: Specialized achievement tests are designed to measure achievement in a single academic area (e.g. reading, writing etc) and different test are prepared for different grades. Rather than measuring the general or overall achievement of individuals; specialized achievement tests measure proficiency of an individual in a single subject or area. Two examples of specialized achievement tests have been presented below:

Iowa Reading Test: This is a test for measuring reading ability. This test consists of seven sub-tests; they are:

1. Test of measuring rate of reading,
2. Comprehension of prose after reading,
3. Test on poetry comprehension,
4. Test on vocabulary,
5. Test of sentence meaning,
6. Test of paragraph comprehension,
7. Test on finding location of information.

Thorndike Handwriting Scale for Children: This test of handwriting was first published in 1910. The test was constructed only for measuring handwriting of children using large samples of handwriting of children. Experienced evaluators separated these handwritings on the basis of general merit. Another two similar tests are: The Ayres handwriting scales, and Conard Manuscript writing Standards.

8.6.2 Teacher-made Achievement Test

As suggested by the name itself, teacher-made tests are prepared locally by classroom teachers to test educational achievement of their students. Sometimes these tests are also constructed to determine the effectiveness of teaching and learning. For constructing such tests the teacher must have mastery over subject matter besides accurate and thorough knowledge about educational objectives. These tests are not standardized tests. Many people say that standardized tests are more useful for achievement testing since they are more valid and more reliable than the teacher made tests. But at the same time it must also be remembered that the teacher cannot always depend upon standardized tests. Many a times standardized tests fail to serve local and immediate needs or purpose and they are costly and not easily available. This is why classroom teachers need to construct their own test to meet the local needs and purpose of testing. Teacher-made tests are generally constructed only through two steps: the step of planning and the step of preparing the test. The main characteristics of a teacher made test are:

- Teacher-made tests are constructed by classroom teachers to meet immediate and local needs.
- They may be used as diagnostic or prognostic tests.
- Steps involved in standardized tests are not followed.
- Less valid and less reliable than standardized tests.
- The construction and administration is easy and economic.
- Scope of teacher-made test is limited and cannot be used with any other group than the one for which it was constructed.

Steps involved in the construction of teacher-made test:

As already mentioned above, teacher-made tests are constructed following only two stages: Planning and preparing the test.

Planning for a teacher made test: At this stage the length and organisation of the items is decided keeping in view the objectives of testing. After this a blue print or test-specification is prepared including the instructional objectives, weightage to different areas of the content, the type of questions or tasks to match the instructional objectives, and the learning domains and levels therein. Constructing a test is a complex task and it can be successfully achieved only with the cooperation from experts, colleagues and other experienced persons. Therefore, at this stage the teacher should consult the concerned persons about the test and collect their views and suggestions. Moreover, the teacher should have good knowledge about the construction and use of different types of test items.

Preparing the test: At this stage items are constructed and are repeatedly studied to analyze properly in order to ensure that the selected items are correct and appropriate. Provisions are made to effect necessary modifications of items and to delete items that are defective. At the same time, it also must be kept in mind that the number and type of items are retained in the test as decided at the time of preparing the blueprint. After selecting the items for the test, they are arranged in ascending order of difficulty and a set of clear instructions as to how to respond to the items is prepared and attached. Moreover, the time stipulated for the test, the procedure for scoring is decided and attached with the test. Thus, a teacher-made test is constructed.

Limitations of teacher made test:

- Constructing a test by a classroom teacher is time consuming and costly.
- Standardization, item analyses etc. scientific procedures are not followed by these tests.
- Every teacher cannot construct a good or useful test.
- These tests are less valid, less reliable and are subjective.

- These tests can fulfil local needs only.
- These tests cannot maintain a uniform standard of testing.

8.6.3 Essay type, Objective type, and Short answer type tests

Teacher-made tests may be of different types depending on the content and different objectives of testing. The major types are discussed below:

Essay type tests: Essay type tests are very popular and traditional tests and answer to these questions resembles long essays. These questions can test the mental abilities like explanation, illustration, presentation, analysis, comparison, justification, synthesis etc. These types of questions are very useful for assessing the educational progress and depth of knowledge of the students. However, these questions are influenced by subjectivity though they are easy to construct and administer. These types of tests are alleged to be less valid, less reliable and subjective.

Strengths of essay type tests:

- Essay type tests an test can test the language skills of students.
- Essay type tests can measure written communication, expression, and understanding.
- Essay type tests can appropriately test mental abilities like explanation, illustration, presentation, analysis, comparison, justification, synthesis etc.
- Construction of essay type test is simple and economic.
- Administration of essay type test is very easy.
- Essay type tests may develop the habit of deeper study among the students.

Limitations of essay type test:

- Essay type tests cannot represent all or the large part of the content since only a few questions can be asked due to their length.

- Examinees may write irrelevant things since the scope of the answer is not clear.
- The validity, reliability and objectivity of essay type tests are doubtful.
- Different examiners or the same examiner at different times may award different marks on the same answer script.
- Essay type tests are time consuming.
- Essay type tests measure the power of memory rather than testing their knowledge and understanding.
- Essay type tests reduce the speed of thinking to the speed of writing of the examinees.

Short answer type tests: These are fairly structured items that require brief answers, usually within few words or sentences. These items can represent wide area of topic. Unlike essay or objective tests, they are neither too long nor too short. During the current times these tests have become very popular. These questions can be highly thought provoking and these questions can be made very specific about the subject to be asked.

Strengths of short answer type test:

- Short answer type tests can represent very wide area of the content.
- These tests are more valid and reliable than essay type tests.
- Short answer type tests objective and the influence of subjectivity is less.
- Examinees cannot write irrelevant answers since the scope and length of the answers are specified.
- The construction, administration and scoring of short answer type test is easy.
- Short answer type test is a synthesis of the essay type and objective type tests.
- Short answer type tests are economic.

Limitations of short answer type tests:

- Short answer type tests restrict the scope of the topic to be discussed.

- Short answer type tests indicate the nature of the desired response and thereby limit originality.

OBJECTIVE TYPE TESTS: These types of tests require short answers which usually consist of one word, a phrase or a sentence. This type of test items consist of factual questions that require very short answers that can be scored quickly and unambiguously by anyone with a scoring key. These tests are called objective since their scoring procedure is objective. Objectivity is the greatest strength of these tests.

Objective type tests can mainly be categorized into two types: Recall type or supply type and Recognition type or selection type. Answers to the recall type of questions must be supplied from examinee's memory, for which these items are also called supply type of items. In recognition type items, the examinees have to select one from the available alternatives as answer to the question; this is why this type of items are called selection type items.

The recall type items may further be divided into two types: simple recall and completion type items. A simple recall type item is given in the form of a direct question, which is presented as a simple interrogative sentence, answer to which must be given in a single word or a short sentence. Examinees have to supply the answer to such questions from their memory. Two examples of simple recall type item have been given below:

Example: Q1. Who was the first President of India?

Example: Q2. From which language the term "Educare" has been derived?

The completion type item is given as a sentence from which an important word or phrase is omitted. The examinee has to fill the blank by supplying the appropriate word or phrase missing in the sentence. Two examples have been given below:

Example: Q1. The capital of India is _____.

Example: Q2. Hero worship is an important characteristic of _____stage.

The retention type of items may be of different types. Some of the important ones are discussed below:

One of the most popular and widely used tests is the Alternate response type item. These items are given in the form of a sentence or statement to which the respondent has to reply by "yes or no", "true or false" or by putting "" or ×".

Example: Q1. Psychology is the science of soul. (True/False)

Example: Q2. Education is a lifelong process. (Yes/No)

Multiple choice types are another popular, widely used and superior type of objective items. A multiple choice item is usually a statement followed by four or five alternative responses from which only the best or correct one must be selected. The statement or question is called the 'stem' and the alternatives or choices are the 'options' and the correct alternative is called the 'key. Incorrect options or alternatives are called 'distracters'.

Example: The name of the capital city of India is -

(a) Mumbai (b) New Delhi

(c) Kolkata (d) Chennai

Matching is another retention type of objective item. This type of items requires linking contents of one list to contents in another list. The test items are presented in the form of two columns of interrelated contents and ask the testee to link or connect the contents.

Example:

Column A	Column B
Hunter Commission	1854
Kothari Commission	1952
Radhakrishnan Commission	1882
Mudaliar Commission	1964
Wood's Despatch	1948

Classification is another retention type of objective items in which some words, phrase, names of objects, or some information is presented, among which one is odd or not related to the others. The examinees have to select the unrelated or odd item out.

Example: 1. Delhi, Mumbai, Rajasthan, Chennai

Example: 2. Brahmaputra, Yamuna, Allahabad, Kaveri

Arrangement is still another retention type of objective items in which some event, information, or steps of some task are given randomly which the examinee has to arrange in their logical or chronological sequence.

Example: September, February, December, July

Strengths of objective type test:

- Objective type items can be used to measure all the objectives of teaching.
- Items are clear, direct and unambiguous.
- More valid, reliable and objective.
- Scoring is fast and objective, which can be done by machines as well.
- Students cannot write irrelevant answers.
- Administration is easy.
- Objective type items can cover the whole content or curriculum.

Limitations of objective type test:

- Construction of objective type items is difficult and costly.
- Measures factual information, not understanding.
- Examinees may guess the correct answers.
- Objective items cannot develop the interest and habit of in depth study.
- Linguistic skills cannot be tested.
- Possibility of examinees using unfair means is higher.
- Higher mental abilities such as explanation, presentation, comparison, description etc cannot be tested.



CHECK YOUR PROGRESS

Q 3: What is diagnostic test?

.....

Q 4: What is a teacher-made achievement test?

.....

Q 5: What is objective type test?

.....

8.6.4 Difference between standardized tests and teacher-made tests

Though the main purposes of standardized and teacher-made achievement tests are same, they have significant differences in their construction, administration and scoring procedure. Some of such important differences are as follows:

Standardized tests	Teacher-made tests
Constructed by test experts	Constructed by classroom teachers
Standardized on a very large sample	Not standardized; prepared only for a single class of students
Only objective items are included	Objective as well as subjective items may be included
Passes through four stages—planning, preparation, try out and evaluation	Passes only through two stages—planning and preparation
Validity and reliability indices are estimated	Instead of estimating validity and reliability more emphasis is laid on covering the content
Item analysis done for selecting items	No analysis of items is done
Norms are developed for interpreting the scores	Norms are not fixed and hence interpretation remains incomplete
Administration, scoring and interpretation is uniform and standard	No uniformity in administration, scoring and interpretation
Construction is complex and costly	Construction is easy and economic

May not serve local and immediate classroom purposes	Constructed to serve local and immediate needs of the classroom
Can be used to compare the achievement of two different groups	Cannot be used to compare since the test is not standardized
Used mainly in research, instruction, guidance, selection and administrative works.	Used to measure achievement of students and institutions

According to Ross, there are four essential aspects in which a standardized test differs from a teacher-made test:

- The content is standardized, i.e., item-selection has been done vigorously, after careful scrutiny and by competent judges.
- Administration is standardized, i.e., directions, time-limit, etc., are worked out carefully.
- Scoring has been standardized, i.e., scoring keys are prepared, definite rules for scoring have been formulated etc.
- Interpretation has been standardized, i.e., norms for various groups are provided.



8.7 LET US SUM UP

The unit can be summarized in the following ways-

- Educational achievement refers to the knowledge, attainment, understanding, skill, performance, etc. acquired by an individual during the process of learning. This means, all desirable changes in student behaviour as a result of learning may be termed as educational achievement.
- Educational achievement is measured by standardised achievement test developed for school subjects. An achievement is a systematic procedure for determining the amount a student has learned through instructions. Moreover, many important educational achievements can be measured through examinations and tests of various types made by teachers.
- The main objective of achievement test is to determine the extent of success of the students in achieving the knowledge, skills or the

experiences that were provided to them. It measures what and how much an individual has learned or attained skills.

- Though achievement tests have been classified differently, all achievement tests can be broadly categorized into standardized tests and teacher-made tests.
- A standardized achievement test is a test that is administered, scored, and interpreted in the same way for all students and in all conditions of testing. A standardized test establishes uniformity of procedures for administration and scoring of the test.
- Standardized achievement tests include Readiness test, Diagnostic tests, Mastery test and General Achievement Tests. The general achievement tests may again be categorized into two types: (i) General Achievement Test Batteries, and (ii) Specialized Achievement Tests.
- Teacher made tests are prepared locally by classroom teachers to test educational achievement of their students. These tests may be of different types such as Essay type, Short answer type and objective type tests.
- Though the main purposes of standardized and teacher made achievement tests are same they differ with regard to their content, administration, scoring and interpretation.



8.8 FURTHER READING

1) Bloom, B.S. (1956): Taxonomy of Educational Objectives. Handbook I, Cognitive Domain, David McKay Co. Inc, New York
2) Bloom, B.S. (1976): Human characteristics and school learning, McGraw-Hill, New York
3) Gupta, S.P. (2009): Modern Measurement and Evaluations with Statistics, Sharda Pustak Bhavan, Allahabad
4) Kelley, T. L. (1927): Interpretation of educational measurements, Macmillan, New York
5) Pradhan Hemant (1996): Essentials of Educational Measurement, Evaluation and Statistics, Saha Publishers and Distributers, Sambalpur

6) Stanley & Hopkins (1966): Education and Psychological Measurement and Evaluation, Prentice Hall Pvt. Ltd. Eaglewood Cliff, New Jersey

8.9 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: An achievement test is a test that measures the attainments and accomplishments of an individual after a period of training or learning.

Ans to Q No 2: The main purpose of achievement test is to determine the extent of success of the students in achieving the knowledge, skills or the experiences that were provided to them.

Ans to Q No 3: A diagnostic test is one that assesses students' understanding of a subject, area or skill along with their specific strengths and weakness.

Ans to Q No 4: A teacher made achievement test is one that is prepared locally by a classroom teacher to meet the needs and purpose of testing educational achievement of his or her students.

Ans to Q No 5: An objective type test is one with items requiring short answers of one word, a phrase or a sentence. These items consist of factual information that can be scored objectively, quickly and unambiguously.

8.10 MODEL QUESTIONS

A. Short Questions (Answer each question in around 150 words)

Q 1: What is educational achievement?

Q 2: How educational achievement can be measured?

Q 3: Define standardized achievement test.

B. Long Questions. (Answer each question in around 300-500 words)

Q 1: Discuss about the different types of standardized achievement tests.

Q 2: Explain the procedure of teacher made achievement tests.

Q 3: Discuss about the relative merits and demerits of objective type and essay type test items.

Q 4: Distinguish between standardized tests and teacher made tests.

*** ***** ***

UNIT 9: INTELLIGENCE TEST AND ITS MEASUREMENT

UNIT STRUCTURE

9.1 Learning Objectives
9.2 Introduction
9.3 Meaning and definitions of Intelligence
9.4 A brief overview of the theories of Intelligence
9.5 Measurement of Intelligence
9.5.1 Types of Intelligence Tests
9.5.2 Uses of Intelligence Test
9.6 Construction of an Intelligence Test
9.7 Let Us Sum Up
9.8 Further Reading
9.9 Answer to Check Your Progress
9.10 Model Questions

9.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- understand the meaning and nature of Intelligence
- explain the different theories of Intelligence
- analyse the different types of Intelligence tests
- understand the uses of Intelligence tests
- explain the procedure of Intelligence test construction

9.2 INTRODUCTION

Intelligence is a unique mental faculty that is the source or basis of all other mental abilities of an individual. Success of an individual in any activity or his work efficiency depends largely on intelligence. It is an established fact that the level of intelligence varies from individual to individual. Intelligence is often used as a basis for evaluating the proficiency of an individual required for a particular job. In education, intelligence is the

most important factor causing differences among individuals in terms of their achievement. Therefore, measurement of intelligence is important not only for the psychologists but also for common people. Keeping this importance in view, the field of psychology has been full of researches conducted to understand the nature of intelligence and to develop means and tools to measure intelligence accurately. In this unit a complete picture of measurement of intelligence and its various types of test will be depicted.

9.3 MEANING AND DEFINITIONS OF INTELLIGENCE

Intelligence is a wide-ranging general mental ability which affects all the psycho-physical structure, strength and ability of an individual. This mental faculty is relatively permanent which determines, controls, and directs all mental functioning and behaviour of an individual. Intelligence is the main source or factor of individual difference. Alfred Binet, who conducted extensive research on the nature and measurement of intelligence, wrote, "It seems to us that in intelligence there is a fundamental faculty, the alteration or the lack of which, is of the utmost importance for practical life. This faculty is judgement, otherwise called good sense, practical sense, initiative, the faculty of adapting one's self to circumstances."

Intelligence may be understood as a capacity to adjust with varying and new situations in life and as a capacity to solve problems that people face in every walk of their lives. It is the ability for abstract thinking and learning. Intelligence helps in making satisfactory adjustment with newer needs, conditions or environments. Intelligence may also be understood as a set of unique characteristics in the psycho-physical make-up of an individual that determine and find expression in the behaviour, thinking pattern, habits, and character of an individual.

However, psychologists differ widely in terms of their definition and explanation of intelligence and hence there is no single definition of intelligence accepted by all. Presented below are a few important definitions of intelligence:

According to Anastasi, "Intelligence is not a single, unitary ability,

but rather a composite of several functions. The term denotes that combination of abilities required for survival and advancement within a particular culture."

M. Anderson views intelligence as". . . that facet of mind underlying our capacity to think, to solve novel problems, to reason and to have knowledge of the world."

W. V. Bingham says, "We shall use the term 'intelligence' to mean the ability of an organism to solve new problems . . ."

According to S. S. Colvin, "A person possesses intelligence insofar as he has learned, or can learn, to adjust himself to his environment."

W. Freeman asserts that intelligence is, ". . . adjustment or adaptation of the individual to his total environment, or limited aspects thereof . . . the capacity to reorganize one's behaviour patterns so as to act more effectively and more appropriately in novel situations . . . the ability to learn . . . the extent to which a person is educable . . . the ability to carry on abstract thinking . . . the effective use of concepts and symbols in dealing with a problem to be solved . . ."

H. Gardner, "Intelligence is the ability to solve problems, or to create products, that are valued within one or more cultural settings."

According to Stoddard, intelligence is ". . . the ability to undertake activities that are characterized by (1) difficulty, (2) complexity, (3) abstraction, (4) economy, (5) adaptiveness to a goal, (6) social value, and (7) the emergence of originals, and to maintain such activities under conditions that demand a concentration of energy and a resistance to emotional forces."

Defining intelligence in comprehensive terms, D. Wechsler wrote, "Intelligence is the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment."

E. Boring, keeping in view the variety of perspectives in defining intelligence wrote, "Intelligence is what is measured by intelligence tests."

The various definitions of intelligence show that there is no uniformity among psychologists in terms of defining the nature of intelligence. However,

an analysis of the definitions leads us to the conclusion that intelligence is a combination of:

- mental or cognitive capacities
- the ability to adjust with new situations
- the ability to learn
- ability to carry on abstract thinking
- ability to solve problems
- innate as well as acquired capacities

9.4 A BRIEF OVERVIEW OF THE THEORIES OF INTELLIGENCE

There are numerous theories of intelligence, but there is hardly any agreement among the theories about the nature of intelligence. Every theory explains intelligence from a different perspective and often contradicting the views of another theory. We present here a very brief overview of the important theories of intelligence:

Faculty theory: This is the oldest theory explaining the nature of intelligence that was very popular during 18th and 19th century. This theory posits that the mind is composed of different faculties like thinking, memory, reasoning, imagination, etc. These faculties are independent of each other and can be trained and sharpened by rigorous exercises.

Spearman's two-factor theory: Charles Spearman developed this theory in 1904. He proposed that intellectual abilities were comprised of two factors: one general ability or common ability known as 'G' factor and the other a group of specific abilities known as 'S' factor. 'G' factor is universal and inborn ability. Greater the 'G' in an individual, greater success he has in his life. 'S' factor is not inborn but acquired and it differs from one activity to another in the same individual.

Multifactor theory of Thorndike: Thorndike proposed the multifactor theory of intelligence and said that there was nothing like General ability. He proposed that every mental activity requires an aggregate of different set of abilities. He identified four main attributes of intelligence which were following:

- **Level**-refers to the level of difficulty of a task that can be solved.
- **Range**-refers to a number of tasks at any given degree of difficulty.
- **Area**-means the total number of situations at each level to which the individual is able to respond.
- **Speed**-is the rapidity with which we can respond to the items.

Theory of Primary mental abilities or the Group factor theory: This theory holds that intelligence is neither composed of innumerable highly specific factors, nor is an expression of a general factor that pervades all mental activities. Instead, intelligence is composed of a number of groups of mental abilities. According to this theory, certain mental operations have in common a 'primary' factor that gives them psychological and functional unity and that differentiates them from other mental operations. These mental operations then constitute a group. There are a number of groups of mental abilities, each of which has its own primary factor, giving the group a functional unity and cohesiveness. Each of these primary factors is relatively independent of the others.

Thurstone has given a list of six primary factors:

- The Number Factor (N) -ability to do Numerical Calculations rapidly and accurately.
- The Verbal Factor (V) - required by tests involving Verbal Comprehension.
- The Space Factor (S) - involved in any task in which the subject manipulates the imaginary object in space.
- Memory (M) - Involving ability to memorize quickly.
- Word Fluency (W)-needed to think of isolated words at a rapid rate.
- Reasoning (R) - needed in tasks that require to discover a rule or principle involved in a series or groups of letters.

Guilford's model of structure of intellect: Guilford proposed a three dimensional structure of intellect model. According to Guilford each task involving intellectual capacity can be analysed according to: (1) content, (2) the mental operation involved and (3) the product resulting from the operation. He also classified content into five categories, namely--Visual, Auditory, Symbolic, Semantic and Behavioural. Similarly, he classified

operations into five categories, namely, Cognition, Memory retention, Memory recording, Divergent production, Convergent production and evaluation. And finally, products were classified into six categories, namely, Units, Classes, Relations, Systems, Transformations and Implications.

Hierarchical Theory of intelligence: Cyril Burt and Vernon put forward this theory of intelligence. They organized the factors constituting intelligence in a hierarchic order. At the top of the hierarchy, the 'G' factor or the general cognitive factor is placed. At the next level two broad group factors, corresponding to verbal-educational (v: ed) and practical-mechanical (k: m.) aptitudes are placed. These major factors may be further sub-divided. The verbal educational factor may be sub-divided into verbal and numerical sub-factors, and the practical mechanical factor into mechanical information, spatial, and psychomotor ability. At the lowest level of the hierarchy are the special factors. This hierarchical arrangement resembles an inverted genealogical tree, with 'g' at the top, 's' factors at the bottom, and increasingly narrower group factors in between. Vernon's theory is a compromise between Spearman's two factors and Thurston's multifactor theory.

Cattell's Theory of Fluid and Crystallized Intelligence: R.B. Cattell theorized that intelligence is of two types- Fluid intelligence, which is the product of heredity and crystallized intelligence, which is the product of environment. Fluid intelligence is general to many fields, and crystallized intelligence is specific to certain fields, like school learning. Fluid intelligence is used more in tasks requiring adaptation to new situations, while crystallized intelligence is used to tasks where habits have become fixed. These two factors are distinct but correlated.

Gardener's theory of multiple intelligence: Howard Gardner put forward a new and different view of human intelligence. His theory of Multiple Intelligences proposed that intelligent behaviour does not arise from a single unitary quality or factor of the mind. Rather, different kinds of intelligence are generated from separate metaphorical pools of mental energy. Each of these pools enables the individual to solve problems, or to create products, that are valued within one or more cultural settings. The eight intelligences

proposed by the theory are: linguistic, logical, spatial, musical, motor ability, interpersonal, intrapersonal and naturalistic intelligence.

9.5 MEASUREMENT OF INTELLIGENCE

Owing to the intangible and abstract nature of the concept of intelligence, it cannot be measured physically or directly. Modern psychologists believe that intelligence can be assessed only when it gets manifested in the behaviour of man. That is why intelligence tests have been constructed and used extensively to measure intelligence. Intelligence tests give an "intelligence quotient" or IQ as a measure of intelligence. However, people tried to measure intelligence using different means since very long.

Psychologists first tried to measure mental ability in the late 19th century. However, no formal measures of intelligence existed at that time and hence the people were evaluated on the basis their fame, judged by encyclopaedia entries, honours, awards, etc. It was thought that intelligence was hereditary. Later, psychologists began to judge intellectual ability on the basis of some skills such as reaction time, sensitivity to physical stimuli and body proportions. The name of Sir Francis Galton, who was a pioneer in the measurement of **individual differences,** is important in this regard. He tried to measure sensory responses or visual and auditory acuity and reaction times and their relationship to differences in ability.These were first attempts to measure intelligence quantitatively and study individual differences.

The abbreviation "I.Q." meaning **"Intelligence Quotient",** most widely used measure of intelligence in modern times, was first coined by the German psychologist **William Stern** in the early 1900s. This term was later used in 1905 by Alfred Binet and Theodore Simon, who published the first modern intelligence test, the Binet-Simon intelligence scale. Later, the Binet-Simon scale was adopted for use in many other countries. The Binet and Simon tests were based on practical knowledge, reasoning, memory, vocabulary and problem-solving that predicted academic success of

students. During the course of development and use of the scales Binet-Simon realized that some children were able to answer more advanced questions that older children were generally able to answer, and vice versa. Based on this observation, they suggested the concept of a **Mental Age.** These tests were based on the premise that all children followed the same course of intellectual development, and hence children may be assigned a Mental Age which corresponded to their ability to perform tasks or solve problems appropriate for that particular age. For example, if a six year-old boy could perform the tasks at the same level as an average eight year-old; his mental age will be eight.

TO LEARN MORE

- German psychologist William Stern first coined the term ***"intelligence quotient"***, in the early 1900.
- In 1905 Alfred Binet and Theodore Simon published the first modern intelligence test- the Binet-Simon intelligence scale.

Later Lewis Terman of Stanford University adapted the Binet Simon tests for American use. This adapted test first published in 1916 and was called the Stanford-Binet Intelligence Scale. He advocated that an individual's intelligence can be measured as a quotient (intelligence quotient). The formula for the quotient was,

$$\text{I.Q.} = \frac{\text{(Mental Age)}}{\text{(Chronological Age)}} \times 100$$

Thus, a five year old boy who answers all questions meant for seven year old students would have an IQ of 140. On the other hand, a six year-old boy achieving at the same level as his six year-old students would score a 100. Thus, score of 100 is the average score which is still in use even today.

At the beginning of the First World War, U.S. Army needed to screen huge number of recruits. Accordingly Robert Yerkes, who was chairperson

of the Committee on the Psychological Examination of Recruits, developed two tests known as the Army Alpha and Army Beta in 1917. The Army Alpha was designed as a paper pencil test, while the Army Beta was an oral test for those who were unable to read. After the end of the First World War, the tests remained in use in a wide variety of situations outside of the military with individuals of all ages, backgrounds and nationalities.

American psychologist David Wechsler developed a new intelligence test known as the **Wechsler Adult Intelligence Scale (WAIS)** in 1955. Wechsler developed two different tests for use with children: the Wechsler Intelligence Scale for Children (WISC) and the Wechsler Preschool and Primary Scale of Intelligence (WPPSI). The adult version of the test has been revised since its original publication and is now known as the WAIS-IV.

