MAEDN-503
Educational Evaluation

MA EDUCATION

EDUCATIONAL EVALUATION

MA [EDUCATION]

Third Semester

MAEDN-503

RAJIV GANDHI UNIVERSITY

Arunachal Pradesh, INDIA - 791 112

BOARD OF STUDIES

Authors

Prof. J C Soni
Dr. C Siva Sankar
Moyir Riba

About the University

Rajiv Gandhi University (formerly Arunachi University) is a premier institution for higher education in the state of Arunachal Pradesh and has completed twenty-five years of its existence. Late Srnt. Indira Gandhi, the then Prime Minister of India, laid the foundation stone of the university on 4th February, 1984 at Rono Hills, where the present campus is located.

Ever since its inception, the university has been trying to achieve excellence and fulfill the objectives as envisaged in the University Act. The university received academic recognition under Section 2(f) from the University Grants Commission on 28th March, 1985 and started functioning from 1st April, 1985. It got financial recognition under section 12-B of the UGC on 25th March, 1994. Since then Rajiv Gandhi University, (then Arunachal University) has carved a niche for itself in the educational scenario of the country following its selection as a University with potential for excellence by a high-level expert committee of the University Grants Commission from among universities in India.

The University was converted into a Central University with effect from 9th April, 2007 as per notification of the Ministry of Human Resource Development, Government of India.

The University is located atop Rono Hills on a picturesque tableland of 302 acres overlooking the river Dikrong. It is 6.5 km from the National Highway 52-A and 25 km from Itanagar, the State capital. The campus is linked with the National Highway by the Dikrong bridge.

The teaching and research programmes of the University are designed with a view to play a positive role in the socio-economic and cultural development of the State. The University offers Undergraduate, Post-graduate, M.Phil and Ph.D. programmes. The Department of Education also offers the B.Ed. programme.

There are fifteen colleges affiliated to the University. The University has been extending educational facilities to students from the neighbouring states, particularly Assam. The strength of students in different departments of the University and in affiliated colleges has been steadily increasing.

The faculty members have been actively engaged in research activities with financial support from UGC and other funding agencies. Since inception, a number of proposals on research projects have been sanctioned by various funding agencies to the University. Various departments have organized numerous seminars, workshops and conferences. Many faculty members have participated in national and international conferences and seminars held within the country and abroad. Eminent scholars and distinguished personalities have visited the University and delivered lectures on various disciplines.

The academic year 2000-2001 was a year of consolidation for the University. The switch over from the annual to the semester system took off smoothly and the performance of the students registered a marked improvement. Various syllabi designed by Boards of Post-graduate Studies (BPGS) have been implemented. VSAT facility

installed by the ERNET India, New Delhi under the UGC-Infonet program, provides Internet access.

In spite of infrastructural constraints, the University has been maintaining its academic excellence. The University has strictly adhered to the academic calendar, conducted the examinations and declared the results on time. The students from the University have found placements not only in State and Central Government Services, but also in various institutions, industries and organizations. Many students have emerged successful in the National Eligibility Test (NET).

Since inception, the University has made significant progress in teaching, research, innovations in curriculum development and developing infrastructure.

About IDE

The formal system of higher education in our country is facing the problems of access, limitation of seats, lack of facilities and infrastructure. Academicians from various disciplines opine that it is learning which is more important and not the channel of education. The education through distance mode is an alternative mode of imparting instruction to overcome the problems of access, infrastructure and socio-economic barriers. This will meet the demand for qualitative higher education of millions of people who cannot get admission in the regular system and wish to pursue their education. It also helps interested employed and unemployed men and women to continue with their higher education. Distance education is a distinct approach to impart education to learners who remained away in the space and/or time from the teachers and teaching institutions on account of economic, social and other considerations. Our main aim is to provide higher education opportunities to those who are unable to join regular academic and vocational education programmes in the affiliated colleges of the University and make higher education reach to the doorsteps in rural and geographically remote areas of Arunachal Pradesh in particular and Northeastern part of India in general. In 2008, the Centre for Distance Education has been renamed as "Institute of Distance Education (IDE)."