The WAIS-IV has 10 subtests along with five supplemental tests. The test provides scores in four major areas of intelligence: a Verbal Comprehension Index, a Perceptual Reasoning Index, a Working Memory Index, and a Processing Speed Index. The test also provides two broad scores that can be used as a summary of overall intelligence: a Full-Scale IQ score that combines performance on all four index scores and a General Ability Index based on six subtest scores. However, unlike Stanford-Binet, which is scored on the basis of chronological age and mental age, the WAIS is scored by comparing the test taker's score to the scores of others in the same age group. The average score is fixed at 100, with two-thirds of scores lying in the normal range between 85 and 115. This scoring method has become the standard technique in intelligence testing and is also used in the modern revision of the Stanford-Binet test.

The modern IQ tests measure a precise mathematical score based on the Normal probability curve, with majority of people scoring the average and very smaller numbers of people at points higher or lower than the average. Roughly 95% of the population scores between 70 and 130 points. Important to remember here that the relationship between IQ score and mental ability is not linear and a person with a score of 120 does not have double the mental ability of a person with a score of 60. The most commonly

used tests are the Stanford-Binet, the Wechsler Intelligence Scale for Children (WISC), the Wechsler Adult Intelligence Scale (WAIS), and the Kaufman Assessment Battery for Children (Kaufman-ABC).



CHECK YOUR PROGRESS

Q 1: What is intelligence?

.....

.....

Q 2: What is an intelligence test?

.....

.....

Q 3: What is Mental Age?

.....

.....

9.5.1 Types of Intelligence Tests

Intelligence tests may be classified on the basis of different criteria. Some of the important types of intelligence tests are as follows:

- **Verbal, non-verbal and performance tests**
- **Individual and group tests of intelligence**

Verbal tests of intelligence: The verbal tests of intelligence are very old tests. Spoken and written language is the chief medium of these tests. Test-takers are usually given written instructions and they respond to the questions using either spoken or written language. Knowledge of language plays a very important role in these types of tests since. Abilities of the students are expressed through the means of language while responding to the items in these tests. Some examples of this category of tests are: Stanford

Binet test, Army Alpha test, Army General Classification test etc.

It is important to note that, verbal tests can be administered either individually or in groups. In other words, verbal tests can be classified again as individual verbal tests and group verbal tests.

The **chief merits of verbal tests** are:

- Individuals with correct knowledge of language can apply their intelligence properly.
- Thinking, judgement, reasoning, abstraction, comparison, analysis, synthesis, problem solving etc. capacities can be effectively measured using verbal tests.
- Any confusion during the administration of the test can be easily removed.

The **demerits of verbal tests** are:

- Verbal tests cannot be used with illiterates or with young children.
- Due to differences in language of different places, same tests cannot be used everywhere.
- The validity and reliability of these tests are not above doubt.
- Verbal tests cannot measure native intelligence of individuals.

Non-Verbal tests of intelligence: Non-verbal tests are those that do not require the use of language at any stage of administration. No spoken or written language is used to construct the items, giving instructions to test-takers or in responding to the items. Instructions to candidates are given using demonstration, pantomime, gestures, charts etc. The items are given in the forms of pictures or structures and the candidates have to select or find out the appropriate alternative by observing the pictures or structures in the question. Though these are non-verbal tests, these tests use paper-pencil as used by the verbal tests. However, the use of pen or pencil is limited only to record signs such as """ or "X", or to record other non-verbal responses such as underlining, encircling the correct alternative. The non-verbal tests are usually administered as group tests of intelligence. Some examples of non-verbal tests are: **Chicago Non-**

verbal test, Army Beta test, Raven's Progressive Matrices test etc. The **merits of non-verbal tests** include:

- These tests can be used with illiterates as well as young children.
- These tests can be used with people having difficulties in using language.
- These tests can be used in different places or countries; since these tests are non-verbal they are free from social and cultural influences to some extent.

The main demerit of the non-verbal tests is that non-verbal tests cannot measure all mental qualities and hence cannot give a composite picture of intelligence of an individual.

Performance tests of intelligence: Performance tests are those that use language only at the stage of giving instructions to candidates and the items are manipulative in nature. Test-takers have to solve problems manually or have to manipulate different objects while responding to the items. Performance tests are used to test some such qualities that cannot be measured through verbal tests. These tests are generally administered individually. Some of the important performance tests are: **Pinter-Patterson scale, Porteus Maze test, Alexander Pass-along test, Bhatia performance test etc.**

Some **merits of performance test** are:

- Performance tests are very effective for use with young children, illiterates and people with difficulties in using language.
- These tests have clinical value, because these tests can be used to diagnose mental or neurological disorders.
- These tests are free from cultural influences to some extent.
- These tests help to identify practical and mechanical skills of students which in their turn help in providing educational and vocational guidance to students.

Some **demerits of performance test** are:

- These tests are time consuming and cannot be used with groups.

- These tests cannot differentiate between the various levels of general intelligence.
- Reliability index of these tests are relatively low.
- These tests can measure only a few of the intellectual capabilities and hence cannot present an overall picture of intelligence of an individual.

Individual tests of intelligence: An individual test of intelligence is one which is administered to only one individual at a time. It is claimed that individual tests can accurately measure intelligence and can give a complete picture of intellectual abilities of an individual. Individual tests of intelligence can be verbal, non-verbal or performance type test. Alfred Binet popularized the individual tests of intelligence. Examples of verbal individual intelligence test are: Binet Simon intelligence test, Stanford Revision of Binet test, Wechsler Bellevue intelligence test etc. Similarly examples of non-verbal and performance individual tests are: Alexander Pass Along test, Koh's Block Design, Wechsler performance test etc.

Some of the **merits of individual test** are:

- Tester can pay more attention to the testee.
- Tester can easily encourage the testee and observe his behaviour during the test more closely.
- Scores on individual tests are not as dependent on reading ability as scores in group tests.
- Individual tests are more valid and reliable.

Some of the **demerits of individual test** are:

- Individual tests are time consuming
- Individual tests require a highly-trained tester.
- Individual tests cost more than the group test.

Group tests of intelligence: Group tests of intelligence were developed to meet pressing needs of testing large number of recruits during the First World War. Group tests of intelligence can be administered to a group of persons at a time. These tests help to test large groups simultaneously and the instruction and

administration procedures are also simplified. These tests do not require highly trained individuals to use the tests. Group tests can be either verbal or non-verbal. Examples of verbal group test are: Army Alpha test, Lorge Thorndike Intelligence test, Army General Classification Test (AGCT), Terman-McNemar Tests of Mental Ability etc. Examples of non-verbal group test are: Raven's Progressive Matrices test, Chicago non-verbal test, SRA test of general ability etc.

Some of the **merits of group test** are:

- Group tests can be administered to very large numbers simultaneously
- the role of the tester is simplified
- scoring of group test is usually more objective
- large, representative samples are often used and hence it leads to better established norms
- A highly verbal group test can have a higher validity co-efficient than an individual test.
- Group tests are economic in terms of time, money and effort required.

Some of the **demerits of group test** are:

- Scores on the group test are generally dependent on the reading ability.
- Information obtained by the group test is generally less accurate than the individual tests
- Tester has lesser opportunity to establish rapport, obtain cooperation, and maintain interest of the testee
- These tests cannot readily detect if testee is tired, anxious, unwell
- There is evidence that emotionally disturbed children do better on individual than group tests
- Test-takers' responses are more restricted.

9.5.2 Uses of Intelligence Tests

Intelligence tests have many uses that have been discussed under:

- **Educational uses:** Intelligence tests are of immense use in educational field. Following points are important in this regard:
  - **Selection of course:** Intelligence test works as a basis for selecting different subjects or courses for students. Different courses or subjects require different levels of intellectual abilities. Hence intelligence tests can help pupils to select courses in tune with their intellectual abilities.
  - **Selection of pupils:** Intelligence tests help to select prospective students at the time of giving admissions to an institution.
  - **Classification of students:** Students need to be classified in the field of education to meet various purposes. Intelligence tests help in classifying students on the basis of their intellectual qualities.
  - **Comparison of students:** Various educational decisions are based on comparison of students on the basis of their abilities. Intelligence tests help to make such comparisons between students.
  - **Prognosis of educational potentialities:** Students come to educational institutions with different inborn talents and qualities. Intelligence tests help to identify such strengths and qualities among the students and help to give them instructions and training to develop their potentialities fully.
  - **Identification of different categories of students:** All students do not have same level of intelligence. Teaching students with different levels of intelligence in a single class is very difficult and harmful. Intelligence test helps to identify students with different levels of intelligence and group them into homogeneous classes so that appropriate content and methods can be used in tune with their intellectual abilities.

Intelligence tests help to identify gifted as well as backward students so that appropriate instructional measures can be arranged for them.

- **Setting appropriate level of aspiration for students:** Setting appropriate level of aspiration is very important to achieve success and to avoid frustration. According to Sawrey and Telford, "One should aim for what one desires. In other words, one should very carefully set one's level of aspiration neither too high nor too low, if one ventures to be successful. It is possible only when one's potentiality can be properly assessed. Intelligence tests can serve this purpose, and hence they are found to assist the individual to establish a level of aspiration that is realistic in terms of intellectual potential."
- **Improving teaching-learning process:** Intelligence tests may help to improve the teaching-learning process. The view of Crow and Crow is noteworthy in this regard. They wrote, "Results of intelligence tests can help a teacher to discover what the child can learn and how quickly he can learn, as well as the teaching methods that should be applied and the learning content that should be utilized to guide the learner to use his mental potentialities to their utmost."
- **Educational guidance:** Guidance is an important educational service to help students to find their place. Intelligence tests help to provide proper guidance to students on the basis of their intellectual capacities.

- **Use in vocational guidance:** Similar to educational guidance, vocational guidance is very important to be successful and to live a contended life. Intelligence tests help to provide guidance to students in their choice and preparation for a particular vocation.
- **Diagnosis of mental disorders:** Various mental disorders may be diagnosed with the help of intelligence tests.

- **Use in child guidance:** In order to provide appropriate guidance to children, their level of intelligence must be known. Very high or very low level of intelligence, both may lead to problem behaviour of the child. Intelligence test helps to know the child properly and to give him guidance accordingly.
- **Studying individual differences:** Intelligence is the greatest factor of individual difference. Tests of intelligence help to understand the nature, depth and dimensions of individual difference among students.
- **Use in research:** Intelligence tests are extensively used in research. Various educational and psychological research problems need intelligence testing for finding their appropriate solutions.



CHECK YOUR PROGRESS

Q. 4: What is a verbal test of intelligence?

.....

.....

Q. 5: Mention any two demerits of group test of intelligence.

.....

.....

.....

9.6 CONSTRUCTION OF AN INTELLIGENCE TEST

In order to construct a standardized intelligence test, one must always remember some important concepts about the nature of intelligence. These concepts are:

- Intelligence is normally distributed among people. This implies that majority of the people are average and only a few are very bright or dull in terms of intelligence.

- Heredity and environment, both influence intelligence of an individual.
- Measurement of intelligence is relative, i.e. when we talk about intelligence of an individual we talk about his intelligence in relation to other individuals.
- Wide differences in terms intelligence exist among individuals.
- Intelligence grows with increase in age; but it does not grow throughout life and there is no definite age at which the growth of intelligence stops.
- Intelligence does not differ on the basis of sex.
- Race, caste or culture does not make difference to intelligence.

The general principles and the steps to be followed while constructing a psychological test (has been already discussed in UNIT-IV) apply to the construction of an intelligence test as well. Additionally, following considerations must govern the construction of a standardized intelligence test.

While planning for an intelligence test it has to be seen for which grade or age the test is to be constructed. Because a test designed for a particular age or stage must be of particular shape and nature. The items must also be developed or selected accordingly. Standardizing an intelligence test implies:-

- Standardizing the condition or situations under which the test is to be administered.
- Deciding what results could be expected from the test-takers in that standardized environment and in proportion to the level of intelligence. (fixing the norms)
- Ensuring that the results obtained are the measures of intelligence only (establishing validity)
- Ensuring that the scoring procedure gives same results on retesting (establishing reliability)
- Ensuring accurate interpretation of results.

After planning for the test, the first task is to write or collect items for the test. This collection of items depends on the nature of the test, i.e. items for a verbal test of intelligence will be different from the items for a

non-verbal test. More items than actually planned for the test should be collected at this stage since in the final stage some items would be deleted on the basis of analysis of the items.

The second task is to carry a preliminary administration of the selected items on a group of individuals selected from the population for which the test is to be constructed. The objective of this preliminary administration is to find out too difficult or too easy items, selecting the defective items and the items that are ambiguous. These defective items are discarded or revised and some other items are prepared if required.

The selected items are then arranged in ascending order of difficulty. It has to be kept in mind that most of the items should be within the difficulty level of at least 50% students and only a few items should be too difficult and too easy.

After finalizing and arranging the items and preparing clear directions to the takers of the test, the test is administered to a large sample of the group for which the test is constructed. After the administration of the test the responses or answers are scored objectively.

Subsequently, scales must be prepared for the appropriate interpretation of the scores. Various statistical means are used for preparing these scales. Some of the important ones are:

Percentile norm: Percentile norms help to judge and interpret the scores of students of a particular age group. Percentile norm is a process of converting raw scores of individuals into percentiles. A percentile is a score below which a given percentage of individuals in a group fall. For example, the 65th percentile is the score below which 65% of the individuals lay on the basis of their scores.

Norm and Mental Age: When the test-takers belong to different age groups then separate norm for each age group is determined. For example, a test consisting 100 items is administered to a group of students ranging from 8 year old to 11 year old. Now Mean of the scores for each of the age group (8, 9, 10, 11) is calculated which is the norm for that particular age group. When the test is administered to another group of students, then their scores are compared with the norm established for that particular

age group. If a student scores equal to the norm for a particular age group then his Mental Age would be the age of that group irrespective of his chronological age. For example, if an eight year-old boy could obtain a score which corresponds to the norm for average ten year-old; his mental age will be ten.

Intelligence Quotient (I.Q.): Only Mental Age cannot give a clear idea about the intelligence of an individual. An eight year old student with a mental age of ten and a six year old student with mental age of eight do not signify the same thing and cannot be compared. In order to make a comparison between the two, both their mental age as well as chronological age must be considered. Intelligence Quotient is a proportion of these two ages which is calculated using the following formula:

$$\text{I.Q.} = \frac{\text{(Mental Age)}}{\text{(Chronological Age)}} \text{ X } 100$$

However, the most important thing is to develop and define the concept of the mental ability or characteristic which the test-maker wishes to measure. Before giving final shape to the test, reliability and validity must be established using proper statistical measures. It must be kept in mind that the group on which the preliminary administration of the test is done, must be representative of the group for which the test is constructed. Moreover, the group must represent the population when the test is administered to establish norms for the test.



9.7 LET US SUM UP

The unit can be summarized as-

- Intelligence is a relatively permanent general mental ability which affects all the psycho-physical structure, strength and ability of an individual.
- This mental faculty determines controls and directs all mental functioning and behaviour of an individual.
- An intelligence test is a test designed to measure intelligence. There are different types of intelligence tests such as verbal and non-verbal, individual and group tests.

- Various intelligence tests are used to measure intelligence. Intelligence tests give an "intelligence quotient" or I.Q. as a measure of intelligence. Though there are numerous theories of intelligence, there is hardly any agreement among the theories about the nature of intelligence. Every theory explains intelligence from a different perspective contradicting the views of another theory.
- Construction of an intelligence test must follow the general principles and the steps that are followed while constructing a standard psychological test.



9.8 FURTHER READING

1) Chauhan, S. S. (1978): Advanced Educational Psychology, Vikas Publishing House Private Limited, New Delhi
2) Gardner, H. (1993): Multiple intelligences: The theory in practice, Basic Books, New York
3) Mangal, S. K. (2014): Advanced Educational Psychology, 2nd Edition, PHI Learning Private Limited, Delhi
4) Stanley & Hopkins (1966): Education and Psychological Measurement and Evaluation, Prentice Hall Pvt. Ltd. Eaglewood Cliff, New Jersey



9.9 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Intelligence is a wide-ranging general mental ability which affects all the psycho-physical structure, strength and ability of an individual. This mental faculty is relatively permanent which determines controls and directs all mental functioning and behaviour of an individual.

Ans to Q No 2: An intelligence test is a test designed to measure intelligence. There are different types of intelligence tests and they measure thinking, reasoning, learning, and other varieties of mental capacities and skills of an individual.

Ans to Q No 3: Mental Age of an individual is the age of that group, questions or tasks meant for which group can be answered by the individual. For example, if a five-year-old child can successfully answer the questions

meant for seven-year-old children, then his mental age would be seven years.

Ans to Q No 4: The verbal tests of intelligence are the tests that use spoken and written language as the chief medium of tests. These verbal tests are very old tests. Test-takers are usually given written instructions and they respond to the questions using either spoken or written language.

Ans to Q No 5: Two demerits of group tests of intelligence are: (i) tester has lesser opportunity to establish rapport, obtain cooperation, and maintain interest of the testee, and (ii) test-takers' responses more restricted in group tests.

9.10 MODEL QUESTIONS

A. Short Questions (Answer within 150 words)

Q 1: Discuss the nature of intelligence.

Q 2: How is intelligence measured?

Q 3: Discuss the concept of Multiple Intelligence.

B. Long Questions. (Answer 500 words)

Q 1: Write a brief note on the factor theories of intelligence.

Q 2: Discuss the uses of intelligence test.

Q 3: Write a note on the different types of intelligence tests.

Q 4: Discuss the procedure for constructing an intelligence test.

*** ***** ***

UNIT 10: MEASUREMENT OF APTITUDE

UNIT STRUCTURE

10.1 Learning Objectives

10.2 Introduction

10.3 Meaning of Aptitude

10.4 Definitions of Aptitude

10.5 Nature or characteristics of Aptitude

10.6 Measurement of Aptitude

10.6.1 Meaning of Aptitude Test

10.7 Types of Aptitude Test

10.7.1 Multiple Aptitude tests

10.7.1(a) General Aptitude Test Battery (GATB)

10.7.1(b) Differential Aptitude Tests (DAT)

10.7.1(c) Test of Primary Mental Ability (PMA)

10.7.2 Measures of Specific Aptitudes

10.7.2(a) Mechanical Aptitude Test

10.7.2(b) Clerical aptitude

10.7.2(c) Musical and Artistic Abilities

10.7.2(d) Measuring Aptitude for Art

10.7.2(e) Tests of Scholastic and Professional Aptitudes

10.8 Importance of Aptitude Test

10.9 Let Us Sum Up

10.10 Further Reading

10.11 Answers to check your progress

10.12 Model Questions

10.1 LEARNING OBJECTIVES

After going through this unit, 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 examples
- describe specialized aptitude tests with examples
- discuss importance of aptitude tests

10.2 INTRODUCTION

We generally observe that some persons are said to possess a certain specific ability or aptitude in addition to intellectual ability or intelligence. Next to intelligence, aptitude is considered to be another important characteristic of an individual which can predict success in a course of study or career. It is asserted that aptitudes are related to vocational success as intelligence is related to success in life in general. By testing aptitude, we are able to know whether an individual will become a good leader, a good musician or a good teacher. Because of these specific abilities people differ from one another and within themselves in their performances in different fields. Thus, along with the measurement of general intelligence, the measurement of aptitude is also necessary because both these measures provide sufficient information about the potentialities of an individual. So, in this unit we will learn about the aptitude tests.

10.3 MEANING OF APTITUDE

Before thinking about aptitude tests it is obvious to know about the meaning of aptitude. The word aptitude is derived from the word **"aptos"** which means **"fitted for"**. Aptitude may be described as a special ability or capacity which helps an individual to acquire the required degree of proficiency or achievement in a specific field. An aptitude 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. 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 selected test which have predictive value. When we say Reeta has an aptitude for music we mean that her present condition or

ability reveals that there is possibility of showing proficiency in music with proper training. It is a condition, a quality, or a set of qualities in an individual of the probable extent to which she/he will be able to acquire some specific knowledge, understanding or skill in different activities under suitable training.

The most appropriate explanation of aptitude is given by **Bingham,** which is included in Warren's Dictionary of psychology. It explains aptitude as- "A condition or set of characteristics regarded as symptomatic of an individual's ability to acquire with training some (usually specified) knowledge, skill or set of responses, such as ability to speak a language, to produce music …" Bingham further describes aptitude as a measure of probabilities of success of an individual with training is certain type of situations. Bingham states that the concept of aptitude rests upon three assumptions:

- An individual's potentialities are not equally strong
- Individuals differ from one another in their potentialities
- Many of these differences are stable.

Thus, aptitude is something more than ability. It is ability combined with suitability of performance. A scholarly person having high verbal ability may not necessarily be a good teacher. Teaching aptitude depends upon not only high scholastic ability, high interest in teaching, but also some personality traits which are wanted in a teacher. Every person with high measure of scientific ability cannot prove to be a good doctor unless he possesses some further traits like perseverance, sociability, love of humanity, dexterity of fingers to perform an operation etc.



LET US KNOW

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

- *Achievement looks to 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*

10.4 DEFINITIONS OF APTITUDE

To explore the meaning and nature of aptitude few definitions are cited below:

- Dictionary of Education - Aptitude is defined as a "Pronounced innate capacity for or ability in a given line of endeavor such as a particular art, school subject or vocation."
- According to Bingham (1937) - "Aptitude is 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."
- Traxler - "Aptitude is a condition, a quality or a set of qualities which is indicative of the probable extent to which an individual may be able to acquire, under suitable training, some knowledge, understanding, or skill."
- Van Dusen has defined the term in a rather strict manner. He states. "Aptitude is a measure of the probable rate of learning which results in interest and satisfaction and is relatively specific and narrow."

10.5 NATURE / CHARACTERISTICS OF APTITUDES

For clear understanding of the meaning and concept of aptitude following points are needed to be mentioned regarding true nature / characteristics of aptitude:

- Aptitude is an inborn and inherent quality which finds expression through training in a particular area.
- It is influenced by both hereditary and environmental factors. Hereditary provides the potential and environment helps to find expression to it. It remains, more or less constant.

- Aptitude is not a general ability but a specific talent needed for different kinds of job. Similarly, it is of a specific nature, as it is not common to all the situations where a particular ability is used. As it cannot be further analyzed into simpler abilities, it is unitary.
- An understanding of one's aptitude helps us to know what he can do in the future. Thus, it is a present condition but with a forward future reference.
- 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 derive a car but may have an aptitude for driving-which means that he has the capacity or ability to acquire proficiency in driving -provided he receives the proper training. Thus -

Aptitude = Ability + Probability of success

- It is symptomatic or indicative of one's ability or potentialities for a particular job. It implies fitness and suitability for the activities in question.
- An aptitude is a unique combination of abilities and personality characteristics, which predisposes a person to do one kind of work better than another and increases his chances of success at it.
- Aptitude are pluralistic. They are not unitary trait of human personality. 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 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.
- Aptitudes are not rigidly constant.

To sum up, aptitude is thought of as a natural tendency, special ability, or capacity or cluster of abilities. Often these natural abilities are looked at in relationship to a person's readiness to learn or their suitability

for a particular career. But it cannot appear or find expression unless the environment is favorable or suitable opportunity is provided.



CHECK YOUR PROGRESS

Q 1: Write two assumptions stated by Bingham regarding concept of aptitude.

.....

.....

Q 2: Mention two characteristics of aptitude.

.....

.....

.....

Q 3: Differentiate between aptitude and ability.

.....

.....

.....

Q 4: Fill in the blank-

Aptitude = — +—

.....

.....

10.6 MEASUREMENT OF APTITUDE

Aptitude can be measured through a multiple evidence which includes past academic achievement of the pupil, the observation of the teacher, the hobbies selected by the pupil, the interview, self-estimate by the pupil ,a standardized aptitude test etc. The most important of the above means is the aptitude test. An aptitude test is designed to assess what a person is capable of doing or to predict what a person is able to learn or do given the right education and instruction. Such tests may be used to assess

either mental or physical talent in a variety of domains to assess academic potential or career suitability. An aptitude test is used to determine an individual's propensity to succeed in a given activity.

10.6.1 Meaning of Aptitude Test

An aptitude test, by definition, is any type of assessment that evaluates the talent or ability or potential of an individual to perform a specific task, with no prior knowledge and/or training. These tests assume that individuals have inherent strengths and weaknesses, and have a natural inclination toward success or failure in specific areas based on their innate characteristics.

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. Aptitude tests indicate the ability to acquire certain behaviour or skills when given opportunity. 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 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 in terms of probabilities only.

In education, akin to intelligence tests, aptitude tests also measure a student's overall performance across a broad range of

mental capabilities. But aptitude tests also often include items which measure more specialized abilities--such as verbal and numerical skills--that predict scholastic performance in educational programs. Similarly, the difference between aptitude and achievement tests is sometimes a matter of degree. Compared to achievement tests, aptitude tests cover a broader area and look at a wider range of experiences. Achievement tests tend to measure recent learning in particular school subjects. Aptitude tests tell us the talent or potential that a student brings to the task regardless of the specific curriculum that the student has already experienced. Some aptitude and achievement tests look a lot alike. In fact, the higher a student goes in levels of education, the more the content of aptitude tests resembles achievement tests as the knowledge that a student has already accumulated is a good predictor of success at advanced levels.

10.7 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 general aptitude test or multiple aptitude test approach. The other approach to aptitude testing is to develop individual tests for each specific ability. Thus, two types of aptitude tests usually employed in psychological testing are-

- Multiple Aptitude Batteries or General Aptitude Tests and
- Specialized Aptitude Tests.

10.7.1 Multiple Aptitude tests

Multiple aptitude test batteries are intended for the independent measurement of each of several kinds of mental operation and that provide a separate score for each. On theoretical level multiple aptitude approach represents an adoption of a group factors

approach to abilities. On practical level, this approach recognizes that testing time will always be limited and that 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 test 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. Thus, an individual's potentialities in a number of academic and vocational areas can be analyzed or appreciated 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)

10.7.1 (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 GATB are as under-

- **Intelligence (G):** It is measured by the combined scores on three tests. Three dimensional space, vocabulary and arithmetical reasoning.
- **Verbal Aptitude:** It is measured by vocabulary tests of synonyms and anonyms.

- **Numerical Aptitude:** It includes both computational and Arithmetic Reasoning tests.
- **Spatial Aptitude:** It is measured by three Dimensional space test, involving the ability to comprehend two dimensional objects and to visualize effect of movement in three dimension.
- **Form Perception:** It test ability to match drawing of tools in one part, and of geometric forms in another.
- **Clerical Perception:** It requires matching of names.
- **Motor Coordination:** It is measured by simple paper-pencil test requiring the subject to make specified pencil marks in a series of squares.
- **Finger dexterity:** It test proficiency in assembling and disassembling rivets and washers.
- **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 21/2 hours of testing time. Raw scores are converted into standard scores for various aptitude 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.

10.7.1 (b) Differential Aptitude Tests (DAT)

Differential Aptitude Tests' was prepared by George K. Bennett, Harold G. Seashore and Alexander G. Wesman and published by U.S Psychological Corporation. This was comprehensive and carefully developed battery. This battery was designed to serve as an aid in the educational and vocational guidance of students in grades eight through twelve. It can also be used for unselected adults.

The test, first published in 1947 and it's another revision was published in 1972 (Form S and T). The DAT is probably the most widely used multiple aptitude battery in secondary school level. The battery consists of eight subtests of some specific abilities. Combination of such abilities is required for success in different occupations. The tests in the battery are power tests (The one exception is the clerical Speed and accuracy Test). Each test should measure level of ability. Each of the abilities represented in the battery should be independently tested. All eight tests of the battery were standardized on the same population. The battery includes the following eight tests:

- **Verbal Reasoning (VR):** This test consist 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 special form of analogies items that requires students to use reasoning rather than mere associations.
- **Numerical Ability (NA):** This test is designed to measure understanding of numerical relationship and facility in handling numerical concepts. Items are essentially computational, rather than problem solving type. Some of the items tests only proficiency in four fundamental process; others require understanding of quantitative concept and relationships.

The items on verbal reasoning and numerical ability are similar to items on general intelligence and scholastic aptitude.

- **Abstract Reasoning: (AR):** This test attempts to measure reasoning without the use of words. Abstract reasoning items measure no language reasoning ability.
- **Spatial Relations (SR):** This is the most ingenious test in the series. This test measures the ability to visualize and 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.

- **Mechanical Comprehensive (MC):** The test measures understanding of mechanical principles by means of a series of pictorially presented situation encountered in everyday life, involving mechanical and scientific principles. Thus, each item includes a picture illustrating a mechanical situation.
- **Language Usage Part I:** It is a spelling test in which the subject indicates the incorrectly spelled words in a list.
- **Language Usage Part II:** It consists of sentences in which the examinee is required to distinguish faculty 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 in so many academic and vocational pursuits.

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

Three of the tests- Verbal reasoning, numerical reasoning, abstract reasoning- measure broad intellectual abilities. Items on verbal reasoning and numerical reasoning use materials commonly taught in U.S. schools. 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 counselling 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 sub tests. 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 and validity of the test is high.

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.

10.7.1 (c) Test of Primary Mental Ability (PMA)

Primary Mental Ability test was developed from Thurstone pioneering research on structure of abilities. Factor analysis of intelligence had led to identification of certain mental abilities or traits 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 Science Research Association. So this addition was named the SRA Primary Mental Abilities. In this revision there was three scale one for ages 5-7, another 7-11 years and the third one 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-

- **Verbal reasoning:** the ability to understand ideas expressed in words or ability to deal with words and verbal concepts.
- **Number facility:** facility in doing simple numerical tasks.
- **Spatial relations:** ability to visualize and conceptualize. The test items are various simple designs and geometric figures, differently rotated.

- **Reasoning:** ability to think logically.
- **Perceptual speed:** ability to distinguish sizes and shapes.
- **Word fluency:** ability to produce words easily.

The verbal meaning, 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

Q 5: What is an aptitude test?

.....

.....

Q 6: What is Multiple Aptitude tests?

.....

.....

Q 7: Write the names of three aptitude test battery.

.....

.....

.....

Q 8: Mention three subtest included in GATB.

.....

.....

Q 9: How many sub test was included in DAT?

.....

.....

Q 10: Write the name of the only speed test of DAT.

.....

.....

Q 11: Who prepared PMA?

.....

.....

10.7.2 Measures of Specific Aptitudes

Multiple aptitude batteries measures abilities within the context of single, integrated test battery. But specific aptitude tests have been devised to measure the aptitudes of individuals in various specific fields of activities. In this approach the test constructor has concentrated on measuring a specific ability rather than attempting to provide a broad picture of abilities. These single aptitude tests includes test 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. We shall discuss the different Special Aptitude Tests under following headings:

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

Some of them are discussed below--

10.7.2. (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 psycho-physical ability of an individual. It is a combination of sensory and motor capacities plus abilities such as manual dexterity, 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 capacity and performance of 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 test this kind may be described below-

(i) 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 with 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.

(ii) 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 "Minnesota Mechanical Aptitude Test." This test, consisting of four sub-tests, was a test battery. The sub-tests are-

- **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 scored in terms of rate and accuracy of work. The test has been found useful in predicting success of junior high school boys in shop courses as well in a number of occupations such as mechanist and auto mechanist, wood worker, tool maker, iron worker etc.
- **Minnesota Paper Form board:** It this test paper form boards are presented in printed form. Each problem is composed of two or more parts of a geometric figure. (see figure) These parts when correctly assembled make a complete figure. The subject is required to identify the correctly assembled figure from among several choices. It presents 64 problems. The test measures

one's capacity to visualize and imaginatively manipulated geometric forms, to discriminate geometric patterns.

- **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 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. Interest measures are highly meaningful for those entering professional or skilled jobs. Thus this test is included in the battery to measure interest from 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 mechanical nature of problem solving and capacity to manipulate small objects with precision.

(iii) Bennett Mechanical Comprehension Test: This test of mechanical comprehension presents mechanical problem in 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 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--

(1) One form is designed for use with high school students, engineering school applicants and with relatively untrained and in experienced persons.

(2) The second form is meant for use with engineering school candidates and

(3) 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, perceptional 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, other by means of performance type materials and some others by means of paper and pencil tests.

10.7.2. (b) Tests of Clerical Aptitude

Clerical aptitude like mechanical aptitude is not a unitary function. This ability refers to the ability of routine clerical work. 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-

1. **Perceptual ability:** Ability to observe words and numbers with speed and accuracy.
2. **Intellectual:** Ability to grasp the meaning of words and symbols and to make correct decision regarding the questions they raise. This intellectual ability becomes increasingly indispensable as the clerical work increases in complexity.
3. **Various Mental Skills:** Ability to add and multiply, to spell correctly, to punctuate, and to use a wide variety of words and expressions.
4. **Motor Ability:** Ability to use various types of machines and tools like a type writer, 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. **Minnesta Clerical Test:** This test measures quickness and accuracy needed in the clerical job. This test was developed by M. Andrew. There are two sub-test in the test--

(i) Number comparison and

(ii) Name comparison.

In the first 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 pair which are alike. In the second subtest there are 200 names instead of numbers. The task is similar in both tests.

The test measures only perceptual speed and accuracy. The test is useful for distinguishing those who have promise for clerical work. A very low score on the test indicates lack of clerical aptitude, but a high score does not necessarily show positive aptitude for clerical occupations specially those involving responsibilities and variety of works. In such cases, the test result should be interpreted by considering the evidence from other sources.

B. **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.

10.7.2. (c) Musical and Artistic Abilities

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

(i) 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.

1. **Judgement of Rhythm:** The rhythm test requires the subject to compare two rhythm patterns and judge whether they are similar or different.
2. **Intensity or loudness discrimination:** Judging loudness of pairs of notes with gradation differences.
3. **Time discrimination:** Judging the duration of the some note sounded a second time.
4. **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 talent of little children. It is also incapable of measuring all components of musical aptitude, as it measures auditory aspect.

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

10.7.2. (d) Measuring Aptitude for Art

Art aptitude is a complex of interest, abilities energy, perseverance traits and other uncommon factors. Two important tests of art aptitude are --

(i) 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.

- **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 reproduction of three dimensional abstract sculpture perceptual function, rather than to make separate analysis of elements that enter aesthetic judgment.

10.7.2. (e)Tests of Scholastic and Professional Aptitudes

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

1. The American Council of Educational Scholastic aptitude test.
2. Scholastic aptitude tests in U.S.A.
3. Moss scholastic aptitude test for medical students.
4. Law School Admission Test.
5. Pre Engineering Ability Test.
6. Minnesota Engineering Analogies Test.
7. Teaching Aptitude test by Moss etc.

The main advantage of constructing 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. A particular specific aptitude test is usually developed for a specific purpose or situation. It is, therefore, more valid indicator of that particular situation.

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 person or 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

Q 12: Write the names of two -

(a) Mechanical Aptitude Test (c) Clerical aptitude

(b) Musical Aptitude Test (d) Scholastic aptitude tests

.....

.....

.....

Q 13: Write the name of a sub-test of Minnesota Mechanical Aptitude Test.

.....

.....

10.8 IMPORTANCE OF APTITUDE TEST

Applicability of aptitude and its testing is quite obvious in modern world. The need and prospect of such measurement in educational and vocational guidance has been increasing day by day. Following are some of the uses of aptitude tests--

- Every individual is not fitted to any and every kind of job. Their innate and inborn abilities and aptitude should be identified and necessary training should be given for use of manpower resources in today's world of industrialization and specialization. Aptitude tests are helpful for identifying abilities.
- Students have to make decision regarding the choice of subjects as various types of courses need specific nature of abilities or aptitudes. We can use an aptitude test battery for identifying the strong and weak points of children and to help the students in the proper choice of courses, subjects and careers accordingly.
- Vocational guidance largely depends on use of different kinds of aptitude tests. Many students have different aptitudes needed for

different kind of vocations. It is necessary to identify the vocation to which they are best fitted on the basis of their aptitude.

- Aptitude tests can be used for vocational selection. With these tests we can reduce the number of probable failures among those who take up a particular job. Thus aptitude tests are used for the purpose of prediction for future success in careers.
- There are certain specialized courses such as Law, Medicine, and Engineering etc. where selection of the trainees may be considered very essential before training is given to them. Hence, students can be selected for admission to these professional courses on the basis of aptitude test results.
- Aptitude tests are used to help the students in the improvement of certain special traits. Knowing something about the aptitude level of students in a given class can help a teacher identify which students are not learning as much as could be predicted on the basis of aptitude scores.

It is thus clear that, aptitude testing is of much importance for any purposeful guidance and counselling program or entrance examination to specialized academic and professional courses or selection for specialized jobs. Aptitude tests, coupled with other criteria like I.Q. interest, aspirations, personality traits etc. can help to a greater extent in avoiding waste of human as well as material resources by placement of individuals in places and lines of work in which they are best fitted. It is important that aptitude test data must be used along with previous achievement data, present interests, leisure time activities and work habits etc. It is worth noting that it estimates future achievement in terms of probability only. Aptitude tests score only provide leads or suggestions or help in career planning and do not automatically match individuals to suitable courses or studies, and occupations.



CHECK YOUR PROGRESS

Q 14: Write three points showing importance of aptitude test.

.....

.....

Q 15: Mention a limitation of aptitude test.

.....

.....



10.9 LET US SUM UP

The unit can be summarized as-

- 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 professional and vocational success of an individual. There are different types of aptitude tests. Generally, two types of aptitude tests usually employed in psychological testing are test (a) Multiple Aptitude Batteries or General Aptitude Tests and (b) Specialized Aptitude Tests.
- Multi aptitude batteries measures abilities within the context of 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 field of activities. (1)

Mechanical Aptitude. (2) Musical and Artistic Aptitude. (3) Professional aptitude tests (4) Scholastic aptitude

10.10 FURTHER READING

1) Anastasi Anne (1976): Psychological Testing 4th Edition, New York: Macmillan Publishing Co. Inc.
2) Anderson, Scarvia B. Samuel Ball, Murphy, R.T. and et.al (1976): Encyclopedia of Educational Evaluation, Jessey-Bass Publishers.
3) Asthana B (2015): Measurement and Evaluation in in Psychology and Education, Binod Pustak Mandir, Agra.
4) Brown Frederick G. (1097): Principles of Educational and Psychological Testing, Holt: Rineheart and Winston, 1976.
5) Boehm, Ann E. and White, Mary Alice (1982): The Parent's Handbook on School Testing. New York: Teacher's College Press.
6) Freeman, F.S: Theory and Practice of Psychological Testing, Oxford and IBH.
7) Gronlund, Norman, E (1968): Readings in Measurement and Evaluation, Education, and Psychology, The Macmillan Company.
8) Goswami M (2010): Measurement and Evaluation in Education and Psychology, Nilkamal Publishing House, Hyderabad, 2010
9) Jensen, Arthur R.(1981): Straight Talk About Mental Tests Their Uses and Abuses Explained in No-nonsense Terms by the World's Foremost Authority, New York: The Free Press.
10) Mehrens, William A. and Lehmann, Irvin J. (1987): Using Standardized Tests in Education, Fourth Edition. New York: Longman, Inc.
11) Mehrens, William, A. and Lehmann, Irvin, J (1978): Measurement and Evaluation in Education and Psychology, Holt, Rinehard and Winston.
12) Sattler, Jerome M. (1988): Assessment of Children, Third edition. San Diego: Jerome M. Sattler, Publisher.

10.11. ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Two assumptions stated by Bingham are-i) an individual's

potentialities are not equally strong, and ii) individuals differ from one another in their potentialities

Ans to Q No 2: Two characteristic of aptitude are-a) they are inborn and find expression through training, b) they are influence by both hereditary and environmental factors

Ans to Q No 3: Aptitude is something more than the ability. It is ability plus suitability of performance.

Ans to Q No 4: Aptitude = Ability + Probability of success.

Ans to Q No 5: 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.

Ans to Q No 6: Multiple aptitude test batteries are intended for the independent measurement of each of several kinds of mental operation and that provide a separate score for each. On theoretical level multiple aptitude approach represents an adoption of a group factors approach to abilities

Ans to Q No 7: The General Aptitude Test Battery (GATB), Differential Aptitude Test (DAT), Primary Mental Abilities (PMA)

Ans to Q No 8: a) Intelligence (G), b) Verbal Aptitude, c) Numerical Aptitude

Ans to Q No 9: Eight tests

Ans to Q No 10: Clerical Speed and accuracy

Ans to Q No 11: Thurstone.

Ans to Q No 12: a) Bennett Mechanical Comprehension Test and Minnesota Mechanical Aptitude Test

- Minnesta Clerical Test and General Clerical Test (GCT)
- Seashore Measures of Musical Talents and Wing Standardized Tests of Musical Talent
- Scholastic aptitude tests in U.S.A. and Moss scholastic aptitude test for medical students.

Ans to Q No 13: Minnesota paper form board

Ans to Q No 14: 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.

Ans to Q No 15: Do not automatically match individuals to suitable courses or studies, and occupations

10.12. MODEL QUESTIONS

A. Short Questions (Answer within 150 words)

Q 1: What do you mean by Aptitude?

Q 2: What are the uses of aptitude tests? Show your acquaintance with any one of the tests of aptitude.

Q 3: What is meant by special ability? Give an account of any test that is used for assessing the mechanical ability of a student.

B. Long Questions. (Answer 300-500 words)

Q 1: What are aptitude tests? Discuss the value of aptitude tests in the field of education.

Q 2: What is aptitude tests? Describe any one of the multiple aptitude test batteries.

Q 3: Write notes no

(a) DAT, (b) SRA: PMA

(c) Musical Aptitude test (d) Clerical Aptitude test.

Q 4: What are the factors involved in the clerical aptitude? Which tests can be used for measuring clerical aptitude?

Q 5: What tests can be used for measuring art aptitude? Discuss.

*** ***** ***

UNIT-11 MEASUREMENT OF ATTITUDE

UNIT STRUCTURE

11.1 Learning Objectives

11.2 Introduction

11.3 Meaning of Attitude

11.4 Definitions of Attitude

11.5 Nature or characteristics of Attitude

11.6 Importance of Attitude Testing

11.7 Measurement of Attitude

11.8 Types of Attitude Test

- 11.8.1 Attitude scales
- 11.8.1(a) Thurstone's scale
- 11.8.1(b) Likert Type scale
- 11.8.1(c) Semantic Differential Scale
- 11.8.1(d) Guttman scales
- 11.8.1(e) Master Scale Method of Remmers and Silence
- 11.8.2 Rating scales
- 11.8.2(a) Rank order scale
- 11.8.2(b) Percentage of Population scales
- 11.8.3 Indirect scales
- 11.8.3(a) Projective Techniques

11.9 Let Us Sum Up

11.10 Further Reading

11.11 Answers to check your progress

11.12 Model Questions

11.1 LEARNING OBJECTIVES

After going through this unit, learner will be able to-

- describe meaning and characteristics of attitude.
- explain measurement techniques of attitude.
- describe Thurstone's scale.

- describe Likert type scale.
- discuss indirect scales.
- describe uses of attitude scale
- describe limitations of attitude scale

11.2 INTRODUCTION

Learners, we know that, we tend to approach and seek out to be associated with things we like and we avoid, ignore or reject things we do not like. Attitudes are expressions of how much we like or dislike various things. In psychology, attitudes constitute valuable elements for the prediction of behaviour. They are pre-conceived and they become part of our nature. Attitude influences an individual's choice of action, and responses to challenges. An understanding of attitudes is also important because attitudes help people adapt to their environment. Learners, in this unit you will learn about attitudes testing.

11.3 MEANING OF ATTITUDE

Learners, now let us discuss about the meaning of attitude. The attitude is the voluntary disposition of a person. It may be explained as a mental preparedness which provides stimulus for an individual in some directions. All people, irrespective of their status or intelligence, hold attitudes. Attitude is a psychological tendency to respond consistently positively or negatively towards certain idea, object, person, or situation. A positive attitude is a strong-willed and optimistic reaction to adversity, and a negative attitude is a harrowing vision, with the consequent reaction apathetic or aggressive

Attitude is a mental and emotional response to the variety of circumstances that occur in life. From this point of view attitude refers to a set of emotions, beliefs, and behaviours toward a particular object, person, thing, or event, one holds towards a person, object, task or idea. These are consequence of a cognitive, affective and behavioral process. Attitudes are generalized ways of behaving and viewing situations which are inferred from the verbal and non-verbal behaviour of the individual. It is expressed

outwardly, through gestures, movements, words, cries, laughs, tears, or sometimes stillness and apathy, but it responds to internal, cognitive and affective stimuli. The attitude is rather a social motivation rather than a biological motivation. The scale of values that constructs a person throughout his life, will contribute to the attitudes that manifests. We cannot have an attitude to what is not known or is not valued as good or bad.

While attitudes are enduring, they can also be subjected to change. Attitudes develop and change in a number of different ways. Psychologists define attitudes as a learned and adopted tendency. Attitudes develop gradually as a result of individual's experiences. From experience, people acquire a certain predisposition that allows them to respond to stimuli. These are results of stimulus and response as well as the muscular preparation for the response set-up in the neuro-muscular system. Thus, attitudes are defined as a state of the nervous and mental disposition that is organized from the experiences and guides or directs the response of a person to certain events. Like most abstract terms in the English language, attitude has more than one meaning.

- It must have definite orientation in the world of objects (or values), and in this respect differ from simple and conditioned reflexes.
- It must not be an altogether automatic and routine type of conduct, but must display some tension even when latent.
- It varies in intensity, sometimes being predominant, sometimes relatively ineffective.
- It is rooted in experience, and therefore is not simply a social instinct.

11.4 DEFINITIONS OF ATTITUDE

The following definitions of attitude forwarded by different psychologists will provide us with a clear understanding of the term attitude-

- Thurstone said, "An attitude denotes the sum total of man's inclinations and feelings, prejudice or bias, pre-conceived notions, ideas, fears, threats and other any specific topic."
- Freeman hold that - "An attitude is a dispositional readiness to

respond to certain situations, persons or objects in a consistent manner which has been learned and has become one's typical mode of response. For example, one's views regarding a class of food or drink (such as fish, liquors), sports, mathematics or democrats are attitudes."

- According to Anastasi - "An attitude is often defined as a tendency to react favorably or unfavorably toward a designated class of stimuli, such as a national or racial group, a custom or an institution."
- According to Allport - "An attitude is a mental and neural set of readiness exerting directive dynamic influence upon the individual's response to all objects and situations with which it is related.
- Kimball Young's opinion about attitude is - "An attitude may be defined as a learned and more or less generalized and affective tendency or predisposition to respond in a rather persistent and characteristic manner, usually positively or negatively (for or against) in reference to some situation, idea, value, material object or class of such objects or person or group of persons."
- Whittaker: "An attitude is a predisposition or readiness to respond in a predetermined manner to relevant stimuli."
- According to N.L. Munn, "Attitudes are learned predispositions towards aspects of our environment. They may be positively or negatively directed towards certain people, service or institution."
- Barr, David, Johnson: "An attitude may be defined as a learned emotional response set for or against something."
- L.L. Bernard : " The behavior which we define as attitudinal or attitude is a certain observable 'set' organism or relative tendency preparatory to and indicative of more complete adjustment."

In short, it can be concluded that the attitude is a state in which man moves to act or do anything in response to a situation or condition of the objects in the surrounding environment. Based on various definitions of attitudes, we can explain attitude as lasting beliefs and cognitions in general, endowed with an emotional charge in favour or against a defined object,

which predisposes to a consistent action with cognitions and emotions relating to that object.

11.5 NATURE OR CHARACTERISTICS OF ATTITUDE

The characteristics or features of attitude include the following--

- Attitudes are the complex combination of things we tend to call personality, beliefs, values, behaviours, and motivations. It indicates the sum total of a man's inclinations and feelings.
- An attitude is an expression of favour or disfavour toward a person, place, thing, or event (the attitude object). It may fall anywhere along a continuum from very favourable to very unfavorable.
- An attitude exists in every person's mind. It guides or motivate the behaviour of the individual in a particular direction. It helps to define our identity, guide our actions, and influence how we judge people.
- Attitudes are learned or acquired.
- Attitudes are mental and neural state of readiness organized as a result of individual's experiences.
- An attitude may be considered as "response consistency." Thus, attitude may be of relatively permanent nature, but changes may be effected in them in suitable and relevant situations. Thus the nature of attitude is dynamic.
- Attitudes constitute one aspect of an individual's personality. It includes certain aspects of personality as interests, appreciations and social conducts.
- The unconscious motive is a very important factor in the creation of attitudes. Sometimes an individual is unaware about the reason of their attitude.
- An evaluative disposition attitudes are based upon cognitions, affective reactions, behavioral intentions and past behaviours.
- Although the feeling and belief components of attitude are internal to a person, we can view a person's attitude from his or her resulting behavior. Various kinds of affective experiences are also attached to attitudes.

- An attitude has different aspects- such as direction, intensity, generality or specificity
- Attitudes are considered undercurrents variables, not be directly observable but subjected to observable inferences

In short an attitude is an affective feeling of liking or disliking toward an object (which can be basically anything) that has an influence on behavior.



LET US KNOW

There are several different components that make up attitudes. The components of attitudes are sometimes referred to as CAB or the ABC's of attitude.

- **Cognitive Component:** your thoughts and beliefs about the subject belief or opinions held consciously
- **Affective Component:** how the object, person, issue, or event makes you feel emotions or feelings.
- **Behavioral Component:** how the attitude influences your behavior inclination for action. And another component is-
- **Evaluative:** positive or negative response to stimuli.

Attitudes can also be explicit and implicit.

- Explicit attitudes are those that we are consciously aware of and that clearly influence our behaviors and beliefs.
- Implicit attitudes are unconscious but still have an effect on our beliefs and behaviors.



CHECK YOUR PROGRESS

Q 1: What is an attitude?

.....

.....

Q 2: Mention two points regarding nature of attitude.

.....

.....

11.6 IMPORTANCE OF ATTITUDE TESTING

An attitude is an important trait of personality. The development of desirable attitude is one of the main objectives of education. Attitudes give clues regarding adjustment and personality. For example, our attitudes towards different situations determine our mode of reaction and adjustment in these situations. Thus, attitude testing is an important part of the comprehensive process of educational evaluation. When the attitudes of an individual is observed, it is feasible to foresee its mode of action. Psychology has undoubtedly contributed greatly to the identification of attitudes and its possible control. Measurement of attitude is a major concern in different areas of education. Personnel at all levels are involved in assessing attitudes and providing activities conducive to the formation of desirable attitudes.

Appropriate attitude formation is an important aspect of education. It is essential in educational setting to study the attitudinal changes that are brought about through the process of education. It is possible to know the extent to which the necessary attitudes have been developed among the students. Attitudes are significant for satisfaction and success in the educational and vocational fields. They play an important role on the success in different vocations. Different types of profession require different attitudes.

The attitude of a person in Armed Forces is different from one in teaching profession. An educational and vocational plan related to the child's ability and aptitude may fall short of realization because it is not consistent with his attitude. Attitudes determine the choice of career plans. Attitudes can be tested through various techniques.

11.7 MEASUREMENT OF ATTITUDE

Attitudes are measurable and as such we can assess how they might affect behaviour. This attitude measurement is a procedure in which individuals' attitudes and responses are assigned quantitative values. Thus attitude measurement refers to assigning numerals to the person's attitude towards the object in question. Attitude is an abstract, complex and ultimately a hypothetical construct that cannot be observed directly. Thus its measurement is a difficult and controversial task. It is being remarked that psychological measurement in general and attitude measurement in particular, are not measurements in the true sense.

Teachers, of course, are more likely to assess attitude with informal instructions or observations etc. than with standardize attitude test. But measurement of attitudes is not a new field, since pioneers as H.H. Remmers and L.L.Thurstone have done work in this field for many years. The critical point in the measurement of the real attitudes of the individual is his verbal behaviour or his non-verbal overt behaviour. Perhaps the most straightforward way of finding out someone's attitudes would be to ask them. However, attitudes are related to self-image and social acceptance (i.e. attitude functions). In order to preserve a positive self-image, people's responses may be affected by social desirability. For this there is the need for instruments which do not destroy the natural form of the attitude in the process of describing it. Psychometricians devised suitable techniques to measure attitudes quantitatively. But here the question arises whether an individual is to be judged by his actions according to his words or according to his attitude scale scores. It is worth mentioning that attitude being a complex affair cannot be wholly described by any single numerical index.

11.7 TYPES OF ATTITUDE TEST

There are several ways of measuring attitudes toward self, others, and a variety of activities, institutions, and situations. These are -

- Attitude scales
- Rating scales
- Indirect scales

11.8.1 Attitude scales

Attitude scales are most widely used tools to measure attitudes. Attitude scales began to be devised which could assess attitudes in a quantitative manner. The Dictionary of Education (1959) defines an attitude scale as "an attitude measuring instrument, the units of which have been experimentally determined and equated, designed to obtain quantitative evaluation of an attitude to be distinguished from attitude questionnaire in which there is no such rational equality of units." First attitude scale was devised by L.L.Thurstone in 1929. Attitude scales reveal the reactions of the individual towards particular things/subjects and from these reactions his attitude can be deduced.

These attitude scales attempt to determine what an individual believes, perceives or feels. These scales provide the relative position of an individual along a unidimensional attitude continuum in quantitative terms. The aims of these scales are to locate upon a measure extending from one limit of acceptance to the other limit of refused. An individual's position is determined in between these two. They yield a score based on the individual's response to the statements pertaining to the issue under consideration. Thus attitude scales are self-report inventories designed to measure the extent to which an individual has favourable or unfavourable feelings towards some persons, objects, institutions or ideas.

Construction of attitude scale is rather different from other

tools. In a test of ability an examinee either passes an item or he does not. Items are either too easy or two difficult for him. Attitude scale presents some unique problems in form of some statements. As a whole, it indicates his attitude in that context. The examinee simply reflects his/her opinions with reference to the statements given. The scale can be a success only when the statements contained in it and the answers which are forth coming to them have a psychological relation with the attitude measured. According to Freeman attitude scales are based upon several assumptions. These are--

- The scale should deal with a controversial question.
- An individual's feelings and insights in regard to the question will determine his responses to the various statements that are made pro and con; and
- The statements can be scaled regarding the degree to which they favour, or are opposed to the question under consideration.