Continuing the endeavor to expand the learning opportunities for distant learners, IDE has introduced Post Graduate Courses in 5 subjects (Education, English, Hindi, History and Political Science) from the Academic Session 2013-14.

The Institute of Distance Education is housed in the Physical Sciences Faculty Building (first floor) next to the University Library. The University campus is 6 kms from NERIST point on National Highway 52A. The University buses ply to NERIST point regularly.

Outstanding Features of Institute of Distance Education:

(i) At Par with Regular Mode

Eligibility requirements, curricular content, mode of examination and the award of degrees are on par with the colleges affiliated to the Rajiv Gandhi University and the Department(s) of the University.

(ii) Self-Instructional Study Material (SISM)

The students are provided SISM prepared by the Institute and approved by Distance Education Council (DEC), New Delhi. This will be provided at the time of admission at the IDE or its Study Centres. SISM is provided only in English except Hindi subject.

(iii) Contact and Counselling Programme (CCP)

The course curriculum of every programme involves counselling in the form of personal contact programme of duration of approximately 7-15 days. The CCP shall not be compulsory for BA. However for professional courses and MA the attendance in CCP will be mandatory.

(iv) Field Training and Project

For professional course(s) there shall be provision of field training and project writing in the concerned subject.

(v) Medium of Instruction and Examination

The medium of instruction and examination will be English for all the subjects except for those subjects where the learners will need to write in the respective languages.

(vi) Subject/Counselling Coordinators

For developing study material, the IDE appoints subject coordinators from within and outside the University. In order to run the PCCP effectively Counselling Coordinators are engaged from the Departments of the University, The Counselling-Coordinators do necessary coordination for involving resource persons in contact and counselling programme and assignment evaluation. The learners can also contact them for clarifying their difficulties in then respective subjects.

SYLLABUS

Objectives:
1. To familiarize the students with the concepts of measurement and evaluation
2. To enable the students to know about the various educational objectives
3. To familiarize the student teachers with characteristics of test.
4. To enable students to know construction of achievement test.

Course Content:
UNIT- I. **Measurement and evaluation:**
- Meaning, nature and scope of measurement and evaluation
- Types of evaluation procedures with special reference to CRE.
- New trends in evaluation: Choice Based Credit
  System, continuous and comprehensive Evaluation, question bank

UNIT-II. **Taxonomy of Educational objectives:**
- Cognitive domain and Formulation of Specific Objectives
- Affective domain and Formulation of Specific Objectives
- Psychomotor domainand Formulation of Specific Objectives

UNIT-III. **Characteristics of Good Test:**
- Reliability and validity
- Meaning of reliability
- Methods of computing reliability, standard error of measurement
- Factors affecting reliability.
- Meaning and types of validity and factors affecting validity
- Relationship between reliability and validity

UNIT-IV. **Construction and Standardization of Achievement Test and Attitude Scale**
- Norms: Meaning and significance of norms, Types of norms and their usability:
  Age norm, Grade norm, percentiles, standard scores-Z-scores, T- Scores, Stanine
- Meaning and Types of Achievement Test- Teacher Made and standardized, Steps of construction of Achievement Test and its standardization
- Meaning and Significance of Measuring Attitude
- Method of construction and standardization of Attitude Scale : Thurstone and Likert method

Practicum:
1. Construction and standardization of an achievement test.
2. Construction of Attitude Scale
3. Practice on taxonomy of educational objectives.

INTRODUCTION

Human beings, irrespective of their race, religion and caste, go through the same cognitive processes in their development. They acquire language, learn to perceive, go through the processes of emotional and moral development, and so on. All of these processes shape an individual's understanding of their own self and help mould their social development. However, sometimes, individuals, especially children, undergo experiences that may prove to be traumatic and affect their ability to cope emotionally with their situations. These experiences may be at school, due to peer pressure or the stress of performing, or they may be at home, in the form of divorce or a death.

In situations like these, the child may require help in the form of guidance and counselling from a professionally qualified individual. Such guidance workers are aware of the words and approach to use for drawing out the problem from the suffering children and provide guidance accordingly.