There are many forms of attitude scale in psychology for attitude testing. In building these scales care is needed to see that these are valid, that is, adequacy in serving purpose for which it is intended. These scales are differed in respect of the method of their construction and kind. There are various methods of attitude scale construction. On the basis of methods of construction, the attitude scales may be classified as below -

11.8.1 (a) Thurstone's scale

This method was developed by Thurstone and his associates. In 1928 Louis Leon Thurstone published an article titled "Attitudes Can Be Measured". In it he proposed an elaborate procedure to assess people's views on social issues. Thurstone and his colleagues constructed some scales for the measurement of opinion and attitude of human groups concerning varied questions like- War, the Church, Negroes, Capital punishment, birth control etc. It seeks

to analyze people on a continuous scale having equal-appearing units. For this reason this technique is also known as the method of **"Equal -appearing intervals."** In this type of scales scale values and other characteristics are largely determined by a panel of judges. Thurstone scale construction method was constituted of the following steps--

1. **Preparing statements:** In the beginning a large number of simple statements which express some kind of opinion (agree/ disagree) about the attitude objects under study are collected. For example, if we wanted to find out people's attitudes towards immigrants, our statements might include:
   - Immigrants drain social services.
   - Immigrants take jobs away from regular people.
   - Immigrants perform low-wage, unpopular tasks.

   These possible statements (related to a particular topic, object or question or institution etc.) are obtained from a group of selected writers, other experts and layman. These items should cover the entire continuum of attitude toward the object in question, from extreme unfavourable on the one hand to extreme favourable on the other. Each item is placed on a slip of paper.
2. **Editing of statements:** Then these statement should be edited.
3. **Judging of statements:** Then these statements are classified by a large number of judges on an eleven-point scale. The judges rate the statements on how effective they would be at uncovering attitudes. This is done by placing each statement in one of eleven-point scale, presumably forming a continuum, according to degree of favourable and unfavourableness of each statement. It is necessary to see whether the statements in the scale are in a definite order or not. The order of statements should be such that they proceed from maximum to minimum acceptance in approximately equal steps. The equal intervals are differences in attitude that appear equal to the judges. The lowest score (1) should indicate an extremely unfavorable attitude and the highest

score (11) should indicate an extremely favorable attitude. They have to place the natural statement of the middle of the scale. Statements that are judged to be ambiguous or irrelevant to the continuum are eliminated.

1	2	3	4	5	6	7	8	9	10	11

Negative Positive

4. **Finding out Scale Value:** The next step is to determine the scale value of the statements. For statements retained, the scale value is the median position assigned by the judges. The scale value for each statement is found out graphically by an ogive or cumulative percentage curve of the number of judges placing the statements in different classes in the eleven point scale. The median of the judges' locations for a statement is its scale value. Thus, the p50 or 50% level of the curve shows the scale value for the statement. To illustrate the procedure for finding out the scale value for each statement, cumulative percentage of the number of judges placing a particular statement at different point of the eleven-point scale is shown in the following table (Number of judges being 30)

Scale points	**F (No. of judges)**	**Cumf**	**Cumf %**
1	5	5	16.7
2	11	16	53.3
3	9	25	83.3
4	4	29	96.7
5	1	30	100
6	0	--	--
7	0	--	--
8	0	--	--
9	0	--	--
10	0	--	--
11	0	--	--

From this table an ogive is prepared. 50% level or p50 of the graph shows the scale value of the statement in question. (See the figure 11.1)

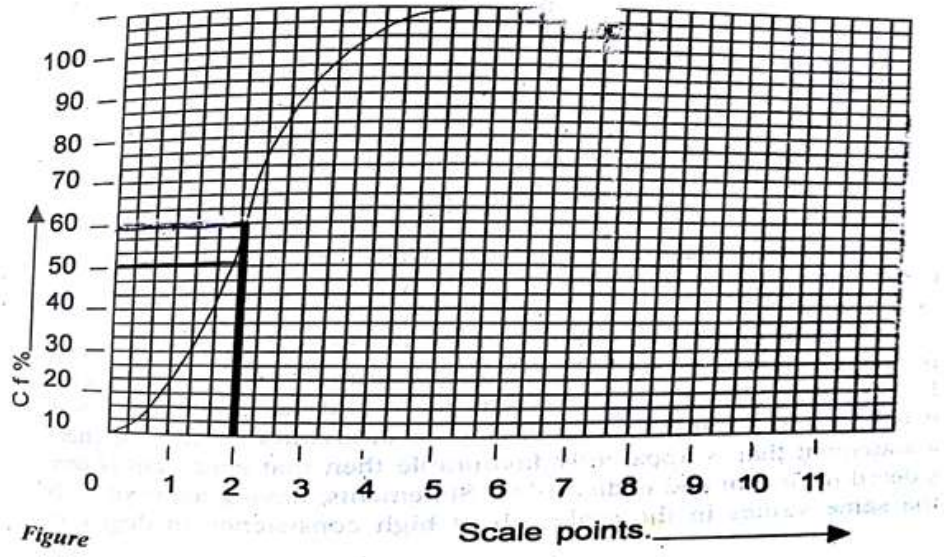


Figure 11.1: Ogive presenting scale value

5. Before inclusion in the final scale, each statement is analyzed for consistency with the general attitude found by the total scale. The statements chosen are then placed in a preliminary check list. This list is given to subjects as means of checking the relevance of items. For example- on a scale to determine attitude towards churches, if it is found that many persons having an unfavourable attitude check a statement that is apparently favourable then that statement is considered irrelevant and is discarded. Statements, having approximately the same values in the scale show high consistency in degree of endorsement. The final scale consists of items which have passed, the tests for ambiguity and relevancy. Final items are selected to represent the entire range of attitudes.

For example following statements are included in the attitude scale toward the church with their scale values (from Freeman)

- I think the church is a parasite on society (11)
- I believe what the church teaches but with mental reservations (4.5)

- I think the teaching of the church is altogether too superficial to have much social significance (8.3)
- I believe in religion but I seldom go to church (5.4)
- I believe the church is the greatest institution in America today (1.7)

To give an idea of the spread of the distribution about the scale value(s) the inter quartile range (IQR) is also calculated. This may be regarded as a measure of ambiguity. Statements that are judged to be ambiguous or irrelevant to the continuum are eliminated.

In a Thurstone attitude scale, the respondent checks those statements with which he agrees, his score being the median of the scale values of the statement he has marked. The attitude score of an individual obtained by this method has an absolute interpretation in terms of the psychological continuum of scale value of the statements making up the scale. If this score falls in the middle range of the psychological continuum, the attitude of the individual is described as ?neutral. If it falls towards the favourable end of the continuum, it is described as 'favourable' and if it falls towards the unfavourable ends, it is described as 'unfavourable'. This method permit direct comparison of the scores of any attitudes so measured. The validity of such comparison, however, has been questionable because the defined 'neutral points' of different attitude are not necessarily the same. The interval of the scale is not equal but only appear equal.

11.8.1 (b) Likert Type scale

This method generally associated with the work of Rensis Likert, a sociologist at the University of Michigan. Likert developed an attitude scaling technique simpler than the Thurstone method. His original report entitled "A Technique for the Measurement of Attitudes" was published in the Archives of Psychology in 1932.His goal was to develop a means of measuring attitudes in a 'scientific' way.

Specifically, he sought a method that would produce attitude measures that could reasonably be interpreted as measurements on a proper metric scale.

A Likert scale taps agreement or disagreement with a series of belief statements. It is an application of item analysis procedures borrowed from test-construction techniques. First of all as in the case of method of equal appearing intervals, a large number of statements relating to the attitude object in question are collected. These items are to be carefully edited to eliminate ambiguous irrelevant, and otherwise faulty items. The items are then drawn up in the form of a questionnaire or attitude test, each item being given multiple response categories of "strongly agree, agree, undecided, disagree, strongly disagree." The statements should be so constructed that for half of them an "agree" response represents a favourable attitude toward the attitude object in question, and for the other half a "disagree" response represents a favourable attitude. Then this is administered to a reasonably large group of subjects and asked to indicate their own attitudes by checking the response to each item which most nearly expresses their feeling on that item. The subject indicates his reaction to each statement usually on a five point scale. To score the scale the item score is obtained by assigning arbitrary weights of 5, 4, 3, 2 and 1 for Strongly Agree (SA), Agree (A), Undecided (U), Disagree (D) and Strongly Disagree (SD) for the statements favouring a point of view. On the other hand, the scoring weights of 1, 2, 3, 4, and 5 are used for the respective responses for statements opposing this point of view. For example, "strongly agree' with a favourable statement would receive a score of 5 (five), and similarly "strongly disagree" with an unfavourable statement would also receive a score of 5.

The items are then analyzed for their discriminatory power with respect to measurement of the attitude in question, by item-analysis procedures. Perhaps, the simplest index is to take the top and bottom of subjects on the distribution of total scores, and

calculate the mean of the responses to each item for each of these groups separately. Those items are the most discriminating which show the greatest discrepancy in mean response between high and low extreme groups. The final attitude scale is then constructed by choosing from the total list which show the greatest discrimination.

The sum of the weights assigned to the responses to the statements represent the individual's total score which represents the person's attitude towards the subject. Thus this scale is also known as **'Method of Summated Ratings'**. Generally, highest mark is considered as the attitude preferred and lower aggregate will be considered a disapproval of the attitude concerned (in case of positive statement).Sometimes this total score is interpreted in terms of empirically established norms. This method measures attitude only in the sense that individuals are given a rank order according to attitude intensity.

The main advantage of Likert's method is that it does not require the use of a group of judges to arrange statements into different categories. Items are selected solely on the basis of the responses of subject to whom they are administered in the course of developing the test.

Comparison of Thurstone and Likert techniques

Several investigators have tried to compare the merits of Thurstone and Likert techniques. There is little difference between these two types of scale. The factors which make for invalid self-report are equally present in both tests. But construction of a Likert type scale requires less time than Thurstone procedure. The Likert method permits response sets to influence the score, which might lower validity. Whereas in Thurstone scale no response sets affect the score. This scale is easy to response. In Thurstone scale, the attitude score obtained by a single subject has an absolute interpretation. Thus the interpretation of an attitudes score on this scale can be made independently of the distribution of scores for a

particular group of subjects. But interpretation of attitude on the basis of a single score is difficult in case of Likert scale.

11.8.1 (c) Semantic Differential Scale

The semantic differential scale uses bipolar adjectives (i.e. with opposite meanings), to measure the meaning associated with attitude objects. Thus, to prepare a semantic differential scale, we must first think of a number of words with opposite meanings that are applicable to describe the subject of the test. In this scale an individual gives quantitative rating to the subject of the attitude scale on a number of bipolar adjectives such as good-bad, friendly-unfriendly, positive-negative etc. The respondent indicates the point on the continuum between the extremes which represents her or his attitude. Respondents tick to indicate how they feel about what is being measured. A single attitude may include different meanings or qualities of evaluation. For example, a negative attitude about cigarette advertising might be made up of several beliefs or values. This can be tested in the following way:

A statement is made followed by a series of scales. E.g. Cigarette advertising is:

- Good 1 2 3 4 5 Bad
- Creative 1 2 3 4 5 Ugly
- Honest 1 2 3 4 5 Dishonest
- Healthy 1 2 3 4 5 Unhealthy

The semantic differential technique reveals information on three basic dimensions of attitudes: evaluation, potency (i.e. strength) and activity.

- **Evaluation** is concerned with whether a person thinks positively or negatively about the attitude topic (e.g. dirty - clean, and ugly - beautiful).
- **Potency** is concerned with how powerful the topic is for the person (e.g. cruel - kind, and strong - weak).

- **Activity** is concerned with whether the topic is seen as active or passive (e.g. active - passive)

11.8.1 (d) Guttman scales

Scales constructed in accordance with the 'scale analysis' techniques devised by Louis Guttman and his co-workers is known as Guttman scales. A Guttman scale, however, attempts to determine whether an attitude is unidimensional. It is unidimensional if it produces a cumulative scale. In an cumulative scale, an individual who agrees with a given statement also agree with all related preceding statements i.e. if one agree with statement 3, also agree with statement 2 and 1. The Guttman scale focuses on items that vary in their degree of psychological difficulty. "Guttmann believes that a genuine scale, capable of legitimate measurement, exists when homogeneity is virtually complete".

11.8.1 (e) Master Scale Method of Remmers and Silence

Remmers and his collaborators have prepared a series of attitude scale similar to Thurstone method to a large extent. The special feature of this method is that it measure an attitude towards a large group of objects, person or institution. Thus, attitude toward different things may be measured with a single scale through this method. Too much generalization of fact has been attempted in this method. However, psychologists point out that this method may not be carried out so efficiently as it possible with specific attitude test method. This scale deals among others, with national and racial groups, vocations, teachers, social actions and school subjects.

Evaluation of Attitude scales

Attitude scales are used to measure the degree of positive or negative feeling associated with any slogan, person, institution, religion, political party etc. Attitude scales are also used in public opinion surveys in order to make some important and crucial

decisions. Educationists, for example, conduct opinion surveys to find out how people feel about educational issues.

An attitude scale is designed to provide a valid, or accurate, measure of an individual's attitude towards different thing or subjects. The validity of decisions is directly related to the truthfulness of the responses made by individuals. However, there are shortcomings in these self-report scales of attitudes which are direct method of attitude testing. However, the most common problem is that of social desirability. There is no guarantee that individuals are expressing their true attitude. Their responses to generally represent "socially acceptable" attitude and often motivated to give replies that make them appear "well adjusted", unprejudiced, open minded etc. Similarly, respondents who harbor a negative attitude towards a particular group may not wish be admit to the experimenter (or to themselves) that they have these feelings. Consequently, responses on attitude scales are not always 100% valid. Thus, whenever attitude scales are used effort must be made to reduce potential problems by giving appropriate directions to the test takers.

11.8.2 Rating scales

In this scale the experimenter gives the individual's attitude, a position on a scale extending from the highest to the lowest quality. In this method the individual is given a relative position on the scale, in comparison with other individuals. Thus, in this method attitudes are evaluated on the basis of judgement or decision of the experimenter. The experimenter employs various sources to collect data to give his decision, such as--

- Nonverbal behaviour toward a particular thing, such as, accepting or rejecting customs. But a person is not able to show his behaviour according to his attitude.
- Verbal statement is a good source for understanding attitude. However, it may be kept in mind that a person should express

his attitude faithfully. Thus, verbal statements at many places throws light on our attitudes.

- Personal documents: Personal diary, letter, autobiography etc. give expression to our attitudes. But these records are not easily available for all
- Secondary expressive clues such as facial expression, voice also help in gaining knowledge about attitude. It is a common belief that "what a man say may be less revealing than how he says it."
- Clinical type interview: In this method the interviewer, the psychologist tries to know the attitude of a person through conversation.
- Projective technique: The dormant attitude lying in the unconscious come to light through this method.

From these sources the individual gets enough data concerning the attitude of an individual. In attitude testing rating scales are of two kinds - **relative and absolute.**

11.8.2 (a) Rank order Scales

In this scale, the experimenter gives the individual's attitude, a position on a scale extending from highest to lowest quality. In this method, the individual is given a relative position on the scale. In comparison with other individuals.

11.8.2 (b) Percentage of Population scales

In this scale, one or more individuals are given a position in the population. In this method a definite population is put to the test and then the opinion of one individual is examined. The examiner places the opinion of the individual on the scale, which indicates the percentage of the population concurring with his opinion.



CHECK YOUR PROGRESS

Q 3: What is an attitude Scale?

.....

.....

Q 4: State the names of two commonly used attitude scale.

.....

.....

Q 5: Write the name of the first attitude scale.

.....

.....

Q 6: Why 'Thurstone scale' is known as Equal -appearing intervals?

.....

.....

Q 7: Explain the scoring procedure of a Likert Type scale.

.....

.....

Q 8: What are the uses of attitude scale?

.....

.....

Q 9: Mention a limitations of attitude scales.

.....

.....

11.8.3 Indirect Scales

To avoid the problem of social desirability associated with direct method of attitude testing various indirect measures of attitudes have been used. Some person tries to hide his attitude consciously or unconsciously. Hence, psychologist have favoured the indirect

methods for the measurement of attitude in which he cannot guess the tendencies being measured. Therefore, the subject answers the questions unhesitatingly. This method reveals the presence or absence of certain attitudes. It, however, does not reveal intensity of the attitude. Moreover, this method require more careful planning on the basis of experiments.

11.8.3 (a) Projective Techniques

Indirect methods typically involve the use of a projective test. They are based on the Freudian psychoanalytic notion that the person will "project" his or her views, opinions or attitudes into the ambiguous situation, thus revealing the attitudes the person holds. Thus, a projective test involves presenting a person with an ambiguous (i.e. unclear) or incomplete stimulus (e.g. picture or words). The stimulus requires interpretation from the person. The person's attitude is inferred from their interpretation of the ambiguous or incomplete stimulus. It is claimed that these tests provides genuine data, unbiased by people guessing intent of measurement. It is argued that the tests can be used to measure such factors as the affective, usually hidden, component of attitudes.

Evaluation of Indirect Methods

The advantages of indirect techniques of attitude measurement are that they are less likely to produce socially desirable responses, the person is unlikely to guess what is being measured and behavior should be natural and reliable.

However, indirect methods only provide general information and do not offer a precise measurement of attitude strength since it is qualitative rather than quantitative. The major criticism of indirect methods is their lack of objectivity. Such methods are unscientific and do not objectively measure attitudes in the same way as a Likert scale.

There is also the ethical problem of deception as often the

person does not know that their attitude is actually being studied when using indirect methods.

In addition to the above scales and tests , there some more techniques, such as, Error Choice Technique, Free Response, Technique, Paired Comparisons, Opinion Polling or Surveying, diaries, auto-biographic etc. used for testing attitude. In short, we have to mention that there are various methods of attitude testing. But only one method is not sufficient to give a reliable and clear picture of the attitude of an individual. Attitude is a complex affair which cannot be wholly measured by any single numerical index. Again people's attitudes are subjected to change. Thus, it becomes essential that the attitudes are tested after some reasonable intervals.



11.9: LET US SUM UP

Dear Learners, after going through this unit, it can be summarized as-

- Attitude testing helps us in getting information about work preference of the students.
- Attitudes are learned tendency to evaluate things, issues, people or events in a certain way.
- The attitude is the voluntary disposition of a person. These are significant for satisfaction and success in the educational and vocational fields. Thus, attitude testing is an important part of the comprehensive process of educational evaluation.
- The methods by which, generally, attitudes have been measured are-(1) Attitude scales (2) Rating scales (3) Indirect scales.
- There are many forms of attitude scale for testing attitude. These scales are differed in respect of the method of their construction and kind. The main among them are the followings - Thurstone's scale: and likert scale.

11.10 FURTHER READING

1) Aggarwal, J.C. (1993) : Educational Vocational Guidance and Counselling, Doaba House New Delhi
2) Anastasi Anne (1982): Psychological Testing 4th Edition, New York: Macmillan Publishing Co. Inc.
3) Anderson, Scarvia B. Samuel Ball, Murphy, R.T. and et.al (1976): Encyclopedia of Educational Evaluation, Jessey-Bass Publishers.
4) Asthana Bipin (2015): Measurement and Evaluation in in Psychology and Education, Binod Pustak Mandir, Agra.
5) Brown Frederick, G. (1976): Principles of Educational and Psychological Testing, Holt: Rineheart and Winston.
6) Freeman, F.S.(1965): Theory and Practice of Psychological Testing, Oxford and IBH.
7) Goswami Marami (2010): Measurement and Evaluation in Education and Psychology, Nilkamal Publishing House, Hyderabad.
8) Gronlund, Norman, E. (1968): Readings in Measurement and Evaluation, Education, and Psychology, The Macmillan Company.
9) Mehrens, William, A. and Lehmann, Irvin, J. (1978): Measurement and Evaluation in Education and Psychology, Holt, Rinehard and Winston.

11.11. ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Attitude is a psychological tendency to respond consistently positively or negatively towards certain idea, object, person, or situation.

Ans to Q No 2: Attitudes are learned or acquire/ guide or motivator of behaviour / an expression of favour or disfavor / mental and neural state of readiness

Ans to Q No 3: Attitude scales are self-report inventories designed to measure the extent to which an individual has favourable or unfourable feelings towards some persons, objects, institutions or ideas.

Ans to Q No 4: Thurstone's scale and Likert scale

Ans to Q No 5: Thurstone's scale

Ans to Q No 6: It seeks to analyze people on a continuous scale having equal-appearing units. For this reason this technique is also known as the method of "Equal -appearing intervals."

Ans to Q No 7: To score the scale the item score is obtained by assigning arbitrary weights of 5, 4, 3, 2 and 1 for Strongly Agree (SA), Agree (A), Undecided (U), Disagree (D) and Strongly Disagree (SD) for the statements favouring a point of view. On the other hand, the scoring weights of 1, 2, 3, 4, and 5 are used for the respective responses for statements opposing this point of view.

Ans to Q No 8: Help to measure the degree of positive or negative feeling associated with different things, uses in public opinion surveys.

Ans to Q No 9: The most common problem is that of social desirability.



11.12. MODEL QUESTIONS

A. Short Questions (Answer within 150 words)

Q 1: What is attitude?

Q 2: How attitude is measured?

Q 3: What is attitude testing?

Q 4: Discuss the need of attitude testing.

Q 5: Write notes on.

(a) Thurstone's attitude scale (b) Likert's attitude scale

(c) Master attitude scale (d) indirect scales.

(e) Strong Vocational Interest Bank (SVIS)

(f) Kuder preference Record (KPR)

B. Long Questions. (Answer 300-500 words)

Q 1: What are attitudes? Show your acquaintance with a standard test of measuring attitude.

Q 2: Discuss different techniques of scaling attitude.

Q 3: Describe the construction procedure of Thurstone's attitude scale.

Q 4: Describe any one method of construction of attitude scale.

*** ***** ***

UNIT-12 MEASUREMENT OF INTEREST

UNIT STRUCTURE

12.1 Learning Objectives

12.2 Introduction

12.3 Meaning of Interest

12.4 Definitions of Interest

12.5 Types of Interest

12.6 Nature of Interest

12.7 Measurement of Interests

12.8 Techniques of measuring Interest

- 12.8.1 Formal Assessment Techniques
- 12.8.1(a) Strong Vocational Interest Blank (SVIB)
- 12.8.1(b) The Kuder Preference Record (KPR)
- 12.8.1(c) Thurston's Vocational Interest Schedule
- 12.8.2 Informal Assessment Techniques

12.9 Evaluation of Interest Inventory

12.10 Importance of Interest Testing

12.11 Let Us Sum Up

12.12 Further Reading

12.13 Answers to check your progress

12.14 Model Questions

12.1 LEARNING OBJECTIVES

After going through this unit, learners will be able to-

- explain meaning and concept of interest
- discuss interest testing
- discuss different interest inventories
- describe the uses of interest inventories
- explain the importance of interest testing
- highlight the cautions to be taken in using Interest inventories

12.2 INTRODUCTION

Learners, you might have observed that some of your friends show more inclination towards mathematics, while others towards computers, some develop interest in literary activities or in painting etc. Interest is an expression of our likes and dislikes or our attractions or aversions .When we are free to do some work or perform something, we will choose only those in which we are interested. An individual chooses the most acceptable, suitable alternative out of many, go after preferred objectives, activities etc. and consequently get satisfaction, success and happiness out of the activities selected. It is because of his or her natural liking for these activities over others. Interest refers to activities that an individual likes to engage and enjoy. Interest are major factors in educational and vocational success. They also forecast satisfaction and dissatisfaction. These are motivator of our behaviour. The study of interest would help the teacher to better understand the students in terms of his/her likes and dislikes. Learners, therefore, in this unit you will learn about interest testing.

12.3 MEANING OF INTEREST

Interests are acquired motivations. Etymologically, the word interest in Latin language means 'it matters'. The pupil is interested in the examination results because he is vitally concerned about it. Interest basically refers to the motivating force of an individual which impels him to attend a person, a thing, an object or an activity. It is a subjective attitude motivating a person to perform a certain task. Interest means what matters or concerns us. Any volitional act can take place when we are interested in it. If someone is interested in 'football', it means 'football' matters to him. It is to some extent permanent aspects of mental life. Being interested in something implies possession of a disposition in the mental structure about a particular object or activity. We can describe interest as a feeling of association. It gives pleasure and satisfaction. It results in curiosity towards the object of interest and strength of will to face difficulties while engaged in

the task of one's interest. It brings a definite change in behaviour in the presence of the object characterized by attention and concentration.

We can explain interest as the mental attitude. The distinction between attitude and interest are very narrow. Interests indicates the degree to which the individual prefers to hold an object before his consciousness, whether he acts approvingly or disapprovingly toward the object, while attitude indicate his reaction in terms of direction. Interests give clues regarding adjustment and personality. Strong who has devised interest test batteries, says that interest is behaviour organized around activities. Interest is similar to attitude as attitudes are organized around concepts (and not activities).

12.4 DEFINITIONS OF INTEREST

Let us now study a few definitions of 'interest'-

- **Jones states,** "Interest is a feeling of likening associated with a reaction, either actual or imagined to a specific thing or situation."
- **Bingham defines:** "Interest is a tendency to become absorbed in an experience and to continue it, while an aversion is a tendency to turn away from it to something else."
- **According to Strong:** "Interests are the sum total of likes and dislikes for a wide range of stimulus objects and activities"
- Murphy **states:** "Interests are conditioned stimuli related to goal objects and expressed as likes and dislikes of activities, objects, characteristics or people in the environment"
- C**row and Crow** state, "Interest may refer to the motivating force that impels us to attend to a person, a thing, or an activity; or it may be the affective experience that has been stimulated by the activity itself".
- **I. L. Russel** describes interest as organismic conditions which lead to continuing stimulation of concern about particular objects, persons and activities.
- **Sawrey and Telford** explain. "Interesting objects and experiences

are those which have satisfied or which promise to satisfy motives".

- **Drever** said, "An interest is a disposition in its dynamic aspect."
- According to **J.S. Ross,** "A thing that interests us is just something that concerns us or matters to us."

When we analyze all the above definitions, we find that the psychological term 'interest' is related to a number of facts. Firstly, interest is a mode of experience of the individual. Stout makes it more specific by saying that it is affective-conative aspects of experience. A person interested in sports, has a particular feeling towards it (affective side) and strives to attend to it (conative side). Secondly, this experience is not a temporary one, lasting only in the presence of the object of interest. It is enduring. The sportsman is interested in sports, even when he is not in the sport field. There rests an enduring disposition in the mental structure of the individual round about the object of interest. Thirdly, interest is the basis of conative act in the form of attention. We attend to a certain act because we are interested in it. "Interest is latent attention and attention is interest in action." While interest is the enduring mental disposition, attention is the experience succeeding it.

12.5 TYPES OF INTEREST

Super has suggested three types of interests-Expressed Interests, Manifested Interests and Tested or measured Interests

- **Expressed Interest:** Expressed interest refers to interest in an object, activity, subject of study or occupation as verbally expressed by an individual. In the expressed interest the person expresses his personal likings through verbal communication with such sentences as 'I love music'. It is the first source of knowing the interest of a person. But much reliance cannot be based on it. Because permanency of such expressions of likings are prone to vary from time to time depending upon the maturity of the person.
- **Manifested Interest:** Manifested Interest refers to the actual participation of an individual in a particular activity or occupation.

Interests are the traits of personality of an individual .These traits are manifested as likes and dislikes, preferences and indifferences. Learners, therefore, quite often we say 'I am interested in literature', 'I am interested in photography' 'I am interested in sports' and so on. Manifest interest is the interest that is not expressed but observed by others while the person is engaged and absorbed in an activity. Newton forgot his meals while engaged in scientific experiments.

- **Tested/ Measured Interest:** It refers to an individual's interest as measured by interest inventories.

12.6 NATURE OF INTEREST

From above discussion we can draw the following importance points regarding interest-

- Interests are expressed as likes or dislikes of activities, objects, characteristics or people in the environment.
- Interest is not an inborn quality. We acquire interest in different situations of our life. But some interest are basically inborn in character.
- Interest is satisfying i.e. pursuits of one's interest is always satisfying.
- Interests are conditioned stimuli related to goal objects.
- Interests are an aspect of personality development shaped by both hereditary and environmental factors.
- Interests never become permanently fixed. They get changed as a result of growth and development, learning and different internal as well as environmental factors.
- Interest is a great motivating force. It is the cause and result of motivation.
- Interests are unique to an individual. Individual differences in interests are evident even from the pre-school years.
- Interests are closely related with abilities. If an individual has got an interest along with an ability there is possibility of more progress.
- Interest of a child is determined by his social environment.

- Interests are considered as latent attentions and attentions are treated as interest in action.
- Interests are measurable with the help of different types of interest test.
- Interests are affective-conative aspects of experience. This experience is not a temporary one. It is enduring.



CHECK YOUR PROGRESS

Q 1: Mention two points indicating nature of interest.

.....

.....

Q 2: Distinguish between attitude and interest.

.....

.....

Q 3: What is expressed and manifested interest?

.....

.....

.....

12.7 MEASUREMENT OF INTERESTS

Interest is classified into three types as - expressed interest, manifest interest and measured interest. While knowing the interests of a person we can rely more on the manifest interest and the measured interest rather than the expressed interest. But it is not always practicable to know the manifest interest for want of frequent situations arrange able wherein the observer can observe the manifestation. Hence measured interest is the usual source of interest testing. The interest measurement is the estimate and account of a person's interest as revealed by some psychological tests

or interest inventories. There are various methods of measuring interests, such as interviews, check lists, direct observations, questionnaires and inventories. Questionnaires and inventories are the most important methods of measuring interests.

12.8 TECHNIQUES OF MEASURING INTEREST

Generally, pupil believe that an individual's interest can be ascertained by asking him about their interest. But psychologist have developed different a number of test to measure interest of a person. There are various methods of measuring interests, such as interviews, check lists, direct observations, questionnaires and inventories. Interest can also be assessed informally in the classroom in the same ways as attitude. The tools for measurement of interest are of two types -

- Formal Assessment Techniques and
- In-formal Assessment Techniques

The formal methods are specialized and standardized measuring instruments such as interest inventories, interest test batteries etc. The informal methods include the person's own statement, a record of his activities and observation by the parents and the teachers. The former i.e., the formal methods are usually supplemented by the informal methods in interest testing.

12.8.1 Formal Assessment Techniques: Interest Inventories

Interest inventories are formal techniques for measuring interests. Inventories are the most important methods of measuring interests. A standardized Interest Inventory asks an individual to indicate personal likes and dislikes such as the kinds of activities she or he prefers to engage in. They are considered more reliable at discovering an individual's interest. The inventories address a large number of questions representing a broad range of careers. An ideal testing programme to measure and identify interest includes a

standardized Interest Inventory. A standardized Interest Inventory asks an individual to indicate personal likes and dislikes such as the kinds of activities she or he prefers to engage in. Responses are generally compared to known interest pattern. Interest indicates a simple level what a person likes to do and enjoy. It is important to remember that interest inventories indicate the student's interest and not his or her strength. There are many inventories now in use and majority of them deal with occupational interest. The emphasis of interest inventories is on self-exploration. One should know about one self and be able to relate this information to what one is doing. Thus, the results of interest inventories should not be seen in an isolated manner but should be supplemented with information from other techniques such as observation, interview, informal talk etc.

There are many inventories now in use and majority of them deal with occupational interest. Two frequently used inventories are Strong Vocational Interest Bank (SVIS) and Kuder preference Record (KPR). These are significant milestones in the area of interest measurement and a number of revisions of the original forms have been undertaken. Three notable formal methods universally employed are:

- **Strong Vocational Interest Blank,**
- **Kuder Preference Record, and**
- **Thurstone's Vocational Interest Schedule.**

A brief description of each has been discussed in the following section.

12.8.1 (a) Strong Vocational Interest Blank (SVIB)

The Strong Interest Inventory is a psychological instrument that began its formation in the years following World War- I through the work of E.K. Strong of Stanford University California. Strong went on to publish the first version of the Inventory in 1927. The Strong Interest Inventory test is an assessment that helps people match their interests with potential educational, career, and leisure activities,

using an individual's preferences in a variety of areas to aid them in discovering what they will most enjoy doing with their work and their free time. The purpose of the Inventory was based around studying the occupational interests of men and women. The SVIB is based on the assumption that a person who has the interest typical of successful people in a given occupation will enjoy and find satisfaction in that occupation. It is most useful with persons at least 17 years old. It is available in two form one form for men and one for women. Prof. Strong of Stanford University California designed this in the form of a check list. The check list contains 400 separate items. The items consist of occupations, amusement activities, school subjects, personal characteristics, etc. It is presented to the individual and he is simply asked to indicate whether he likes, dislikes or is indifferent, on a three point scale.

Thus, SVIS measures interest in a number of professional and business fields. As such it is useful for both educational and vocational guidance. These 400 items include 100 occupations, 49 recreations, 36 school subjects, 48 activities and 47 peculiar interests. The test reveals the interest maturity of the individual, his masculinity and of femininity, and his occupational level. It presents numerous activities and the respondent indicates whether she or he likes, dislikes, or has no opinion regarding each activity. Hence it is called category-response type of assessment. These responses are then compared to those of persons in various professionals and business occupations. The inventory can be scored for 40 occupations. Major revisions of the inventory occurred in 1981, 1985 and in 1994. These major revisions created opportunities to update and revise the main sections of the instrument. The Strong Interest Inventory Test is further broken down into four categories:

- **The General Occupational Themes:** Created to provide an economical organization for the Strong Interest Inventory Profile, detailing an individual's general attitude toward specific broad interest fields.

- **The Basic Interest Scales:** Created to provide additional means of explaining high and low scores on the General Occupational Theme Scales through measuring an individual's penchant for certain subject areas and activities.
- **The Occupational Scales:** The original scales developed by E. K. Strong that form the foundation of the Strong Interest Inventory test, which relate specific occupations with various interests and qualities that individuals in each occupation identify with.
- **The Personal Style Scales:** The newest additions to the inventory starting with the 1994 edition. Added to provide a means to measure comfort with broad styles of living and working, including areas such as how a person prefers to be taught, how an individual finds that they best accomplish tasks related to work and whether they prefer working alone or with other people, how an individual feels about taking a management role, and how an individual feels about taking risks.

A final component of the Strong Interest Inventory is a series of indexes that allow interpreters and individuals to understand more atypical profiles that may not conform directly to the scales of the Strong Interest Inventory. The information gathered from these scales comes together to provide us with a picture of our interests and which jobs or school majors match these interests. Each scale is ranked by the separate score. There is no provision for calculating total score for the whole scale. These scales give us a bird's eye view of our interests.

12.8.1 (b) Kuder Preference Record (KPR)

Kuder Preference Record has been prepared by G. Frederic Kuder. Kuder inventories are designed for use with children from grade 9 onwards and with adults. They are in the forms of three Preference Records. In 1938, Kuder compiled Kuder Preference Record. It

was revised in 1951.It has served as a useful interest inventory since then. The KPR constructed by G. Frederic Kuder, assess preferences for specific activities. This record measures preferences or interest in broad occupational areas. This test covers a wider field, comprising of nine separate scales of occupations, viz. mechanical, computational, scientific, persuasive, artistic, literary, musical, social and clerical. Kuder presupposes three major interests viz. mechanical, literary and artistic. So when the same task is presented to the subject, with three related activities, the subject will select the activity that relates one of the three interests that he possesses. It is made up of 504 items. Each item lists three activities and respondents indicate which activity they would most like to engage in. Each item contains three choices. The subject has to select one of the three choices as his first choice, and another as his third choice. An example of the choices given in an item are:

- Build bird houses
- Write articles about birds
- Drawn sketches of birds.

This particular item aims to test three types of interests - mechanical, literacy and an artistic as mentioned above. If the subject opts for the first, his interest is mechanical. Another example is presented. Here the subject is asked to select the activity that he would prefer the most, and the activity he would prefer the least out of the following three:

- Visit an art gallery.
- Browse in a library.
- Visit a museum.

A triple activity regarding collections is:

- Collect autographs.
- Collect coins.
- Collect butterflies.

A detailed scoring system is employed for analysis and interpretation. A percentile of 75 or above is considered significantly

high. If a person goes beyond P75 in any of the areas, all the occupations in that area are attractive for him.

The main difference in the two types of inventories, Kuder Preference Record (KPR) and Strong Vocational Interest Blank (SVIB) is that in Kuder preference record the subject is forced to say which one of the three activities he likes best and which one he likes least, even though he may not like any of one of these choices. Hence it is called forced choice pattern type. In Strong Vocational Interest blank, the subject has to differentiate between L (like), I (indifferent) or D (dislike). Hence it is called category-response type of assessment. Both the interest inventories however better in terms of approaches in construction and purpose for which they are used. If you want a description of a person's so that inferences could be done regarding suitability for one of the very large number of jobs, Kuder Preference Record is your preferable tool of interest assessment. On the other hand if the purpose is to appraise the individual interest for a limited number of specific jobs, then strong Vocational Interest Blank is more useful. It is important to remember that the SVIB and KPR tell nothing about the ability or aptitude of an individual. These remains to be assessed by other methods.

In India Allahabad Bureau prepared a vocational interest preference reckoner in 1956 on the line of Kuder Preference Record. In the same years, Educational and Vocational Bureau in Bihar compiled an interest preference reckoner again on the lines of Kuder Preference Record.

12.8.1 (c) Thurston's Vocational Interest Schedule

This test has been devised by Thurstone. He administered a comprehensive test to 3400 college students who expressed their Likeness (L), Indifference (I) and Dislike (D) to each of the items in the test. He analyzed the test scores and through the techniques of factor analysis, arrived at 8 factors of interest viz.; (i) Commercial

Interest, (ii) Legal, (iii) Athletic, (iv) Academic, (v) Descriptive, (vi) Biological, (vii) Physical Science, (viii) Art. This interest schedule is more effective in inducing thoughtful self-appraisal on the part of the student. This is more informative, convenient to use and is more economical of time and money.

Some less used interest tests are Hepner's Vocational Interest Quotient and Lufburrow's Interest Blank. The latter is of the same design as Strong's Blank. It gives eleven families instead of nine, viz. Artistic, Commercial, Constructions, Industrial, Scientific, Humanistic literary, Transpositional, Mechanical and Technical. Clecton's vocational Interest Inventory deal with 9 groups of occupations and contains 630 items. There are two more inventories which are specially prepared for women. Manson's Occupational Interest Bank is for guiding women. Stewart and Brainard have prepared specific Interest Inventories of types, one each for girls, women, boys and men. Any type (say for girls) contains 20 groups of five questions dealing with no different interests.

Some interest inventory developed in India are -

- **P.S.M. Inventory:** This inventory has been prepared by Vocational Guidance Bureau, Prantiya Shikshan Mahavidyalaya, Jabalpur. It is a verbal inventory which can be given both to individuals and groups in the age range of 13- 18 years.
- **Chatterjee's Non-language Preference Record Form 962:** This is a nonverbal record which can be administered to groups. It measures interest in ten interest areas-fine arts, literary, scientific, agriculture, technical, craft, outdoor, sports, medical and household works.
- **R.R Singh's Interest Record Card:** This self-administering Interest Record measures interest in seven areas-mechanical, business, scientific, social, clerical, aesthetic and outdoor.



CHECK YOUR PROGRESS

Q 4: Which type of interest is the source of interest testing?

.....

.....

Q 5: Distinguish between formal and informal techniques of interest testing.

.....

.....

Q 6: Write the name of two interest inventory.

.....

.....

Q 7: Write the name of the scales included in Kuder Preference Record.

.....

.....

Q 8: What factors are included in Thurston's Vocational Interest Schedule?

.....

.....

Q 9: Mention the name of two Indian interest inventory.

.....

.....

Q 10: Write name of two Indian agency which develop interest inventory in India the line of Kuder Preference Record.

.....

.....

12.8.2 Informal Assessment Techniques

You have read about different types of interest inventories. Besides, these some tools are also available for use to test interest of an individual. It includes expression and observation of interests of an individual. The interviews of an individual regarding his/her interests provide a good beginning for understanding his/her client. For example, the books s/he reads, and the amount of time spent on reading provides an indication of interests of the person concerned. Learners, you can supplement this information by using qualitative techniques of assessment such as observation of the activities a person in or information gathered from anecdotal records and written works as sources informal information regarding interests of a person.

It is important to remember interests indicate the clients and not his or her strength. The emphasis of interest inventories is on self-exploration. One should know about one self and be able to relate this information to what one is doing. Thus the results of interest inventories should not be seen in an isolated manner but should be supplemented with information from other techniques such as observation, interview, informal talk etc. interest results are of little value themselves. They should be considered with achievement and aptitude scores to help individuals make educational and career choices.

12.9 Evaluation of Interest Inventories

Uses of Interest Inventories:

Results of interest inventories or tests are particularly useful to the counsellor as-

- Interest inventories are needed as they enable students to have a better choice in ease of adopting subjects, courses, occupations and recreational activities, etc.

- They are very much useful for employers to select a student for a favourable occupation in the concerned occupational set-up.
- They are highly useful for the teachers and guidance workers to ascertain students' performance in his field of interest.
- They are also useful for the parents as they insist them to help their children for their liking field.
- They assist the counselor, teacher and guidance worker in dealing with problems of the children.
- In comparison to other psychological devices they are economical.
- The items of interest inventories are non-threatening and so the student is more likely to respond honestly.
- The student, with the assistance of the counselor, is helped to see the discrepancy, if any, between his/her view of one's interests and those assessed by the interest inventories.
- It also provides the counsellor a view into whether the clients expressed interests are shifted to what s/he is pursuing.

Limitations of interest inventories:

- One cannot rely on scores on interest forever be also cause preferences, liking may change. Intensity of interest may vary. The tests presuppose that the subject possesses a particular interest. But it can reveal the interest that is present at the time of test, and not afterwards. The interests revealed may not remain permanent. Moreover the interests are cultivable also. At the time of testing a particular interest may not have developed fully, but it may develop afterwards. It has been seen that some interests develop during the vocation.
- Some of the tests reveal ability rather than interest. But interest is not the same thing as ability. So some tests are not fully valid or reliable.
- Interest inventories are not very satisfactory in predicting job success, job satisfaction, personality, and adjustment and academic success.
- The questions in the inventories deal with certain types of activities, and not all these lead to clear-cut vocations. Again, there is much

overlapping between one activity and another. An occupation is not one interest but a combination of activities or interests.

- The interest inventories reveal facts on the basis of the report given by the subject. The accuracy of the report is still a problem. Some people do not reveal facts.
- Interest inventory can be affected by the ambiguity of questions asked.

Cautions on the use of Interest inventories:

Learners, you should keep in the following cautions in mind while using the results of interest inventories.

- Interest data of an individual provides a profile of the pattern of their interests (what they like doing) and should not be taken as the strength of the individual.
- Interest results tell us only about the likelihood of an individual finding enjoyment in a particular type of activity or occupation and not s/he is capable of.
- A high score on an interest inventory so should not be considered analogues to success in a career.

12.10 IMPORTANCE OF INTEREST TESTING

Interests give clues regarding adjustment and personality. The knowledge and understanding of a child's interests are a major factor in educational and vocational guidance. Thus, interest testing is important in educational and vocational fields. Knowing student's interest can be of great value for a teacher in promoting motivation and making learning activities relevant to students. The study of interest would help us to better understand the students in terms of his/her likes and dislikes. Interest assessment is used for several different purposes. Career exploration, that leads to decisions such as choosing a major, selecting a career, career choice predictions or career changes etc. Within this context, college and high school counseling services are the most typical providers of interest assessment and career counseling experiences. However, employment

agencies, vocational rehabilitation services, social service agencies, corporations, consulting firms, and community agencies also incorporate interest assessment. Interest testing is done to achieve some purpose such as:

- To help teachers, counselors and parents to assist the students to prepare his educational and vocational terms consistent with his interest.
- To help in the selection of the right person for the right work, and thus reduce frustration, unhappiness and disappointment in the life of the individual and increase productive capacity of individual.
- To assist students to have a better choice in ease of adopting subjects, courses, occupations and recreational activities, etc.
- To enable teachers, counselors assist the counselor, teacher and guidance worker in dealing with problems of the children.
- They are also useful for the parents as they insist them to help their children for their liking field.



12.11 LET US SUM UP

Dear Learners the unit can be summarized as-

- Assessment of interest indicates a simple level what a person likes to do and enjoy.
- Interest testing is important in educational and vocational fields. As it helps a teacher in promoting motivation and making learning activities relevant to students.
- We can use two types of techniques to discover a student's interest - (i) formal assessment techniques and (ii) informal assessment techniques.
- Interest inventories are formal assessment techniques of interest. Three frequently used inventories are Strong Vocational Interest Bank (SVIS) and Kuder preference Record (KPR) and Thurston's Vocational Interest Schedule.
- The emphasis of interest inventories is on self-exploration.

- Interest inventories are needed as they enable students to have a better choice in ease of adopting subjects, courses, occupations and recreational activities, etc.
- One cannot rely on scores on interest or attitude testing forever because preferences, likings or disliking may change.
- Interest data of an individual provides a profile of the pattern of their interests (what they like doing) and should not be taken as the strength of the individual.

12.12 FURTHER READING

1) Aggarwal, J.C. (1993) : Educational Vocational Guidance and Counselling, Doaba House New Delhi
2) Anastasi Anne (1982): Psychological Testing, 4th Edition, New York: Macmillan Publishing Co. Inc.
3) Anderson, Scarvia B. Samuel Ball, Murphy, R.T. and et.al (1976): Encyclopedia of Educational Evaluation, Jessey-Bass Publishers.
4) Asthana B (2015): Measurement and Evaluation in in Psychology and Education, Binod Pustak Mandir , Agra.
5) Brown F. G. (1976): Principles of Educational and Psychological Testing, Holt: Rineheart and Winston.
6) Goswami M (2010): Measurement and Evaluation in Education and Psychology, Nilkamal Publishing House, Hyderabad.
7) Gronlund, Norman, E. (1968): Readings in Measurement and Evaluation, Education, and Psychology,The Macmillan Company.
8) Mehrens, William, A. and Lehmann, Irvin, J (1978): Measurement and Evaluation in Education and Psychology, Holt, Rinehard and Winston.

12.13 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: i) Interest is shaped by both hereditary and environmental factors, ii) It is closely related with abilities etc.

Ans to Q No 2 : Interests indicates the degree which the individual prefers to hold an object before his consciousness, whether he acts approvingly

or disapprovingly toward the object, while attitude indicate his reaction in terms of direction.

Ans to Q No 3: Expressed interest refers to interest in an object, activity, subject of study or occupation as verbally expressed by an individual and Manifested Interest refers to the actual participation of an individual in a particular activity or occupation.

Ans to Q No 4: Measured interest is the usual source of interest testing.

Ans to Q No 5: : The formal methods are specialized and standardized measuring instruments such as interest inventories, interest test batteries etc. The informal methods include the person's own statement, a record of his activities and observation by the parents and the teachers.

Ans to Q No 6: Strong Vocational Interest Blank and Kuder Preference Record,

Ans to Q No 7: Nine separate scales of occupations, viz. mechanical, computational, scientific, persuasive, artistic, literary, musical, and social and clerical

Ans to Q No 8: 8 factors are (i) Commercial Interest, (ii) Legal, (iii) Athletic, (iv) Academic, (v) Descriptive, (vi) Biological, (vii) Physical Science, (viii) Art.

Ans to Q No 9: P.S.M. Inventory and Chatterjee's Non-language Preference Record Form 962

Ans to Q No 10: : Allahabad Bureau and Educational and Vocational Bureau in Bihar

12.14. MODEL QUESTIONS

A. Short Questions (Answer each question in around 150 words)

Q 1: Why is measurement of interest so important?

Q 2: Write a note on Interest Inventor

Q 3: Write short notes on

(a) Strong Vocational Interest Bank (SVIS)

(b) Kuder preference Record (KPR)

Q 4: Show your acquaintance with a test of measuring interest.

B. Long Questions. (Answer each question in around 300-500 words)

Q 1: What is interest inventories? State the uses and limitation of interest inventory.

Q 2: Describe any one commonly used Interest test Inventory in educational measurement.

*** ***** ***

UNIT-13 MEASUREMENT OF PERSONALITY

UNIT STRUCTURE

13.1 Learning Objectives

13.2 Introduction

13.3 Methods of Assessing Personality

- 13.3.1 Subjective Methods
- 13.3.2 Objectives Methods
- 13.3.3 Projective Methods

13.4 Let Us Sum Up

13.5 Further Reading

13.6 Answers to check your progress

13.7 Model Questions

13.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- identify the different methods of personality assessment
- discuss the different techniques used in personality assessment
- Identify different subjective, objective and projective methods of personality assessment

13.2 INTRODUCTION

You must be curious to know about your own personality or that of others. It is natural from our part to know what type of personality or personality traits are possessed by us or others. For that you should have the knowledge and skill for the assessment of personality. It is easy to understand some traits of someone's personality but assessing his or her personality is a difficult task. In your earlier semester, you have come across the concept of personality and its theories. As you know that personality is a dynamic phenomenon and it is ever in a process of change and modification. In spite of that there are various methods and techniques for assessing personality which can help us make this difficult task easy. In this unit, we

will discuss various subjective, objective and projective methods or techniques used for assessing personality.

13.3 METHODS OF ASSESSING PERSONALITY

As mentioned above, personality assessment is a difficult task. Nevertheless, psychologists have been trying to develop various methods or techniques for assessing personality. These methods may be broadly classified as below:

i) Objective methods
ii) Subjective methods
iii) Projective methods

Methods and techniques of personality assessment may also be categorised as standardised and non-standardised. Some experts also use the terms such as testing and non-testing techniques to assess the personality. Self-reporting methods like autobiography, interviews, questionnaires and personal inventory are also used in which the person to be assessed, reports about himself. On the contrary, some methods such as rating scales, biographies, case history, etc., are those in which assessment of personality is done on the basis of what others feel about an individual. In some cases, we may use behavioural measures for assessing personality in which the actual behaviour of an individual is observed in real life situations. We may also use projective techniques by which we can find how an individual reacts to an imaginative situation involving fantasy. Lastly, there are techniques by which we can determine some personality variables in terms of physiological responses by machines and technical instruments. We may categorise the different personality assessment methods as depicted below in the Table-13.1

Table 13.1: List of Personality Assessment Methods

Standardised Tests (Objective Methods)	Non-standardised Tests (Subjective Methods)	Projective Techniques
1. Achievement Tests 2. Aptitude Tests 3. Intelligence Test 4. Interest Tests 5. Personal Inventories	1. Anecdotal Records 2. Autobiography 3. Case study 4. Cumulative Record Card 5. Interview 6. Observation 7. Parent's Views 8. Personal Data Blanks 9. Questionnaire 10. Rating Scale or Attitude Scale 11. Schedule 12. Score Card 13. Sociometry	1. BG Test 2. Blocky Picture Test 3. Children's Apperception Test 4. Clay Modelling 5. Cloud Picture Test 6. Draw-a-Man Test 7. Graphology Test 8. Mosaic Test 9. Psychodrama 10. Rorschach's Ink Blot Test 11. Sentence Completion Test 12. Socio-drama 13. Szondi Test 14. Thematic Apperception Test 15. Toy and Doll Playing Test(word game) 16. Word Association Test (Free and Controlled)

Source: Aggarwal, J.C.(2015). Essentials of Educational Psychology. New Delhi:Vikash Publishing House Pvt. Ltd. P.376

From the Table-13.1, it is obvious that a large number of methods and techniques are available to assess personality. But, it is not possible to give details of all the methods or techniques here. So, we are going to discuss only a few select methods and techniques in the following sub-sections of this unit. We shall discuss subjective methods first.

13.3.1 SUBJECTIVE METHODS

The subjective methods are those in which the individual is allowed to express or disclose what s/he knows about himself/herself as an

object of observation. They are based on what the subject himself or herself to say about his/her traits, attitudes, personal experiences, aims, needs and interests.

(a) Case-Study Method

A Case-study thoroughly investigates the family history, medical history, social environment and other relevant facts. This method helps to diagnose the causes of mal-adjustment. This is basically a clinical method which helps the psychiatric social worker to deal with the patient. Some important headings of data collection for case-study are as below:

- General information: Name, age, sex, address
- Present Problems: Symptoms and diseases
- Health History: Diseases from which s/he suffered in the past
- Educational History: progress done in school and colleges
- Inter-personal relationship: the subject's behaviour towards others
- Psycho-sexual relationship: Sexual habits, sexual relation, disease, if any
- Personality Traits: Moods, sentiments, and reaction of the subject.

(b) Interview

Interviewing is a commonly used method of collecting information directly from subject or interviewee. It is a dynamic face-to-face relationship. The purpose of personality assessment interview of a student or a child is to find out the status of development of his/her personality. It can be used effectively to collect information about an individual. For that the interviewer should have a balanced outlook towards the interaction and have a good plan for the interview. The interviewer should prepare for the interview by collecting all data concerning the interviewee. S/he should set clear objectives for the interview. Establishing a good rapport with the interviewee in the beginning of the interview is very important for the interviewer to

extract real information from the respondent. Rapport does mean here the feelings of friendliness, security and mutual confidence between the interviewer and the interviewee. The interviewer should frame questions with proper phrasing and asks sequentially to the interviewee. The interviewer should give useful comments and elaboration during the interview in need. The environment of the interview must be very warm, reciprocal and congenial one. Imposition of discipline during the interview makes the interaction very mechanical. So, no discipline should impose to the interviewee. To conduct an interview successfully, the interviewer needs adequate training and practice.

Types of Interview

Interview may be of many types. In the context of personality assessment we may classify it into the following types:

(i) Introductory Interviews: This is like an ice-breaking session in which a good rapport develops between the interviewer and the interviewee. This is basically a pre-interview interaction to familiarise to each other, build up an atmosphere of friendliness and removes emotional blocks for setting the platform for future interaction.

(ii) Fact-finding Interviews: The main purpose of these interviews is to extract such information which cannot be obtained by other methods. Some important information such as the interviewee's attitude, interest and feelings may be found out through this kind of interaction.

(iii) Informative Interviews: The interviewer collects information about the interviewee from different sources. The purpose of this interaction is to inform this information to the interviewee. For example, interviewee may be informed about his/her abilities and interests.

(iv) Therapeutic Interviews: These types of interview are used for purpose of treatment. There are three types of such interviews:

- **Counsellor-directed interview:** Counsellor himself plans the solution of the patient on the basis of the interaction.
- **Non-directive interview:** The subject (interviewee) himself takes the lead and during the course of interview identifies his/her own desires and wishes.
- **Electric interview:** In this particular interaction both the above approach is used to treat the interviewee.

(c) Observation

Observation is widely used in assessing personality. It helps to study the behaviour pattern of an individual in actual life situation. Normally, an observer first decides what personality traits or characteristics of a particular person or a group to be observed and accordingly s/he observes relevant activities of the subject in real life situation. This observation may be done in two ways- participant observation and non-participant observation. In participant observation, the observer more or less becomes part of the group under observation. On the other hand, in non-participant observation the observer takes a position at place for observation where his presence is not disturbing or influencing the behaviour of the individual under observation. The observer may use tape-recorder, photographic camera, video camera, telescope to get clear idea of the observation. The reliability of the observation may be established by going for repeated observation of the individual or group under observation in the same situation a number of times. The subject may be observed by a number of observers and results may be pooled together. While making observation, the following principles to be kept in mind:

- The whole situation should be observed.
- Observe one student at a time.
- Students should be observed in different situations, such as in classroom, on the play ground or while interacting with their peers etc.
- Observation should be made over a period of time.