Intelligence is one of the most intriguing aspects of medicine. Not only have researchers failed to describe it in a nutshell, it remains till date one of the most fascinating areas of inquiry. Educational Evaluation and Guidance in Education introduces the concept of testing and evaluation to its readers and the importance of guidance and counselling. This book is crucial in explaining how, even though different evaluation scales have been designed, the success is not entirely determined by the efficiency of the test. More often than not, the right test administered to the right individual makes all the difference. The way in which tests are constructed and comprehended is crucial to the process of evaluation. Different tests are designed to examine different attributes of intelligence and its evaluation is contingent on how well these tests have been understood. In other words, the lucidity of a test, inclusive of the kind of questions, is an important factor in determining its success and accuracy.

The book *Educational Evaluation and Guidance in Education* has been designed keeping in mind the self-instruction mode (SIM) format and follows a simple pattern, wherein each unit of the book begins with **Introduction** followed by **Unit Objectives** to the topic. The content is then presented in a simple and easy-to- understand manner, and is interspersed with **Check Your Progress** questions to test the understanding of the topic by the students. A list of **Questions and Exercises** is also provided at the end of each unit, and includes short-answer as well as long- answer questions. **Key Terms** and the **Summary** section is a useful tool for students and arc meant for effective recapitulation of the text.

The book is divided into eight units:

Unit 1: Familiarizes the students with the concepts of measurement and evaluation

Unit 2: Deals with taxonomy of educational objectives

Unit 3: Introduces the various types of tests and their characteristics

Unit 4: Analyses the construction and standardization of achievement tests

UNIT 1   MEASUREMENT AND

EVALUATION

Structure

1.0   INTRODUCTION

The process of measurement is an essential component of the learning system. The measurement has more administrative utility in education than evaluation. An evaluation process is useful in modifying and improving the learning system and instructional procedure. Measurement is more precise and objective in approach as compared to evaluation.

Measurement is always done of a quality, attribute or variable of a thing or a person. Psychologists and educationists are mainly concerned with the variables and attributes. The process of measurement converts the variables into variety which is used for drawing the inferences. For example, intelligence is quantified in terms of EQ (Intelligent Quotient) and achievement variable is measured in terms of scores.

In this unit, you will be introduced to the concept of measurement and evaluation in education. You will also learn about the types of evaluation, with special reference to CRE (criterion referenced evaluation) and the new trends in evaluation.

1.1 UNIT OBJECTIVES

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

- Discuss the meaning, nature and scope of measurement and evaluation in education
- Identify the various types of evaluation
- Describe the new trends in evaluation

1.2 FUNDAMENTAL PRINCIPLES OF MEASUREMENT

The fundamental principles of evaluation in education is discussed in detail in this section.

1.2.1 Meaning

Measurement refers to the process by which the attributes or dimensions of some physical object are determined. When used in the' context of learning, it would refer to: applying a standard scale or measuring device to an object, series of objects, events or conditions, according to practices accepted by those who are skilled in the use of the device or scale.

American psychologist E. L. Thorndike stated that 'anything that exists at all, exists in some quantity; and anything that exists in some quantity is capable of being measured'. Measurement of any kind is a matter of determining how much or how little, how great or how small, how much more than or how much less than. The *Encyclopedia of Educational Research* explains measurement in more refined terms; to measure means 'to observe or determine the magnitude of a variant'. Measurement answers the question of 'how much'.

In our day-to-day life, we measure the height, weight, miles travelled, etc.; the tailor measures the dimensions of an individual's body to prepare dress according to required size; the shopkeepers weigh different commodities like rice, wheat, sugar, fruits and vegetables. We have a wrong notion that measurement takes place only with tapes and scales. The ranking of contestants in a debate competition can be considered as measurement, while rating of human behaviour comes under measurement. Hence, measurement may be defined as, 'the assignment of one of a set of numbers to each of a set of persons or objects according to certain established rules'.

Professor James M. Bradfield defined measurement as 'the process of assigning symbols to the dimension of phenomenon in order to characterize the status of phenomenon as precisely as possible'. J. P. Guilford defined measurement as the 'assignment of numerals to objects or events according to certain rules'. According to Norman E. Gronlund, 'Measurement results are some score or numerical value and quantitative descriptions of the pupils'.