How to plan and execute the observation?

(i) Planning Stage

Proper planning of observation is the key for successfully getting desired results from it. So, a good planning for observation is necessary before going out to observe somebody. The following points should be kept in mind while someone plan for observation to assess an individual's personality:

- Specific activities or behaviours that are to be observed during observation should be identified and defined well in advance.
- An appropriate group of subjects should select for observation.
- Decide whether go for individual observation or group observation.
- The length of the observation period, number of periods and interval between periods should be fixed.
- The form of recording should be decided.
- Decided the instruments to be used for observation.
- Observer should decide the place to take the physical position from where s/he observes the activities/behaviour of the individual or group under observation.
- The tools to be used for recording should be kept handy.

(ii) Execution of Observation

To translate the plan of observation into the field well, the observer should have appropriate skills and resourcefulness, which depends upon on the following points:

- Properly arrange the special condition under which the subjects are being observed.
- Take proper physical position by the observer for observation so that neither his presence influences the behaviour of the subject under consideration nor he finds any difficulty to observe the subjects.
- The observer should minutely observe the each specific activity or behaviour that is shown by the subjects under observation.

- Observing directly the length and the number of periods and interval decided upon.
- Properly utilise the instruments used for recording the observation.
- Taking instant decision during observation in need and execute that decision without breaking the flow of observation, for example, changing the physical position of the observer or introducing something new stimulus to the observed condition etc.

(iii) Recording of Observation

According to the nature of the activities or behaviour of the individual or group, the method for recording should be employed. Normally, two methods may be employed for recording. The skill of the observer also plays a key role in deciding on the method.

The first method is to record the observation on the spot during observation. But, at times, it may not be feasible when the action or activity performed is very swift. This may likely to distract the subject or subjects.

In the second method, facts may be recorded immediately after the observation is over. This method is useful because this does not distract the mind of the subjects. One risk involves in this method is that the investigator may not be able to recall facts accurately after the interval of a few minutes.

(iv) Device Used in Observation

Check-lists, Rating scale, Score Cards, Blank form of tallying frequencies

(v) Interpretation:

Results of the observation should be cautiously interpreted after taking into account various limitations of planning, sampling or procedure.

(d) Rating Scales

A rating scale is a technique which is used to know from others where an individual stands in terms of certain personality traits. The

ratings are done by parents, teachers, a board of interviewers and judges and by self as well. It reflects the impression the subject has made upon the person who rates him or her. The three basic things involved in this technique are:

- The specific trait or traits to be rated.
- The scale by which degree of possession or absence of the trait has to be shown.
- The appropriate person or judges for rating.

The very first step requires the selection of the traits or characteristics, which have to be evaluated by the judges, are to be stated and defined clearly. Following this, a scale for rating work is to be constructed. Suppose you wish to have rating on the 'Level of Need for Achievement' among the students of a class. You can make division of this quality into degrees such as very high, high, average, low, very low etc. Now, arrangement of these divisions along a line at equal intervals, from high to low or otherwise will be named as Rating Scale for assessing the level of Need for Achievement. Usually the degrees are indicated by numbers, 1 to 3 to 5 to 7, comprising three points, five points or seven points scale. For example, the seven points scale is of the following type-

7	6	5	4	3	2	1
Excellent	Very high	High	Average	Below average	Low	Very Low

Now the raters (say teachers), who are in a position to rate the individual (say students) properly, can be asked to rate them and give them scores, rating from 1 to 7, according to the degree of Need for Achievement they (students) possess.

Principles Governing Rating Scales

- The specific trait or mode of behaviour to be rated must be defined properly.
- The scale should be clearly defined, i.e., whether ratings are at a 3-,4-, 5- or 7- point scale.
- The trait to be evaluated should be readily observable.

- Uniform standards of rating scale should be observed.
- The observer should observe rates in different situations for the traits to be rated in.
- The number of characteristics to be rated should be limited.
- In the rating scale card, some space may be provided for the rater to write some supplementary materials.
- The direction of using the rating scales should be clear and comprehensive.
- Several judges may be employed to increase the reliability of any rating scale.
- Well-informed and experienced persons should be selected for rating.



CHECK YOUR PROGRESS

Q 1: What is the purpose of using therapeutic interviews?

.....

.....

Q 2: Briefly describe how do you record the observation.

.....

.....

.....

Personality is sometimes measured by inventories. A personality inventory is basically a standard set of questions about some aspects of the individual's life history, feeling, preferences or activities, presented in a standard way and scored with a standard scoring key. It is like a questionnaire in several aspects except two aspects: Firstly, the questionnaire is a general device and can be used for collecting all kinds of information. On the other hand, a personality inventory is specially designed to collect information about an individual personality. Secondly, the questions set in the

questionnaire are generally addressable to the second person like:

Do you often feel anxiety? Yes/No?

In case of the personality inventory, questions are usually addressed to the first person such as:

I often feel anxiety. Yes/No.

Description of Some Famous Personality Inventory

A large number of personality inventories have been developed since 1918. R.S. Woodworth, for the first time prepared an inventory namely 'Personal Data Sheet' which was consisted of 116 questions in 1918 for finding neurotic tendency among soldiers. R. G. Bernreuter prepared an inventory containing 125 items to measure neurotic tendency, self-sufficiency, introversion, extroversion, dominance and submission. Likewise, Flabagan, Guilford and many others have also prepared inventories. These inventories have basically been developed in America and Western Countries, notably in Britain etc. Personality inventories have developed and used in India also, for example, Aligarh University Vocational Guidance Centre has used adjustment inventories. Now we shall describe a few of the most important inventories in brief.

i) **Bell Adjustment Inventory:** There are two forms of this inventory available, one for high school and college students, and other for adults. At the preparation stage of this inventory, 223 items of Thurston Personality Schedule were used and additionally 188 items added. The final inventory consists of 140 items after eliminating unsuitable items. This inventory has been designed to measure adjustment of students in four areas: (i) home adjustment, (ii) health adjustment, (iii) social adjustment and (iv) emotional adjustment. There are 35 items in each area. The reliability of this inventory is very high. The inventory contains items such as, "Do you day-dream frequently?"

ii) **Allport and Allport's A-S Inventory:** G.W. Allport and F.H. Allport conducted a study with a view "to discover the disposition of an individual to dominate his fellow (or to be dominated by them) in

various face to face relationships of everyday life." The study is available in two forms- for man and women. The study for men has 33 items and that for women 34 items. Each item begins with a brief description of a situation which we usually encounter at a meeting, in school, on a bus or in other familiar settings. In these items, some situations are presented verbally. The responses to these items are given. The subject is asked to indicate one of the two or four alternative ways, listed for meeting the situation for each item. The total score gives an idea whether ascendance is dominant in the subject or submission. The study is self-administered and no time limit is prescribed. The purpose of the study should not disclosed to the subject till the administration is complete, otherwise responses may be affected.

Reliability and validity coefficients have been provided in the manual of the inventory. Split-half reliability for men was .85 and for women .90. Test-retest reliabilities are also high. Moderate validity coefficients have been obtained. Rating was used as a criterion. The study has proved helpful in self-knowledge and in vocational selection and scores guidance.

iii) The Minnesota Multiphasic Personality Inventory (MMPI):

This inventory was published in 1943. It is one of the most widely used inventories in America. It consists of 550 items and can be applied to persons of 16 years of age or above. Each item is printed on a separate card and the subject sorts it into three groups: 'True, False, and Cannot Say.' Items are classified into 26 heads such as family, general health, religious attitudes, phobias, delusions, etc. Items can be grouped in separate scales to measure nine personality traits- hypochondriasis(Hs), depression(D), hysteria(D), psychopathic deviate(Pd), masculinity-femininity(Mf), paranoia(Pa), psychasthenia(Pt), schizophrenia(Sc), and hypomania(Ma). Their clear description have been given- (1) Hs-physical complaints, seclusive,

apathetic not orderly, indecisive; (2) D- shy not energetic moody, not relaxed; (3) Hy-physical complaint, flattering, friendly, trustful; (4) Pd-sociable, frivolous, incoherent, clever, dishonest, not practical, not friendly; (5) Mf- unrealistic, not popular, unemotional etc., are symptoms of masculinity, while worldly, sensitive, shy, etc., those of feminity; (6) Pt- kind, quite, soft-hearted, dependant, irritable; (7) Sc- depressed, constantly in the same mood, escapist; (8) Pa- arrogant, hard-hearted, sociable, shrewd; (9) Ma- boastful, enterprising, jealous, popular, energetic, self-confident.

Some items of the inventory are: "I wish I could be as happy as others seem to be." Items are answered as 'T', 'F' or 'Cannot Say'. Various keys have been supplied such as psychiatric discriminative keys for various scale, control keys and supplementary keys.

The M.M.P.I. is mainly used for clinical and diagnostic purpose. It is also suitable to distinguish between pathological and normal person. It has high predictive validity. Details for administration and scoring are given for both individual (card) form and group (booklet) form separately. Scoring in the booklet form can be done by hand as well as machine. In machine scoring, seventeen keys are used. It is desirable that all items be answered. The shortened version consisting of 336 items is also available for emergency use.



CHECK YOUR PROGRESS

Q 3: What do you mean by personality inventory?

.....

.....

Q 4: Briefly describe about Bell Adjustment Inventory.

.....

13.3.3 PROJECTIVE METHODS

In our earlier discussion, we have discussed various methods or techniques which evaluate the conscious or over behaviour of an individual. Bur covert or unconscious behaviour of an individual is also very significant to understand his/her personality in totality. In this regard, Sigmund Freud believes that our conscious behaviour is only one-tenth of our total behaviour. So, there should be some techniques which not only stress on the observable part of human personality but also can reveal a person's inner or private world and go deeper in the unconscious of an individual and consequently ascertain the repressed feelings, wishes, desires, fears, hopes and ambitions, etc. These techniques, on account of their using projection phenomena are called projective techniques. According to Anne Anastasi, "It is expected that the test material of projective techniques will serve as 'a sort of screen' upon which the subject projects his characteristics ideas, attitudes, aspirations, fears, worries, aggression and the like.' Ruth Strang observed, "Projective techniques are a method of understanding the inner world of the individual."

Projective techniques are very useful with young children, illiterates and person with language handicaps or speech defects. Non-verbal media is readily applicable to all these groups.

Common Characteristics of Projective Techniques

The most distinguishing feature of projective techniques is in their assignment of relatively unstructured task to the subject. The task in projective technique permits almost unlimited variety of possible responses. Brief general instructions are provided in such tests to allow free play to the individual's fantasy. That is why, the test stimuli is ambiguous and weakly structured. The subject is expected to provide meaning, significance, organisation, or some other ways leave the impression of his personality on the situation. An attempt is made to explore the psychological reality or the

underlying basic personality features of the individual such as his hopes, aspirations, needs, motives, moods, attitudes, conflicts, complexes, and fear, etc.

Projective techniques represent disguised testing procedure. As the subjects or examinees do not know how their responses will be psychologically interpreted, Projective techniques are characterised by a global approach to the appraisal of personality. These techniques are regarded very effective in revealing covert, latent or unconscious aspects of personality.

Projective techniques are not only concerned with emotional, motivational and inter-personal characteristics, but also with some intellectual aspects of behaviour.

Kinds of Projective Techniques

There are many standardised projective tests which are used widely by the psychologists across the world. They include Rorschach's Ink Blot Test, Thematic Apperception Test and various other picture tests like Children's Apperception Test, Blocky Pictures, the Michigan Picture Test, Rosenzweig Picture-Frustration Study, etc. Out of these tests, Rorschach's Ink Blot Test and Thematic Apperception Test (TAT) are very popular and widely used for personality assessment. Now, we discuss in details below about these two most popular tests.

(i) **Rorschach's Ink Blot Tests:** Harmann Rorschach was a Swiss psychologist who experiment with the use of ink blots as a means of diagnosing mental disorders in 1921. Perceptual approach was the basis of this test. The perception of an individual was influenced by the emotional and social make-up when s/he was asked to perceive a figure which was not well defined.

Material of the test consists of 10 cards with ink blots. Five of them are in black and white and five are multicoloured. These ink-blots are commonly unstructured, that is the shapes of the blots do not have any specific meaning as shown in Fig. 13.1:



Fig. 13.1: Two inkblots of the type used in Rorschach inkblot test

ADMINISTRATION OF THE TEST:-

The test is administered individually. Each card is given to the subject, who in turn is asked what he sees in the ink blots, what that means to him and what that might be. The subject has to tell the movement or motion if perceived.

The subject is allowed as much time as s/he wants for a given card and is permitted to give as many responses to it as s/he wishes. The subject is also allowed to turn the card around and look at it from any angle to find things in it.

The examiner notes down the time taken for each response, position in which cards are being held, spontaneous remarks, emotional expression and other incidental behaviour of the subject during the test session.

After all the cards have been presented, the second phase of inquiry follows. It is intended to seek clarification or addition to original responses.

SCORING:-

There are several systems of scoring and interpreting the Rorschach. But the most common scoring categories include location, response determinants and content.

1. **Location:** This includes the total area perceived by the examinee. Whether s/he perceives the whole ink-bolt (W) sub-division (D) or unusual detail (Dd).

2. **Determinant:** This category includes characteristics of the bolt, e.g. movement and colour etc. Different determinants are depicted by following symbols:
   - **Form (F):** The intellect, thinking capacity, contact with reality etc. are known through form. When a good form is perceived F+ is given. In case of blurred form F- is given.
   - **Colour (C):** The emotions are depicted through colour. When form is main and colour is also mentioned then FC category is scored, when colour dominates the response and form becomes secondary then CF category is given, and in case of pure colour only C is given.
   - **Shading (Y):** Shading may be perceived as representing depth, texture, hazy forms such as clouds. It also tells about the depression, joy etc. of the examinee.
   - **Vista (V):** Under this category, FV, VF and V are scored, depending on the performance of figure or height and depth.
   - **Texture (T):** This aspect represents the want for love.
   - **Movement (M):** M+ is indicative of creativity, while M- point towards fantasy.
3. **Contents:** Certain major scoring categories are employed under this head. Chief among these are human figure, human details (or part human figures), animal figures, animal details, and anatomical diagrams.

In addition to these three categories, popular response and organisation responses are also scored.

INTERPRETATION:-

The interpretation of this test is quite different. Movement responses show intellectualism and inter-psychic, i.e., inwards living. Persons with strong M responses are inventive and introspective, while those who low M responses are stubborn and practical-minded. Colour responses indicate emotion to outward stimuli. 'Form' denotes realistic outlook. These interpretations should not, however, be made in an absolute way. The whole family background, training, health,

past life etc. should be taken into account. The interpreter should be trained and should have long experience. Inference, analogy, impression, all may be useful in the interpretation.

(ii) Thematic Apperception Test (TAT): This test was devised by C. D. Morgan and Henry Murray. It was first introduced by Henry Murray to measure the need for achievement. Later, it was fully developed for the assessment of personality with the help of the psychologist Morgan. The test consists of perception of certain picture in a thematic manner, i.e. revealing imaginative things. The test involves a systematic use of pictures representing a number of dramatic events. It consists of 30 pictures.

Test Materials and Administration: As stated above, the test material consists of 30 pictures which portray human beings in a variety of actual life situations. In the original 30-picture tests, 10 pictures are for men, 10 pictures are for women and 10 pictures are for both sex. The maximum number of pictures used on any one subject is thus, twenty. The test is usually administered in two sessions, using ten pictures in each session. Generally, it takes about an hour to administer the test. The pictures are presented to subject one at a time. Pictures are vague and indefinite. At the very outset, the subject is told clearly that this is a test of creative imagination and that there is no right and wrong response. After that pictures are shown to the subject one after another and the subject is asked to make up a story for each of the picture keeping in mind the following aspects while knitting the story:

- What is going on the picture?
- What has led to this scene?
- What would happen in such a situation?



Fig. 13.2 Sample pictures from TAT

The Examiner encourages the subject to create as dramatic story as possible. It is important for the story to have a title, a beginning, a body, a hero and the end. The general idea is that with his/her story, the subject reveal his/her needs, desires, along with their hidden motivations, expectations and intentions, which otherwise would not be out for display on the open.

Scoring and Interpretation: The present system of scoring and interpretation of TAT stories has been conducted by taking into account the following aspects:

- **Hero of the story:** What type of personality does he have?
- **Theme of the story:** What is the nature of the theme or plot used in making the story?
- **Style of the story:** Length of the story, language used, direct or indirect expression, forced or poor expression, organisation of the contents, originality and creativity etc.
- **The content of the story:** What interests, sentiments and attitudes do they depict? In which manner (reality or fantasy) has the behaviour been expressed? What inner state of the mind does the story reveal?
- **Test situation as a whole:** The subject's reaction to be listed as whole.
- **Particular emphasis or omissions:** The omission, addition, distortion, and attention to the particular details,

- Subject's attitude towards authority and sex.
- Outcome: Conclusion of the story-happy, unhappy, comedy etc.

The T.A.T. test is more useful in knowing general personality dynamics, rather than the diagnostic aspects. The test has good reliability. Stability coefficient are, however, not very high. It has moderate validity.



CHECK YOUR PROGRESS

Q 5: How do you administer the Rorschach's Ink Blot Tests?

.....

.....

Q 6: What are the three aspects the subject of T.A.T. has to keep in mind while creating his/her story?

.....

.....



13.4: LET US SUM UP

The unit can be summarized as-

- These methods personality assessment may be broadly classified under Objective methods, Subjective methods, and Projective methods.
- The subjective methods are those in which the individual is allowed to express or disclose what s/he knows about himself/herself as an object of observation. They are based on what the subject himself or herself to say about his/her traits, attitudes, personal experiences, aims, needs and interests.
- Case-study thoroughly investigates the family history, medical history, social environment and other relevant facts.
- The purpose of personality assessment interview of a student or a child is to find out the status of development of his/her personality.

- In Non-directive interview, the subject (interviewee) himself takes the lead and during the course of interview identifies his/her own desires and wishes.
- Observation is widely used in assessing personality. It helps to study the behaviour pattern of an individual in actual life situation.
- According to the nature of the activities or behaviour of the individual or group, the method for recording observation should be employed
- A rating scale is a technique which is used to know from others where an individual stands in terms of certain personality traits.
- A personality inventory is basically a standard set of questions about some aspects of the individual's life history, feeling, preferences or activities, presented in a standard way and scored with a standard scoring key.
- Bell Adjustment Inventory has been designed to measure adjustment of students in four areas: (i) home adjustment, (ii) health adjustment, (iii) social adjustment and (iv) emotional adjustment.
- The Minnesota Multiphasic Personality Inventory (MMPI) was published in 1943. It is one of the most widely used inventories in America.
- The M.M.P.I. is mainly used for clinical and diagnostic purpose. It is also suitable to distinguish between pathological and normal person. It has high predictive validity.
- Sigmund Freud believes that our conscious behaviour is only one-tenth of our total behaviour.
- Ruth Strang observed, "Projective techniques are a method of understanding the inner world of the individual."
- The most distinguishing feature of projective techniques is in their assignment of relatively unstructured task. The task in projective technique permits almost unlimited variety of possible responses.
- Thematic Apperception Test (TAT) was devised by C. D. Morgan and Henry Murray
- The T.A.T. test is more useful in knowing general personality dynamics, rather than the diagnostic aspects.

13.5 FURTHER READING

1) Aggarwal, J.C. (2015): Essentials of Educational Psychology, New Delhi: Vikash Publishing House Pvt. Ltd.
2) Asthana, B. (2015): Measurement and Evaluation in Psychology and Education, Agra: Shri Vinod Pustak Madir.
3) Mangal, S.K. (2006): Advanced Educational Psychology, New Delhi: Prentice-Hall of India Pvt. Ltd.
4) Mangal, S.K. (2000): Essentials of Educational Psychology, New Delhi: Prentice-Hall of India Pvt. Ltd.
5) Sing, A.K. (2008): Tests, Measurement and Research Methods in Behavioural Sciences, New Delhi: Bharati Bhawan.

13.6 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Therapeutic interviews interview are used for purpose of treatment.

Ans to Q No 2: According to the nature of the activities or behaviour of the individual or group, the method for recording should be employed. Normally, two methods may be employed for recording. The skill of the observer also plays a key role in deciding on the method.

- The first method is to record the observation on the spot during observation. But, at times, it may not be feasible when the action or activity performed is very swift. This is also likely to distract the subject or subjects.
- In the second method, facts may be recorded immediately after the observation is over. This method is useful because this does not distract the mind of the subjects. One risk involves in this method is that the investigator may not be able to recall facts accurately after the interval of a few minutes.

Ans to Q No 3: Personality inventory is basically a standard set of questions about some aspects of the individual's life history, feeling, preferences or activities, presented in a standard way and scored with a standard scoring key.

Ans to Q No 4: Bell Adjustment Inventory is available in two forms, one for high school and college students, and other for adults. At the preparation stage of this inventory, 223 items of Thurston Personality Schedule were used and additionally 188 items added. The final inventory consists of 140 items after eliminating unsuitable items. This inventory has been designed to measure adjustment of students in four areas: (i) home adjustment, (ii) health adjustment, (iii) social adjustment and (iv) emotional adjustment. There are 35 items in each area. The reliability of this inventory is very high. The inventory contains items such as, "Do you day-dream frequently?"

Ans to Q No 5: The test is administered individually. Each card is given to the subject, who in turn is asked what he sees in the ink blots, what that means to him and what that might be. The subject has to tell the movement or motion if perceived.

- The subject is allowed as much time as s/he wants for a given card and is permitted to give as many responses to it as s/he wishes. The subject is also allowed to turn the card around and look at it from any angle to find things in it.
- The examiner notes down the time taken for each response, position in which cards are being held, spontaneous remarks, emotional expression and other incidental behaviour of the subject during the test session.
- After all the cards have been presented, the second phase of inquiry follows. It is intended to seek clarification or addition to original responses.

Ans to Q No 6: The subject is asked to make up a story for each of the picture keeping in mind the following aspects while knitting the story: (i) What is going on the picture? (ii) What has led to this scene? (iii) What would happen in such a situation?

13.7. MODEL QUESTIONS

A. Short Questions (Answer each question in around 150 words)

Q 1: Briefly describe about Case-study method.

Q 2: How fact-finding interviews help in assessing personality?

Q 3: What do you mean by rating scale?

Q 4: What do you understand by projective techniques?

B. Long Questions (Answer each question in around 300-500 words)

Q 1: Discuss briefly about the Minnesota Multiphasic Personality Inventory (MMPI).

Q 2: Discuss about the common characteristics of Projective techniques.

Q 3: Explain how to conduct the Thematic Apperception Test (TAT).

*** ***** ***

UNIT-14 MEASUREMENT OF CREATIVITY

UNIT STRUCTURE

14.1 Learning Objectives

14.2 Introduction

14.3 Meaning of creativity

14.4 Criteria of Creative Personality

14.5 Measurement of Creative Thinking

14.6 Measurement of Creative Thinking for Different Levels of Students.

14.7 Let Us Sum Up

14.8 Further Reading

14.9 Answers to check your progress

14.10 Model Questions

14.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- understand the meaning of creativity
- know the different criteria of creative personality
- comprehend the measurement process of creative thinking
- understand the process of measuring creative thinking for different level of students

14.2 INTRODUCTION

In this unit, you will be introduced to different psychological attributes of human beings and how they are measured. Measurement of such attributes of a person reveals his overall personality. In our surroundings or in a classroom environment, we find that some learners or persons are quite different from others in terms of their thinking and ideals. They do not think or perceive the world in a generalized view. Their perception towards the world is somewhat different from the general. The psychological attributes that identifies people as somewhat different in terms of his thinking

and perception is creativity. As we know that creativity is the most valued qualities of human beings which enormously effect all spheres of human live. Thus this unit will introduce you with creativity, different criteria of creative personality and measurement of creative thinking for different level students.

14.3 MEANING OF CREATIVITY

Creativity is an ability to be innovative, unusual, or be different from others. It is an ability to give novel responses, new answers and establish new relationships. One of the most important and accepted skills of our day is originality and usefulness. Creative thinking is problem solving from a different sense. Because there is a problem as the problem cannot be solved by conventional means. Its solution requires creating new ways, or divergent thinking and side thinking. However, it is hard to have a systematic creative thought and this is related to it being divergent. Some selected definitions of Creativity are as follows-

PISA 2021 defined Creativity thinking as the competence to engage productively in the generation, evaluation and improvement of ideas that can result in original and effective solutions, advances in knowledge and impactful expressions of imagination.

Amabile and Pratt (2016) explains that, "Creativity is the production of novel and useful ideas and innovation as implementation of creative ideas".

Plucker, Beghetto and Dow (2004): "Creativity is the interaction among aptitude, process, and environment by which an individual or group produces a perceptible product that is both novel and useful as defined within a social context", reflecting its multidimensional and social nature.

Sternberg and Lubart (1999): "Creativity is the ability to produce work that is both novel (i.e., original, unexpected) and appropriate (i.e., useful, adaptive concerning task constraints)"

Feist (1998): "Creative thought or behaviour must be both novel-original and useful-adaptive"

H. Gardner (1993): "The creative individual is a person who regularly solves problems, fashions products, or defines new questions in a domain

in a way that is initially considered novel but that ultimately becomes accepted in a particular cultural setting".

Welsch (1980): Creativity is the process of generating unique products by transformation of existing products. These products, tangible and intangible, must be unique only to the creator, and must meet the criteria of purpose and value established by the creator.

Torrance (1966): "Creativity is a process of becoming sensitive to problems, deficiencies, gaps in knowledge, missing elements, and disharmonies; identifying the difficulty; making guesses, searching for solutions; or formulating hypotheses about the deficiencies: testing hypotheses, and finally communicating the results".

Mednick (1962): "Creativity involves the process by which ideas already in one's mind are associated in unusual but original ways to form new ideas ".

Guilford (1975): "Creativity is a form of problem solving and distinguished between two types of cognitive operations: divergent production and convergent production".



CHECK YOUR PROGRESS

Q 1: State the meaning of Creativity.

.....

Q 2: State how the concept of "creativity" is viewed as by the researchers.

.....

14.4 CRITERIA OF CREATIVE PERSONALITY

J. P. Guilford, well-known in research and writing on creativity, visualizes creative thinking as a subclass of thinking in general. Creativity is seen as intelligent action that creates multiple solutions to problems by original and unique views (divergent) rather than limiting solutions by

intelligence and logic (convergent).Guilford and other educators distinguish between convergent and divergent thinking in the following way. Convergent thinking is using logic and knowledge to decrease the amount of possible outcomes whereas divergent thinking is foreseeing multiple solutions to a problem, having original and unique ideas. Convergent thinking (the kind most frequently encountered in our schools) is aimed toward a single correct answer. Divergent thinking is inquiring, searching around, often leading to unconventional and unexpected answers (the kind which would probably not bring you a high grade on the usual school examination). Guilford said both thinking method being together would be beneficial, but said that creativity thinking is more of a divergent thinking.

In his work Art of Thought, published in 1926, Wallas presented one of the first models of the creative process. In the Wallas stage model, creative insights and illuminations may be explained by a process consisting of 5 stages:

i) Preparation (preparatory work on a problem that focuses the individual's mind on the problem and explores the problem's dimensions)
ii) Incubation (where the problem is internalized into the unconscious mind and nothing appears externally to be happening)
iii) Intimation (the creative person gets a "feeling" that a solution is on its way)
iv) Illumination or Insight (where the creative idea bursts forth from its preconscious processing into conscious awareness)
v) Verification (where the idea is consciously verified, elaborated, and then applied)

Through the earlier discussion in the earlier semester, you have come to know about the taxonomy of educational objectives. As per the taxonomy of educational objectives under cognitive domain, evaluation falls on the top or higher level of the ladder. But, in 1990 Lorin Anderson, a former student of Bloom and a group of cognitive psychologists revised the earlier Bloom's taxonomy was revised and as per the new taxonomy of Bloom, there are six levels of cognitive learning. Each level is

conceptually different. There six levels are-remembering, understanding, applying, analyzing, evaluating and creating. Thus creativity falls on top of cognitive area. Creativity here has four dimensions. Those dimensions also show a levelled structure.

1. **Imitation:** Imitation is the first step of creativity. In this stage people and works done are imitated. A good example can be seen on artistic creativity. A person first imitates the art of famous artists. Then he/she tries to capture the fine techniques there. Creativity is new for this person, but not for the society or history.
2. **Relocation:** In this stage the person starts to use a known object, method or action somewhere else. In a sense this stage is the redemption step from function that is one of the obstacles of creativity. This type of creativity can be seen in children. Children take an object and use it something else, thus having creative behaviour.
3. **Making connections:** this stage is a more complex level of relocating. The person connects two objects that are not related, far from each other. He makes a never-before connection. This is the group which daily life considers as real creativity.
4. **Method development:** In this stage the person has created a new method with creativity and others start to imitate him with that method. This stage makes him a master.

Along with that creativity has eight properties, four cognitive, four affective. Those four cognitive properties are set by Guilford. They are-

1. **Rationalism:** Creating as much solutions as possible
2. **Flexibility:** Thinking in different ways.
3. **Originality:** Creating solutions that are different from others.
4. **Elaboration:** Detailing.

These cognitive properties are supported by affective properties.

5. **Curiosity:** Creative people are curious people.
6. **Complexity:** They are intellectually complex. They handle situations in a different and complex manner.
7. **Risk taking:** They look for unconventional solutions and take risks. They take risks by looking for the yet unknown.

8. **Imagination:** They have strong imaginations



CHECK YOUR PROGRESS

Q 3: Distinguish between convergent thinking and divergent thinking.

.....

Q 4: State the five stages of creative process development.

.....

Q 5: State two cognitive and affective properties of Creativity.

.....

14.5 MEASUREMENT OF CREATIVE THINKING

The growth of creativity research has been associated with looking for dependable instruments to measure it. Based on different definitions of creativity, researchers have developed different instruments for measuring it. Each instrument reflects the conception of its developer regarding the nature of creativity. These instruments are typically classified into four approaches representing the four main categories of creativity definitions: process, product, person and press.

1. **Process approach:** The process approach in creativity measurement focuses on the specific cognitive processes and structures that are conducive to creative production. Divergent thinking tests have been most widely used for measuring creative processes or creativity-relevant skills. Examples of these tests include

i) Wallach-Kogan Creativity Tests (WKCT; Wallach &Kogan, 1965)
ii) Torrance Tests of Creative Thinking (TTCT; Torrance, 1966, 2008)
iiii) Structure of the Intellect Divergent Production Tests (SOI, Guilford, 1967)
iv) Creativity Assessment Packet (CAP; Williams, 1980).

These tests, which are also known as measures of ideational fluency, include open or ill-structured problems that require individuals to generate as many responses as possible, which are then scored to capture fluency (number of responses), originality, flexibility (number of different categories) and elaboration.

2. **Person approach:** This approach for measuring creativity relies heavily on using self-report questionnaires that enquire about personality traits related to creativity. Instruments in this area were designed by investigating the characteristics or interests of individuals who produced creative achievements. This generates a set of distinctive personality traits for creative individuals, including attraction to complexity, high energy, behavioural flexibility, intuition, emotional variability, self-esteem, risk taking, perseverance, independence, introversion, social poise and tolerance to ambiguity. An individual who exhibits these characteristics is more likely to act creatively than an individual without these characteristics. Several questionnaires have been developed in this area, such as
   - How Do You Think? (HDYT; Davis &Subkoviak, 1975),
   - Creative Personality Scale (CPS; Gough, 1979)
   - How Creative Are You? (HCAY; Raudsepp, 1981).

   Other self-report questionnaires ask individuals about their past creative production instead of personality traits, based on the assumption that past accomplishments can provide a dependable indication of subsequent creative accomplishments. Such questionnaires include
   - Creative Behaviour Inventory (CBI; Hocevar, 1979c)
   - Creative Achievement Questionnaire (CAQ; Carson, Peterson, & Higgins, 2005)
   - Biographical Inventory of Creative Behaviours (BICB; Batey, 2007).
3. **Product approach:** The product-based assessment of creativity has been widely recommended for assessing creativity. It has been argued that a comprehensive assessment of an individual's creativity cannot be attained without a measure of its product. The focus within

this approach lies on rating individuals' creative products in different areas, such as writing, art, music, science or mathematics. The Consensual Assessment Technique (CAT), suggested by Amabile (1982), has been frequently used in creativity studies that employed a product-based assessment. The CAT is not built upon any creativity theories; instead, its main idea is that the best evaluators of the quality of creative products are recognized experts in the relevant domain. Accordingly, the validity of the CAT is not subject to the acceptance or rejection of creativity theories whenever additional evidence is gathered. As a result, it is easy for the CAT, unlike other creativity measures, to accommodate diverse perspectives on many of the controversial theoretical issues in creativity research.

4. **Press approach:** The press approach concentrates on the analysis of the work environment or climate where creativity is realized. The measurement of these environmental factors is based on the idea that there is an indirect relationship between these factors and creativity. Previous studies have revealed that environmental factors can significantly affect individuals' intrinsic motivation to engage in an activity, which in turn affects their creativity. Numerous factors have been found to influence the development of creativity in learning environments. In a systematic review of studies relating to creative learning environments, Davies et al. (2013) determined key characteristics of school environments that enhance creativity, including an appropriate physical environment, availability of various materials, using outdoor environments, an exciting pedagogical environment, employing games-based approaches, flexible use of time and supportive relationships between teachers and students. Commonly used instruments in this area include
   - Siegel Scale of Support of Innovation (SSSI; Siegel &Kaemmerer, 1978),
   - College and University Classroom Environment Inventory (CUCEI; Fraser, Treagust, & Dennis, 1986)

- Work Environment Scale (WES; Moos, 1986)
- Work Environment Inventory (WEI; Amabile&Gryskiewicz, 1989)
- KEYS: Assessing the Climate for Creativity (KEYS; Amabile, Conti, Coon, Lazenby, & Herron, 1996)
- Team Climate Inventory (TCI; Anderson & West, 1998)
- Situational Outlook Questionnaire (SOQ; Isaksen, Lauer, &Ekvall, 1999).

Generally, these instruments share some dimensions such as members' perceptions of organizational vision, resources, leadership, freedom, interactions, support for creativity and in-novation and continuous development.



CHECK YOUR PROGRESS

Q 6: State the four approaches of measuring creativity.

.....

Q 7: Name any two tools that measure personality traits related to creativity.

.....

Q 8: Name any two characteristics of school environment that enhance creativity.

.....

14.6 MEASUREMENT OF CREATIVE THINKING FOR DIFFERENT LEVELS OF STUDENTS

There is no single ability or trait to creativity. Indeed, creativity involves a combination of cognitive (information processing), conative (personality traits, motivational aspects), and emotional factors (affective state, trait) that are interacting dynamically with the environment (which stimulates or inhibits the expression of creative potential) resulting in the uniqueness of the creative process and product.

Measurement of creativity are generally grouped into three broad categories:

A. Measurement of Creativity as a Status
B. Measurement of Creativity as a Level
C. Measurement of Creativity as a Style

A. Measurement of Creativity as a Status - "Is this Student Creative?" This question the implicitly refers to creativity as an exceptional process that is rarely observed in the general population; a process which is thus either present or absent. Measures addressing precisely this question therefore focus on the creative outcomes that one may or may not demonstrate. Biographical inventories (e.g., Biographical Inventory for Students) and achievement questionnaires (e.g., Creative Achievement Questionnaire) are also good examples of this category. Typically, these self-report measures focus on past accomplishments in creative domains. The main assumption underlying the measurement of creative achievement is that, if one has already demonstrated creative outcomes, the probability to demonstrate other creative outcomes in the future is high. This type of approach is thus extremely limited when it comes to assess creativity in school settings, because it is rare to find creative achievement per se at the school age level (e.g., a published book or novel or released musical composition). Clearly, these measures are more adapted for adults who already demonstrated valuable creative outcomes, showed identifiable creative accomplishment, talents, expertise or special training in one or several creative domains.

B. Measurement of Creativity as a Level - "How Creative is this Student?" This question the implicitly refers to creativity as a continuous and unidimensional construct with the implicit idea that creativity is present in each individual to various degrees. There are many candidate factors that serve the indicators of creative potential. These factors are of different nature: cognitive, conative, or environmental. Product-based assessment is also commonly used.

- **Measuring Cognitive Aspects:** Cognitive factors involved in creativity refer to basic processes of thought that lead to creative production, which include identifying, defining and redefining the problem, selective encoding (i.e., the ability to input environmental information related to a problem), divergent thinking, evaluation of ideas, associative thinking, and flexibility (see Lubart et al.). However, a common definition of creativity from the cognitive approach makes reference to divergent thinking only. Divergent thinking tests such as the Torrance Tests of Creative Thinking or the Wallach and Kogan tests are standardized tasks which evaluate the capacity of the individual to generate many ideas from a simple starting point, in a limited amount of time.

Other measures of cognitive abilities involved in creativity include measures of problem-solving such as the Purdue Elementary Problem-Solving Inventory which consists of tasks for sub-process aspects of creative problem solving (e.g., problem identification, problem definition, generating questions and guessing causes, clarifying the goal, analyzing details, or verifying solutions).

- **Measuring Conative Aspects:** Conative factors refer to preferred ways of behaving (personality and motivationally related). Conceptually, it is conceived that these factors facilitate (or inhibit) the effective use of cognitive factors involved in the creative process. Many self or informant rating scales and inventories are used such as, the Barron-Welsh Art Scale, or the Creative Perception Inventory in which individuals are asked to auto-evaluate themselves and check personality traits relative to creative personality or characteristics that typify their behaviour. Gough's Adjective Checklist is also used that includes a creative personality scale which captures more broadly the creative personality attributes.

The motivational dimension involved in creativity is also a factor often targeted to capture aspects of creative potential. Moneta and Siu showed that intrinsic motivation in school (measured using the Work

Preference Inventory) predicted greater creativity in a writing task of story composition. Other measures include, for example, the Torrance Creative Motivation Scale.

- **Measuring Environmental Aspects:** Several environmental factors have been found to contribute to creative potential. These include factors such as parental practices, school environment, the effects of teaching methods and even birth order. There are quite a number of observational and/or rating scales for assessing environmental characteristics, classroom atmosphere and learning environments favourable to individual creativity. For example, the Classroom Activities Questionnaire.
- **Product-Based Assessment:** A common way to assess creative potential involves the achievement of a creative product based on initial elements or constraints. Typically, these products are then evaluated by experts of the domain using the Consensual Assessment Technique. Technically, domain-appropriate judges are asked to score independently a series of productions resulting from the same task, using a Likert scale (e.g. from "1- not creative at all", to "7-highly creative"). A statistical consensus is then estimated to ensure inter-rater agreement. The idea behind this procedure is that, if every expert agrees to "classify" the same creative products in the same order for their level of creativity (resulting in high inter-rater agreement), it is possible to conclude that raters have identified the same quality (that is, creativity). When the consensus is reached (as indicated by a statistical criteria), the scores given by each judge to each production can be averaged to derive a composite score of creativity. In other words, the accumulation and convergence of expertise by multiple informants for the same "subjective" quality results in a consensual measure.

C. **Measurement of Creativity as a Style - "How (and where)** is this Student Creative?" This question, built upon the previous tradition of creativity assessment that goes beyond the notion of "status", "degree", or "level" of creative potential, as emphasized in the

"componential-unitary" measurement approach, and even more as in the "eminence" approach, which respectively put emphasis either on a dichotomous or on a unidimensional approach to creativity. Rather, it underlies the "form" that creative potential can take and it implicitly includes both questions "Where is this student creative?" (Creativity as a domain specific ability) and "How the creative potential of this student is expressed across various dimensions?" (Multifaceted concept of creative potential). Evaluation of Potential for Creativity (EPoC) is a modular, domain-specific tool, which presently includes verbal and graphic subtests that measure the two key modes of creative thinking-divergent-exploratory thinking and convergent-integrative thinking-in elementary and middle-school students. It is the first tool in the history of creativity assessment that combines an approach by domain of creative expression and by mode of thinking, in order to capture the diversity of creative abilities of children.

Another approach to evaluating student creativity is the assessment framework developed by Lucas, Claxton, and Spencer (2012) in harmony with the habits and sub-habits of creativity. Lucas, Claxton, and Spencer (2012) proposed a "Five Dimensional Model of Creativity" that identifies 5 core creative habits and 15 sub-habits of creativity. In addition to being a comprehensive model-incorporating aspects of both creativity and critical thinking-there is empirical evidence suggesting that this framework has the potential to be useful for teachers for assessing and tracking the development of students' creativity in schools.

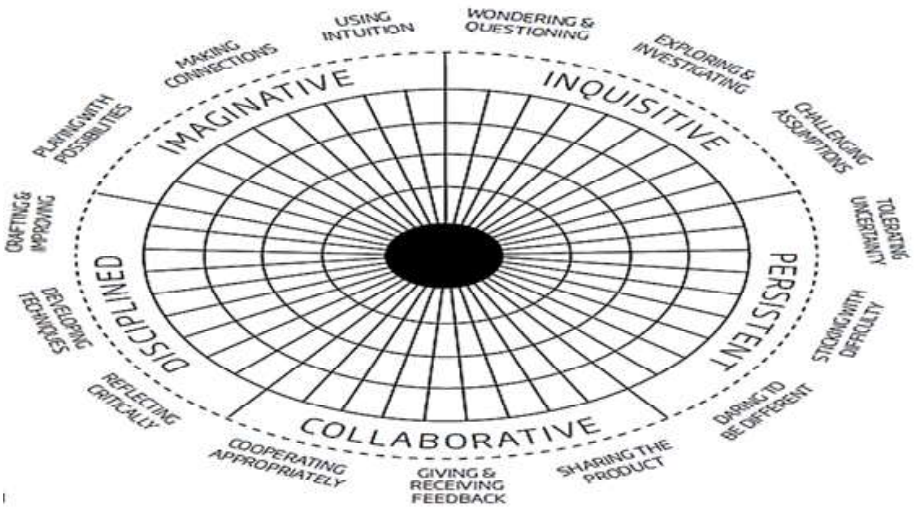


FIGURE 14.1. Creativity Assessment Tool (Lucas, Claxton, & Spencer, 2012,)

1. **Inquisitive:** Creative individuals are good at uncovering and pursing interesting and worthwhile questions in their creative domain.

- **Wondering and Questioning:** Beyond simply being curious about things, the questioning individual poses concrete questions about things. This enables him, and others, to think things through and develop new ideas.
- **Exploring and Investigating:** Questioning things alone does not lead to creativity. The creative individual acts out his curiosity through exploration, and the investigating individual follows up on her questions by actively going out, seeking, and finding out more.
- **Challenging assumptions:** A degree of appropriate scepticism is an important trait of the creative individual. This means not taking things at face value without critical examination.

2. **Persistent:** Creative individuals do not give up easily.

- **Sticking with difficulty:** Persistence in the form of tenacity is an important habit of mind enabling an individual to get beyond familiar ideas and come up with new ones.
- **Daring to be different:** Creativity demands a certain level of self-confidence as a prerequisite for sensible risk-taking as well as toleration of uncertainty.

- **Tolerating uncertainty:** Being able to tolerate uncertainty is important if an individual is going to move 'off of the starting blocks' on a project or task where actions or even goals are not fully set out.

3. **Imaginative:** At the heart of a wide range of analyses of the creative personality is the ability to come up with imaginative solutions and possibilities.

- **Playing with possibilities:** Developing an idea involves manipulating it, trying it out, improving it.
- **Making connections:** This process of synthesising brings together a new amalgam of disparate things.
- **Using intuition:** The use of intuition allows individuals to make new connections and arise at thoughts and ideas that would not necessarily materialise given analytical thinking alone.

4. **Collaborative:** Many current approaches to creativity, such as that of John-Steiner (2006), stress the social and collaborative nature of the creative process.

- **Sharing the product:** This is about the creative output itself impacting beyond its creator.
- **Giving and receiving feedback:** This is the propensity to want to contribute to the ideas of others, and to hear how one's own ideas might be improved.
- **Cooperating appropriately:** The creative individual co-operates appropriately with others. This means working collaboratively as needed, not necessarily all the time.

5. **Disciplined:** As a counterbalance to the 'dreamy', imaginative side of creativity, there is a need for knowledge and craft in shaping the creative product and in developing expertise.

- **Developing techniques:** Skills may be established or novel but the creative individual will practise in order to improve. This is about devoting time to a creative endeavour.

- **Reflecting critically:** Once ideas have been generated, evaluation is important. We could call this 'converging'. It requires decision-making skills.
- **Crafting and improving:** This relates to a sense of taking pride in one's work. The individual pays attention to detail, corrects errors, and makes sure the finished article works perfectly, as it should.

This tool, pictured in Figure 14.1, invites the teacher to map students' habits graphically, thereby providing a formative assessment of creative thinking. Students also use the tool to self-assess their creative habits. Both teachers and students are provided with supporting materials for using the tool, including an online presentation and video explaining the construct of creativity, why assessment can be beneficial, and how they were expected to use the tool.



CHECK YOUR PROGRESS

Q 9: What are the different factors that serve as the indicators of creative potential?

.....

Q 10: Name the five core creative habits identified in the Five Dimensional Model of Creativity.

.....



14.7 LET US SUM UP

After going through this unit, it can be summarized as-

- Creativity is an ability to engage productively in the generation, evaluation and improvement of ideas that results in production of novel and useful ideas and establishment of new relationships.
- Creativity is seen as intelligent action that creates multiple solutions to problems by original and unique views (divergent) rather than limiting solutions by intelligence and logic (convergent).

- According to Wallas stage model, creative process is developed through 5 stages: preparation, incubation, intimation, illumination or insight and verification.
- Creativity here has four dimensions: Imitation, Relocation, Making connections, Method development.
- Creativity has eight properties; four cognitive and four affective.
- Characteristics of creativity are categorised into four categories: Generating Ideas, Digging Deeper into Ideas, Openness and Courage to Explore Ideas, and Listening to One's "Inner Voice."
- Four approaches are considered for classification of measurement tools based on the four main categories of creativity definitions: process, product, person and press.
- Measurement of creativity are generally grouped into three broad categories: Measurement of Creativity as a Status, Measurement of Creativity as a Level, Measurement of Creativity as a Style
- Assessment framework developed by Lucas, Claxton, and Spencer (2012) identifies 5 habits and 15 sub-habits for formative assessment of creative thinking.



14.8 FURTHER READING

1) Bruno-Faria, M., Fleith, D., Alencar, E. (2014): Theory and Practice of Creativity Measurement. United States: Sourcebooks.
2) Baer, J., Kaufman, J. C., Plucker, J. A. (2008): Essentials of Creativity Assessment, Germany: Wiley.
3) Creative Thinking of School Students. (2009): India: Discovery Publishing House.
4) Handbook of Creativity. (1999). United Kingdom: Cambridge University Press.
5) Spencer, E., Lucas, B. (2017): Teaching Creative Thinking: Developing Learners who Generate Ideas and Can Think Critically, United Kingdom: Crown House Publishing.

6) Spencer, E., Claxton, G., Lucas, B. (2012): Progression in Creativity - Developing New Forms of Assessment: A Literature Review, United Kingdom: Unknown Publisher.

7) Sarsani, M. R. (2006): Creativity in Schools, India: Sarup Book Publishers

14.9 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Creativity is an ability to be innovative and produce novel and useful ideas.

Ans to Q No 2: The researchers viewed the concept of "creativity" as (a) creative person,

(b) creative process, (c) creative product, and (d) creative environment.

Ans to Q No 3: Convergent thinking is using logic and knowledge to decrease the amount of possible outcomes whereas divergent thinking is foreseeing multiple solutions to a problem, having original and unique ideas.

Ans to Q No 4: Fivestages of creative process development are

- Preparation
- Incubation
- Intimation
- Illumination
- Verification

Ans to Q No 5: Two cognitive properties of Creativity are Flexibility and Originality.

Two affective properties of Creativity are Curiosity and Imagination.

Ans to Q No 6: The four approaches of measuring creativity are Process approach, Person approach, Product approach and Press approach.

Ans to Q No 7: Name of two tools that measure personality traits related to creativity:

- How Do You Think? (HDYT; Davis & Subkoviak, 1975)
- Creative Personality Scale (CPS; Gough, 1979)

Ans to Q No 8: Two characteristics of school environment that enhance creativity are physical environment and pedagogical environment.

Ans to Q No 9: Cognitive, Conative and Environmental factors serve as the indicators of creative potential.

Ans to Q No 10: Five core creative habits identified in the Five Dimensional Model of Creativity are: Inquisitive, Imaginative, Persistent, Collaborative and Disciplined.



14.10 MODEL QUESTIONS

A. Short Questions (Answer each question in around 150 words)

Q 1: Define the term creativity.

Q 2: Describe the five stages of creative process development.

Q 3: With the help of suitable example illustrate the four dimensions of creativity.

Q 4: List the characteristics of creative personality.

Q 5: State the properties of creativity.

B. Long Questions. (Answer each question in around 300-500 words)

Q 1: Explain the nature of creativity.

Q 2: Make a comparison between four approaches used for classification of measurement tools based on the four main categories of creativity definitions.

Q 3: Discuss the methods used to measure creativity as a status, creativity as a level, and creativity as a style.

Q 4: Discuss the relevance and application of assessment framework developed by Lucas, Claxton, and Spencer (2012) in school education.

*** ***** ***

UNIT-15 EMERGING TRENDS IN EVALUATION

UNIT STRUCTURE

15.1 Learning Objectives

15.2 Introduction

15.3 Question Bank

- 15.3.1 Concept of Question Bank
- 15.3.2 Purposes of Question Bank
- 15.3.3 Construction of Question Bank
- 15.3.4 Use of Question Bank

15.4 Grading System

- 15.4.1 Concept of Grading System
- 15.4.2 Methods of Grading System
- 15.4.3 Advantages and Disadvantages of Grading System

15.5 Semester System

- 15.5.1 Concept of Semester System
- 15.5.2 Components of the Semester System
- 15.5.3 Advantages and Disadvantages of Semester System

15.6 Use of Computer in Evaluation

15.7 Let Us Sum Up

15.8 Further Reading

15.9 Answers to check your progress

15.10 Model Questions

15.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- Know the concept and development of question bank.
- Understand the meaning of grading system.
- Develop understanding about the methods of grading.
- Know the advantages and disadvantages of grading.
- Know the concept and different components of semester system.
- Know the advantages and disadvantages of grading.
- Understand the use of computer in evaluation.

15.2 INTRODUCTION

Evaluation is an integral part of education. At present, evaluation is mainly used in the context of examination. The concept of evaluation is vast in comparison to other concepts such as examination and measurement. Evaluation mainly includes all the factors related to quality and measurement of an individual. To remove different barriers of examination like narrowness and lack of reliability, the concept of evaluation emerged now-a-days. Different educationists of the contemporary time applied innovative measures in evaluation system to widen what is defined as examination system. These innovative measures have been successfully applied in the present day examinations. Thus, in this unit, we shall discuss about the tools such as question bank, grading system, semester system and the use of computer for the process of evaluation.

15.3 QUESTION BANK

Question bank is an important tool in the field of education. Almost every student in their academic life use question bank to develop their academic excellence. Not only students, teachers also use question bank for different purposes. Question banks are very popular among the students who appear for board or university examination.

15.3.1 Concept of Question Bank

The question bank is a collection of questions which are statistically sound and have technical worth. Question bank is also termed as model question paper and item bank. This bank facilitates to select proper questions for a well-designed question paper. A question bank is a planned library of test items designed to fulfil certain predetermined purposes. Question bank should be prepared with utmost care so as to cover the entire prescribed text. It should be exhaustive and cover entire content with different types. It is a collection of relevant questions prepared for a given subject and

useful for students and teachers. They are often made available on the internet.

A question bank should be established only after due planning. Objectives of question bank should be clearly visualised in the question bank for better understanding. Question bank mainly manage to provide utility services to all those who are interested. And for that, suitable space for sorting and classifying of questions are required.

The item bank is a term for a repository of test items that belong to a testing program, as well as all information pertaining to those items. In most applications of testing and assessment, the items are of multiple choice format, but any format can be used.
In the words of Rudner, "An item bank is defined as an organised collection of test items that can be assessed for test development."
According to B. H. Choppin, "An item bank or question bank is a collection of test items organised, classified and catalogues the order to facilitate the construction of a variety of achievement and other types of mental test".

On the basis of meaning and definitions of question bank, some of the characteristics of question bank are mentioned below:

- Question bank is mainly based on sound ideas, purposes and adapted to the particular objective of mind.
- It should be within the range of the students' experience and knowledge.
- Question bank is a continuous process.
- Question bank should present a challenge which stimulates an educative response from the users in keeping with the objectives.
- Question bank should contain only one idea. If it contains many, then it confuses the students.
- Question bank is always based on action.
- Question bank maintains records about the number of question with their answers in all the languages with their corresponding courses, categories and levels.

- Every question bank is mainly prepared for any particular course, level, language, selected category and types of questions.

15.3.2 Purposes of Question Bank

A quality work is not possible without a specific purpose or objective. A Question bank has some purposes which are discussed below:

- One important purpose of question bank is to improve the teaching learning process. Question bank is an important tool through which teacher as well as students will be benefited.
- Another purpose of question bank is to observe the growth of the students. Teachers can develop their instructional efforts with the help of question bank to develop the capability of pupils.
- Enhancement of the evaluation process of education system is another purpose of question bank.
- The question bank can also allow the pool of test items for formative and summative evaluation of the pupil's performance.
- To assess the predetermined objectives of educational system is another purpose of question bank. It is a pool of readymade quality questions which is made available to teachers and examiners. Accordingly teachers or examiners may select appropriate questions to assess predetermined objectives.

15.3.3 Construction of Question Bank

There are certain steps in development of question bank. These steps are discussed below:

- **Blue Print of Question Bank:** For development of question bank, preparation of blue print is essential. If the items are not relevant to the objectives of the programme for which the question pool is being developed, then it results in hodgepodge of question. Blue print always helps to generate a quality question pool.