Measurement involves the process of quantification. Quantification indicates to what extent a particular attribute is present in a particular object. It has been observed that measurement in any field always involves the following three essentials:

(i) Identifying and defining of quantity, attribute or variable that is to be measured

(ii) Determining the set of operations by which the attribute or variable may be made perceivable

(iii) Establishing a set of procedure for translating observations into quantitative statement of degree, extent or amount

1.2.2 Types of Measurement

Measurement is of two types: (i) physical measurement and (ii) mental measurement/psychological measurement/educational measurement.

(i) **Physical measurement:** Physical measurement is the measurement of the object which has absolute existence. For example, we measure height of individuals, and the weight of rice. Here, we directly measure the height or weight of an individual and all the measuring tools of physical measurement start from zero. Physical measurement is always accurate and quantitative, and there are some set of tools for physical measurement all over the world.

(ii) **Mental measurement:** Mental measurement is also known as 'educational measurement' or 'psychological measurement'. It is always relative and there is no absolute zero in case of mental measurement. For example, for measuring the intelligence of a person, we have to take the help of intelligence test, which is a subjective one. Through his response, we can know the level of intelligence of the person concerned. Mental measurement is both qualitative and quantitative in nature, and there are no fixed tools for such measurements, i.e., the same set of tools may not be applied to different types of persons.

The application of the principles of measurement in the field of education is known as 'educational measurement'. In the educational system, measurement is the quantitative assessment of performance of the students in a given test. It can be used to compare performance between different students and to indicate the strengths and weaknesses of the students. It helps in classifying students into homogenous group to assign educational and vocational guidance and to provide remedial measures to the low achievers. Measurement is a tool in the hands of the educational psychologists to study human behaviour. The educational psychologists take the help of different valid and reliable psychological tests to know the level of different traits within an individual. The different kinds of such tests are: intelligence test, achievement test, attitude test, aptitude test, interest inventory, personality test, etc. The methods used for these tests are: observation, interview, checklist, rating scale, examinations, cumulative record card and anecdotal records, etc.

In the teaching-learning situation, teachers should be competent enough to measure the student's achievement, intelligence, attitude, aptitude, etc. To develop competency among the teachers in educational measurement, Ebel has suggested the following measures:

(i) Know how to administer a test properly, efficiently and fairly.

(ii) Know how to interpret test scores correctly and frilly, but with recognition of their limitations.

(iii) Know how to select a standardized test that will be effective in a particular situation.

(iv) Know how to plan a test and write the test questions, to be included in it.

(v) Know the educational uses as well as the limitations of educational tests.

(vi) Know the criteria by which the quality of a test should be judged and how to secure evidence relating to these criteria.

1.2.3 Nature of Measurement

To measure the psychological traits with validity and reliability, the measuring instrument or tests should be far away from the aspects like personal errors, variable errors, constant errors and interpretative errors. The important characteristics of good measuring tools are as follows:

(i) **Should be valid:** Validity of a test refers to its truthfulness. It refers to the extent to which a test measures what it actually wishes to measure. Suppose we want to know whether a Numerical Reasoning Test is valid. If it really measures the reasoning ability, the test can be said to be valid.

(ii) **Should be reliable:** Reliability refers to the consistency of a measuring instrument (how accurately it measures). It refers to the faithfulness of the test. To express in a general way, if a measuring instrument measures consistently, it is reliable. For example, a test is administered on English to the students of class VI. In this test, Ram scores 50. After a few days, the same test is administered and Ram scores 50 again. Here, the test is reliable because there is consistency in the result.

(iii) **Should be objective:** Objectivity of a test refers to two aspects: (a) item objectivity (i.e., objectivity of the items), and (b) scoring objectivity (i.e., objectivity of scoring). By 'item objectivity', we mean that the items of the test must need a definite single answer. If the answer is scored by different examiners, the marks would not vary. Ambiguous questions, lack of proper direction, double barrelled questions, questions with double negatives and essay- type questions must be avoided because they lack objectivity. By 'objectivity of scoring', we mean that by whomsoever scored, the test would fetch the same score. Thus, mostly the objective-type questions should be framed to maintain the objectivity of the test.

(iv) **Should be usable and practicable:** 'Usability' refers to the practicability of the test. In the teaching-learning situation, by usability we mean the degree to which the test (or the measuring tool) can be successfully used by teachers and school administrators.