- **Writing of Questions:** Writing of questions is another important step of question bank development. Following points should be kept in mind while writing questions for question bank.
  - Questions are written on collection from various sources.
  - Readymade questions may be collected from old question papers, from standardised tests and from review exercise of good text books. Specification of questions should be given along with answer key, besides objectives and content area.
  - While writing new questions, experience teachers, examiners and paper setters should be invited and some honorarium should be provided to them for their work. In case of new questions also, specification of questions should be given along with answer key, besides objectives and content area.
  - For reviewing the newly framed questions, a workshop should be organised and question setters should be invited. This get together helps to discuss the questions fact-to-face and get quality questions. In this workshop, subject area and the objective may be allotted to the participants on the basis of their competencies.
- **Screening of Questions:** After writing questions, question sheets are passed on to another group of experts for their comments. These comments are passed on the author of the questions for modification if any. Again another type of screening is done with the help of subject experts. These expert group consist of subject specialist, class teacher and evaluation expert. Subject specialist may pass judgement on the authenticity of the subject matter of the questions. Class teacher helps to judge the suitability of the question for a particular grade level and evaluation experts may help to improve the format of the questions in the light of the objectives to be tested.
- **Preparing the Question Cards:** Final form of questions may be transferred into cards made of thick chart paper (size 8"×5").

Chart paper may be of different colours if more than one card is needed and if one card is needed, then white papers may be used. These colouring facilitates identification and sorting of questions arranged in alphabets. Each card will carry the question data.

- **Coding of Questions:** After the preparation of question cards, they are to be coded and classified in such a manner that they facilitate their use. This is necessary for quick location and sorting of cards. Coding can be done for the subject, instructional objectives and forms of questions. In this regard, Dewey's decimal system of cataloguing is the most convenient system.
- **Filling and Storage of Questions:** Question cards may be sorted and arranged topic wise after looking at code number written on question cards. All question cards and other material may be sorted in duplicate, so that one copy may be issued to users as and when required. Original copy may be kept intact for official use and reference.

15.3.4 Uses of question Bank

A well prepared and developed question bank is used for different purposes. The uses of a question bank are stated below:

- Question bank may be used for pre-testing, development, review and revision of lesson.
- Students can use question bank for their self-evaluation in their spare time. It is also helpful for them to revise the lesson.
- Question bank can be utilized for preparing review exercises of text books.
- It is useful to evaluate pupil's progress by sorting and grouping the individual questions in unit and periodical tests.
- Question card can also be used for interaction or class quiz.
- Private tutor of students can use the question bank for clearing the doubts of students regarding the lesson.



CHECK YOUR PROGRESS

Q 1: Question bank is a _____ collection of questions. (Fill in the gap).

Q 2: Define question bank.

.....

Q 3: Mention three uses of question bank.

.....

15.4 GRADING SYSTEM

Grading is very important to assess the comprehensive performance of the students. At present, to improve the evaluation system and especially the examination system, grading is used instead of marks in every stages of education. National Curriculum Framework or NCF (2005) also suggested that grades be used in place of marks in determining the level of pupil performance and proficiency.

15.4.1 Concept of Grading

The word 'grade' is derived from the Latin word 'gradus' which means 'step'. Grading generally means any step, class or stage. Grading is a process where in subjects or pupils may be classified on the basis of predefined standards and aimed at minimizing misclassification. In grading, students are categorized into a few ability groups on the basis of their performance and proficiency. It involves the use of a set of specialized symbols or numerals whose meaning ought to be clearly defined and uniformly understood by the students, teachers, parents and all other stakeholders. The grading symbols must have the same meaning for all who use them to serve the purpose of communication meaningful and precise. The grading process depends on many things such as nature of

the subject matter, the difficulty of the question paper, the different abilities such as intellectual, physical, emotional, personal, social etc assessed by tools of evaluation and precision is required in the evaluation programme.

Grading is the process by which a teacher assesses student learning through classroom tests and assignments, the context in which good teachers establish that process and the dialogue that surrounds grades and defines their meaning to various audiences. Grading serves four roles:

- It evaluates the quality of a student's work.
- It communicates with the student, as well as employers, graduate schools, and others.
- It motivates how the students study, what they focus on, and their involvement in the course.
- It organizes to mark transitions, bring closure, and focus effort for both students and teachers.

It is said that instead of marks, grading is a scientific system. Grades are considered as statements of value. They are much like merit lists. Grades are related to percentage of people. Marks are statements of quantity whereas grades are statements of quality. The marks secured by students in different subjects are converted in to grades and addition of all the grades of different subject creates the value of Grade Point Average (GPA). Generally, 5 point or 7-point grade series are used to evaluate the performance of students. For example:

5 Point Grade Series

Grade	Class	Grade Score
A	Very Good	5
B	Good	4
C	Average	3
D	Poor	2
E	Very Poor	1

7 Point Grade Series

Grade	Class	Grade Score
O	Outstanding	6
A	Very Good	5
B	Good	4
C	Average	3
D	Fair	2
E	Poor	1
F	Fail	0

15.4.2 Methods of Grading System

Grading can be provided by applying different methods. Two main methods of calculating grades are discussed below:

- **Direct Grading:** In direct grading the performance exhibited by the examinees is assessed in qualitative terms. It may be in the form of mode of behaviours expected from the response of the question or activity to be assessed or by using certain indicators expressed in the form of specific behaviours on the basis of the impression so obtained by the examiner is directly expressed in terms of letter grades. This method may be profitably used for the assessment of both cognitive and non-cognitive learning outcomes.
- **Indirect Grading:** In this technique, marks are assigned to the answers as usual. Then these marks are converted into grades. This is done mainly in two ways: Relative Grading and Absolute Grading.

a) Relative Grading: In this type of grading, students secured a grade on the basis of the marks obtained by other students in the same test or examination. This method is also known as norm referenced method. In this method, grade is determined by the student's relative ranking in the total group rather than by some absolute standard of achievement. Because the grading

is based on relative performance, the grade is influenced by both the student's performance and the performance of the group. The assignment of relative grades is essentially a matter of ranking the students in order of overall achievement and assigning letter grades on the basis of each student's rank in the group. This ranking might be limited to a single classroom group or might be based on the combined distributions of several classroom groups taking the same course. The most common version of this approach is grading on the normal curve.

There are two different techniques used in grading on the normal curve. They are prefixed proportion approach and prefixed interval approach. In **prefixed proportion approach,** before assigning letter grades proportion of each grade to be assigned must be determined. For example, if five point grading scale is used on the normal curve, then there must be equal percentage of A's and F's and also B's and D's. Thus, regardless of the group's level of ability, the proportion of high grades is balanced by an equal proportion of low grades. An example of this type of assigning grade is as follows:

Scale of 5 grade

A	B	C	D	E
7%	24%	38%	24%	7%

Scale of 7 grade

O	A	B	C	D	E	F
3%	7%	22%	36%	22%	7%	3%

Scale of 9 grade

O	A+	A	B+	B	C+	C	D	E
4%	7%	12%	17%	20%	17%	12%	7%	4%



High Group than Mean Group | Mean Group | Lower Group than Mean Group

From the above diagram, it can be said that from the scale of 5 grade in any examination, 7% of the students securing highest mark are included in A grade category, next 24% students are included in B grade category, next 38% students are included in C grade category, next 24% students is included in D grade category and last 7% students are included in E grade category. In the same way, students are included in 7 grade scale and 9 grade scale.

Another technique used in grading on the curve is Prefixed Interval Approach. Here, first find out the mean and standard deviation of the marks of the students in a group. Then divide the group into different categories by taking equal distance from mean and standard deviation and assigning grades from excellent to poorest performance. For example, arithmetic mean and standard deviation of marks obtained by particular class of students is 50 and 10 respectively. Then according to prefixed interval approach we can categorize pupils in a five point grading scale.

b) **Absolute Grading:** Another grading method used in evaluation is absolute grading. This grading is also known as criterion referenced grading. Here, grades are assigned on an absolute basis by comparing a student's performance with specified standards set by the authority, preferably with reference to agreed-on performance standards and for the school or board. These standards may be concerned with the degree of mastery to be achieved by students and may be specified as (a) tasks to be performed or (b) the percentage of correct answers to be obtained for a test designed to measure a clearly defined set of learning tasks. With a standard based system, letter grades are assigned on the basis of an absolute standard of performance rather than a relative one. If all students demonstrate a high level of mastery consistent with the established performance standards, all will receive high grades. Here the grade is assigned according to a student's level of achievement or performance, regardless of other student's

achievement or performance. For attaining mastery in learning, percentage of marks can be set as absolute standards. The assignment of grades can be based on the range of marks as shown below:

Mark Range	Grade	Grade Point
91-100	A+	10.0
81-90	A	9.0
71-80	B+	8.0
61-70	B	7.0
51-60	C+	6.0
41-50	C	5.0
31-40	D	4.0
0-30	E	-

For using this system of grading the teachers should have some prior knowledge of what level of achievement or performance is reasonable to expect. Teachers should also have knowledge about the difficulty level of test and each of the items in the test.

15.4.3 Advantages and Disadvantages of Grading System

There are certain advantages and disadvantages of using grading system. These are mentioned below:

Advantages:

- Grading minimize errors in the classification of students on the basis of marks.
- It eliminates unhealthy competition among high achievers.
- It reduces societal pressure on the students.
- It provides flexible learning situations to make learners more comfortable.
- It makes the learning environment more effective.
- It makes learning stress free and enjoyable.

Disadvantages:

- It does not instil a sense of competition among students.
- It is not an accurate representation of the performance and the knowledge gained.
- There is a chance of misclassification which increases the variability in the neighbourhood of cut scores.
- General people have no idea about the interpretation of grading system.



CHECK YOUR PROGRESS

Q 4: Write the derivative meaning of grading.

.....

Q 5: Mention the role of grading.

.....

Q 6: Write one difference between relative grading and absolute grading.

.....

15.5 SEMESTER SYSTEM

Semester system is one of the demands of the globalised world and recently colleges and universities have switched over to this system. The switch over from annual to semester system is a strategic decision since it requires a lot of change in the structure, budget, curriculum and behaviour of stakeholders i.e. students, faculty and administrators.

15.5.1 Concept of Semester System

Etymologically, the word 'Semester' is derived from the German language which means half year. It means one academic year is divided into two. The Shorter Oxford English Dictionary provided the following meaning of the term Semester "A period of term six

months, especially in German University and some United State colleges, the college half year". It is thus clear that the semester signifies the division of academic year into two parts. Division of the academic year also implies courses are designed to cover one semester (instead of a year) and that final examinations are held twice a year.

According to Dictionary of Education edited by C. V. Good, semester means half the academic year usually 16 to 8 weeks. The main aim of semester system is to put emphasis on continuous and comprehensive and depth learning to build capacity of the students by developing required knowledge, skill and attitude to become an efficient and effective citizen in every field.

The semester system is not only the examination system rather it is an education system which places more emphasis on learning rather than teaching. In simple words, it is a learner centred system of education which puts more emphasis on building learning potential of students rather than imparting knowledge through rigid system of instruction. The course is designed with specific learning objectives and outcomes, keeping in view the learning needs of the target group students. The course contents, learning strategy, learning methodology and learning process with specific techniques of assessment and evaluation of students are decided on course by course basis. To meet the learning objectives of each course for specific group of students, autonomy of teacher and internal system of assessment through variety of methods including tests/ examinations become vital for the implementation of semester system. Here, role of the teacher should be facilitator rather than a teacher in its strictest and traditional sense.

15.5.2 Components of the Semester System

The main components of semester system are described below:

- Division of academic year into two terms of 16-18 weeks duration i.e. having 90-100 working days (500-600 hours).

- Existing courses are restructured into unit models of different credit according to semester wise.
- According to need of the pupils and future requirements, there is flexibility to choose the semester course.
- There is a provision of carrying forward of the credits earn in different semester.
- Instructional technology should be reformed.
- 15 hours of instruction in theory course per week in a semester may carry one credit and in case of practical course, 21 hours of practical may carry one credit.
- Almost total 108 credit courses will be done in a semester course which includes continuous and comprehensive external evaluation, periodical test and remedial instruction for the needy one so that all students attend optimal levels of achievement.
- One-year course is divided into two semesters and two year course into four semesters. Naturally, in the semester system the students study half of the course and appear in the examination. Then, they will study next of the course and thereafter again they appear in the examination.

15.5.3 Advantages and Disadvantages of Semester System

There are some views regarding advantages and disadvantages of semester system. These are:

Advantages:

- The quantities of content in various subjects would be proportioned suitably in semester system.
- It helps students to achieve their objectives in a better way.
- Continuous assessment being an important component of semester system encourages the students to work systematically throughout the course.

- There is increased self-reliance on the part of students in respect of schedule evaluation and performance improvement.
- Rapport and relationship between teachers and students are increased in semester system.
- Semester system include improved methods of assessing students' performance in examination thus increasing the use of objective type testing.
- Students have the opportunity to discuss their awareness in all accepted tools of internal evaluation. It adds to reliability.

Disadvantages:

- Both the teachers and students merely have clear and proper understanding of it.
- Internal assessment is an integral part of semester system so there are chances of favouritism and subjectivity in it.
- A number of students develop the habit of guessing the answers as maximum number of questions are objective type.
- It creates an unhealthy competition among different institutions because in each institution favouritism in internal assessment is there.



CHECK YOUR PROGRESS

Q 7: What is semester system?

.....

Q 8: Write three components of semester system.

.....

15.6 USE OF COMPUTER IN EVALUATION

Computer has a vital role to play in effective and efficient evaluation of students' learning. Computer offers educators a variety of new tools that can be used in the classroom. Computer can help teachers assess their

students' learning as well as their performance in the classroom. Use of computer in evaluation involves the use of digital devices to assist in the construction, delivery, storage or reporting of student assessment tasks, responses, grades or feedback.

Geoffrey Crisp (2011) in the Teacher's Handbook on e-Assessment stated that the ICT based evaluation can be undertaken with many devices, such as traditional desktop computers or laptops, with portable communication devices such as smart mobile phones, with digital devices such as iPads or through the use of electronic gaming devices. Computer based evaluation can use a multitude of formats, including text documents or portable document formats, multimedia formats such as sound, video or images; it can involve complex simulations or games; it can also be undertaken by students in groups or individually and it can occur with large numbers of students in a synchronous or asynchronous manner.

Teachers can use computers for various purposes of evaluation. Some of them are mentioned below:

- Computer can be used to construct the evaluation tasks of teachers, to deliver these tasks to relevant students and to record and provide feedback and grades to these students.
- Computers can also be used to analyse students' responses, both to provide feedback to the student on the quality and relevance of their response, as well as to provide feedback to the teacher on whether the task can differentiate between students with different abilities.
- ICT based assessment can be used to test many different capabilities and skills that are developed by students.
- There are only a few tasks that might not be suitable for completing and recording electronically, but the number of such tasks is rapidly diminishing as technology becomes more sophisticated and widespread.
- In many disciplines laboratory equipment can be manipulated remotely and students can undertake real time physical performances that are able to be recorded and used for assessment

purposes. We are quickly approaching the stage where our imaginations will be the limiting factor in designing e-assessment tasks.

Christine, R. (2013) identified two conceptually different approaches to assessing Key Competencies using computer. These are:

- **Computer-Based Assessment (CBA):** This approach has been employed for more than two decades and now goes beyond simple multiple choice test formats. With this new "Generation Re-Invention" or "transformative" testing, questions are increasingly integrated into more complex and authentic problem contexts, so that the full range of Key Competencies can be assessed. Additionally, due to technological advances, a wider range of answer formats, including free text and speech, can be automatically scored.
- **Technology-enhanced learning environment:** It offers a promising avenue for embedded assessment of the more complex and behavioural dimensions of Key Competencies, based on Learning Analytics. Many of the currently available technology-enhanced learning environments, tools and systems recreate learning situations which require complex thinking, problem-solving and collaboration strategies and thus allow for the development of generic skills. Some of these environments allow learners and teachers to assess performance, understand mistakes and learn from them.



CHECK YOUR PROGRESS

Q 9: What is the full form of CBA?

.....

Q 10: Mention one purpose of using computer in evaluation.

.....



15.7 LET US SUM UP

After going through this units, you can summarize it as-

- Question bank is also termed as model question paper and item bank.
- A question bank is a planned library of test items designed to fulfil certain predetermined purposes. Question bank mainly manage to provide utility services to all those who are interested.
- Question bank is mainly based on sound ideas, purposes and adapted to the particular objective of mind.
- Question bank should present a challenge which stimulates an educative response from the users in keeping with the objectives.
- One important purpose of question bank is to improve the teaching learning process. Question bank is an important tool through which teacher as well as students will be benefited.
- There are certain steps in development of question bank. These steps are discussed below: blue print of question bank, writing of questions, screening of questions, preparing the question cards, coding of questions and filling and storage of questions.
- Grading is a process where in subjects or pupils may be classified on the basis of predefined standards and aimed at minimizing misclassification.
- The grading process depends on many things such as nature of the subject matter, the difficulty of the question paper, the different abilities such as intellectual, physical, emotional, personal, social etc. assessed by tools of evaluation and precision is required in the evaluation programme.
- In direct grading the performance exhibited by the examinees is assessed in qualitative terms.
- In this type of grading, students secured a grade on the basis of the marks obtained by other students in the same test or examination. This method is also known as norm referenced method.

- Absolute grading is also known as criterion referenced grading. Here, grades are assigned on an absolute basis which involves comparing a student's performance with specified standards set by the authority, preferably in reference to agreed-on performance standards and for the school or board.
- Etymologically, the word 'Semester' is derived from the German language which means half year. It means one academic year is divided into two.
- The main aim of semester system is to put emphasis on continuous and comprehensive and depth learning to build capacity of the students by developing required knowledge, skill and attitude to become an efficient and effective citizen in every field.
- Almost total 108 credit courses will be done in a semester course which includes continuous and comprehensive external evaluation, periodical test and remedial instruction for the needy one so that all students attend optimal levels of achievement.
- Semester system include improve methods of assessing students' performance in examination making increased use of objective type testing.
- Internal assessment is an integral part of semester system so there are chances of favouritism and subjectivity in it.
- Computer can help teachers assess their students' learning as well as their performance in the classroom.
- Use of computer in evaluation involves the use of digital devices to assist in the construction, delivery, storage or reporting of student assessment tasks, responses, grades or feedback.



15.8 FURTHER READING

1) Aggarwal, J. C. (1997): Essentials of Examination System. New Delhi, Vikash Publishing House.
2) Bhatia, K. K. (1995): Measurement and Evaluation in Education. Ludhiana, Prakash Parishad.

3) Das Sarma & Kalita (2015): Sikshat Parimapan, Mulyayan Aaru Parisankhyabignan, Guwahati: Shanti Prakashan.
4) Goswami, Marami (2015): Educational Measurement and Evaluation. Guwahati, Ashok Book Stall.
5) Sharma, R. A. (2005): Essentials of Measurement in Education and Psychology, Meerut: R. Lall Book Depot.
6) Talukdar, Dipankar (ed.) (2014): Examination Reforms in Higher Education, Guwahati: R. G. Baruah College.

15.9 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Organised.

Ans to Q No 2: According to B. H. Choppin, "An item bank or question bank is a collection of test items organised, classified and catalogues the order to facilitate the construction of a variety of achievement and other types of mental test."

Ans to Q No 3: Three uses of question bank are:

- Question bank may be used for pre-testing, development, review and revision of lesson.
- Students can use question bank for their self-evaluation in their spare time. It is also helpful for them to revise the lesson.
- Question bank can be utilized for preparing review exercise of text books.

Ans to Q No 4: The word 'grade' is derived from the Latin word 'gradus' which means 'step'.

Ans to Q No 5: Grading serves four roles:

- It evaluates the quality of a student's work.
- It communicates with the student, as well as employers, graduate schools, and others.
- It motivates how the students study, what they focus on, and their involvement in the course.
- It organizes to mark transitions, bring closure, and focus effort for both students and teachers.

Ans to Q No 6: In relative grading, other students' achievement or performance is necessary whereas in absolute grading, grade is assigned according to a student's level of achievement or performance, regardless of other student's performance or achievement.

Ans to Q No 7: The Shorter Oxford English Dictionary provided the following meaning of the term Semester "A period of term six months, especially in German University and some United State colleges, the college half year". It is thus clear that the semester signifies the division of academic year into two parts.

Ans to Q No 8: Three components of semester system are mentioned below:

- Division of academic year into two terms of 16-18 weeks duration i.e. having 90-100 working days (500-600 hours).
- Existing courses are restructured into unit models of different credit according to semester wise.
- According to need of the pupils and future requirements, there is flexibility to choose the semester course.

Ans to Q No 9: Computer Based Assessment.

Ans to Q No 10: Computer used to construct the evaluation tasks of teachers, to deliver these tasks to relevant students and to record and provide feedback and grades to these students.



15.10 MODEL QUESTIONS

A. Short Questions

Q 1: What is grading?

Q 2: What is question bank?

Q 3: Write the full form of GPA.

Q 4: What is semester system?

B. Short Questions (Answer each question in around 150 words)

Q 1: Write the characteristics of question bank.

Q 2: Write a note on relative grading.

Q 3: Write about the advantages and disadvantages of grading system.

Q 4: Write about the advantages and disadvantages of semester system.

C. Long Questions. (Answer each question in around 300-500 words)

Q 1: Describe the developmental procedure of a good question bank.

Q 2: 'Grading is very effective in evaluation system'- discuss the statement.

Q 3: Discuss the issues and challenges of implementing semester system in our university from your own point of view.

Q 4: What type of role is played by computer in evaluation? Discuss.

*** ***** ***



REFERENCES

1. Aggarwal, J.C. (2015): Essentials of Educational Psychology, New Delhi: Vikash Publishing House Pvt. Ltd.
2. Anastasi Anne (1982): Psychological Testing, 4th Edition, New York: Macmillan Publishing Co. Inc.
3. Anderson, Scarvia B. Samuel Ball, Murphy, R.T. and et.al (1976): Encyclopedia of Educational Evaluation, Jessey-Bass Publishers.
4. Asaad, A.S. & Hailaya, W.M. (2004): Measurement and Evaluation; concepts and principles, Rex Book Store. Inc. Manila
5. Asthana B (2015): Measurement and Evaluation in in Psychology and Education, Binod Pustak Mandir, Agra.
6. Baer, J., Kaufman, J. C., Plucker, J. A. (2008): Essentials of Creativity Assessment, Germany: Wiley.
7. Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001): The Autism-spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians, Journal of Autism and Developmental Disorders, 31, 5-17.
8. Bean, K.L.: Construction of Educational and Personal Tests, McGraw Hill Book, Inc. New York
9. Berg, Kris E. and Latin, Richard Wayne (2008): Essentials of Research Methods in Health, Physical Education, Exercise Science, and Recreation, Lippincott Williams & Wilkins
10. Bhatia, K. K. (1995): Measurement and Evaluation in Education. Ludhiana, Prakash Parishad.
11. Bloom, B.S. (1956): Taxonomy of Educational Objectives. Handbook I, Cognitive Domain, David McKay Co. Inc, New York.
12. Boehm, Ann E. and White, Mary Alice (1982): The Parent's Handbook on School Testing. New York: Teacher's College Press.
13. Brown Frederick, G. (1976): Principles of Educational and Psychological Testing, Holt: Rineheart and Winston.

14. Bruno-Faria, M., Fleith, D., Alencar, E. (2014): Theory and Practice of Creativity Measurement. United States: Sourcebooks.
15. Chauhan, S. S. (1978): Advanced Educational Psychology, Vikas Publishing House Private Limited, New Delhi
16. Concepts and Principles, 1st Edition, Manila, Rex Book Store, Philippines.
17. Creative Thinking of School Students. (2009): India: Discovery Publishing House.
18. Cronbach, L. J. (1990): Essentials of psychological testing,5th Edition, Harper Collins Publishers: New York
19. Cronbach, Lee J. (1960): Essentials of psychological testing, Harper & Brothers, New York
20. Das Sarma & Kalita (2015): Sikshat Parimapan, Mulyayan Aaru Parisankhyabignan, Guwahati: Shanti Prakashan.
21. Ebel, R.L. and Frisbel, D.A. (1979): Essential of Educational Measurement, 3rd Edition, Prentice Hall of India, New Delhi.
22. Freeman, F.S. (1962): Theory and Practice of Psychological Testing, Holt, Rirehant & Winston, New York
23. Gardner, H. (1993): Multiple intelligences: The theory in practice, Basic Books, New York
24. Goswami M (2010): Measurement and Evaluation in Education and Psychology, Nilkamal Publishing House, Hyderabad, 2010
25. Gronlund, Norman, E (1968): Readings in Measurement and Evaluation, Education, and Psychology, The Macmillan Company.
26. Gupta, S.P. (2009): Modern Measurement and Evaluations with Statistics, Sharda Pustak Bhavan, Allahabad
27. Handbook of Creativity. (1999). United Kingdom: Cambridge University Press.
28. Jensen, Arthur R.(1981): Straight Talk About Mental Tests Their Uses and Abuses Explained in No-nonsense Terms by the World's Foremost Authority, New York: The Free Press.
29. Kelley, T. L. (1927): Interpretation of educational measurements, Macmillan, New York

30. L. J. (1990): Essentials of psychological testing,5th Edition, Harper Collins Publishers: New York
31. Linn, R.L. and Gronlund, N.E. (1995): Measurement and Assessment in Teaching, 7th Edition. Prentice-Hall, New Jersey
32. Mangal, S. K. (2014): Advanced Educational Psychology, 2nd Edition, PHI Learning Private Limited, Delhi
33. Mehrens, W.A. and Lehmann, I.J. (1962): Standardized tests in Education, Holt, Rinehart and Winston, New York
34. Michael, A. Leonard (2016): Characteristics of a Good Test, Department of Library and Information Science, Delta State University, Abraka
35. NCERT (1963): The Concept of Evaluations in Education, New Delhi.
36. Newton, Paul & Shaw, Stuart (2014): Validity in Educational and Psychological Assessment, Sage Publications, London.
37. Nunnally, J. C. (1967): Psychometric Theory, McGraw-Hill, New York
38. Pradhan Hemant (1996): Essentials of Educational Measurement, Evaluation and Statistics, Saha Publishers and Distributers, Sambalpur
39. Safrit, Margaret J. And Wood, Terry M. (1995): Introduction to Measurement in Physical Education and Exercise Science, Mosby, Missouri
40. Sandfeord, Peter (1939): The Educational Psychology on Objective Study, Longmans Edition,
41. Sarsani, M. R. (2006): Creativity in Schools, India: Sarup Book Publishers
42. Sattler, Jerome M. (1988): Assessment of Children, Third edition. San Diego: Jerome M. Sattler, Publisher.
43. Sax, G. (1970): Principle of Educational Measurement and Evaluation, Wordsworth Pub. California.
44. Sharma, R. A. (2005): Essentials of Measurement in Education and Psychology, Meerut: R. Lall Book Depot.

45. Sing, A.K. (2008): Tests, Measurement and Research Methods in Behavioural Sciences, New Delhi: Bharati Bhawan.
46. Spencer, E., Claxton, G., Lucas, B. (2012): Progression in Creativity - Developing New Forms of Assessment: A Literature Review, United Kingdom: Unknown Publisher.
47. Stanely, Julian C. & Ross, Clay Campbell (1964): Measurement in today's schools, 4th Edition, Prentice Hall
48. Stanley & Hopkins (1966): Education and Psychological Measurement and Evaluation, Prentice Hall Pvt. Ltd. Eaglewood Cliff, New Jersey
49. Talukdar, Dipankar (ed.) (2014): Examination Reforms in Higher Education, Guwahati: R. G. Baruah College.
50. UGC (1972): Examination Reforms: A Plan of Actions, New Delhi