(v) **Should be comprehensive and precise:** The test must be comprehensive and precise. It means that the items must be free from ambiguity. The directions to test the items must be clear and understandable. The directions for

administration and scoring must be clearly stated so that a classroom teacher can easily understand and follow them.

(vi) **Should be easy in administering:** If the directions for administration are complicated, or if they involve more time and labour, the users may lag behind. For example, Wechsler Adult Intelligence Scale (WAIS) is a good test, but its administration is very difficult.

(vii) **Should be economical:** A measurement tool should be less time consuming. The cost of the test must be reasonable so that the schools/educational institutions can afford to purchase and use it.

(viii) **Should be easy in scoring:** The scoring procedure of the test should be clear and simple. The scoring directions and adequate scoring key should be provided to the scorer so that the test is easily scored.

(ix) **Should be easily available:** Some standardized tests are well known all over India, but they are not easily available. Such tests have less usability. It is desirable that in order to be usable, the test must be readily and easily available.

(x) **Should have clear and attractive appearance:** The quality of papers used, typography and printing, letter size, spacing, pictures and diagrams presented, its binding, space for pupil's responses, etc., should be very good and attractive.

1.2.4 Scales of Measurement

In general, measurement is a process of assigning numbers to objects or observations, the levels of measurement being a function of the values under which the numbers are assigned. In mental measurement, our objective is to measure a trait as precisely as possible. However, due to the subjectivity and relativeness of the psychological concepts, it is very difficult to measure it absolutely. For example, it is very difficult to measure the extent to which an individual is obedient. Yet, we can order or rank the individuals as far as the trait in question is concerned. American psychologist Stanley Smith Stevens has attempted a classification of different measurement levels. Also called 'levels of measurement', they are expressed in the following manner:

(i) **Nominal scale:** Nominal scale is used to classify objects. It is the least precise or lowest order scale. Here, the numbers assigned to individuals are not bound by any rule and there is no logical consideration. For example, the pin codes provided to different places and the serial numbers given to cricket players, are only to identify the places or players, respectively. In other words, these values or numbers are simply assigned without any logic where each class is represented by a letter, a name, a number or even a geometrical design. Each number or symbol is like a category name; it has no quantitative significance. This scale is used for students' classification, job classification, and for different types of categorization.

(ii) **Ordinal scale:** In this scale, ranking or grading takes place. It tells us that one observation represents more or less of a variable than the other; but it does not tell how much more or less. Thus, ordinal scale corresponds to quantitative classification of a set of objects with reference to some attribute or rule. For example, when we serially arrange five bags of wheat according to the weight,

the heaviest bag is represented by a numeral 1, the next by 2, and so on. More precise comparison is not done with ordinal scale.

(iii) **Interval scale:** Interval scales have an arbitrary zero, not an absolute zero. Interval scale is also known as 'equal interval scale' because these scales show that a person or item is so many units larger or smaller, heavier or lighter, brighter or duller, etc., than the other. In interval scales, the intervals (i.e., the difference between two consecutive points on the scale) are assumed to be equal over the scale.

All the mental measurement deals with interval scales. In this case, there is no absolute zero point of reference. In mental measurement, the concept of zero is not there; but in case of physical measurement, the concept of absolute zero is well conceived. For example, zero centimetre means absence of length; but in education and psychology, it is difficult to visualize a true zero in any scale used.

For example, a student who scores zero in an achievement test is not supposed to have any knowledge about the subject. In a similar way, an IQ of zero has no meaning. Due to the absence of a true zero point, we cannot say that a child with an IQ of 1 50 is twice as bright as a child with IQ of 75. Fahrenheit and centigrade temperature scales are examples of interval scale. In Fahrenheit scale, 0°F does not imply absence of heat; so this zero is not absolute and, hence, it is an interval scale. Similar is the case with centigrade scale because 0°C does not imply absence of heat.

(iv) **Ratio scale:** The ratio scale is most concrete and refined among the four basic scales. It has an absolute zero point representing complete absence of the property being measured. Here, a boy with a height of 6 feet is considered to be twice as tall as a boy with a height of 3 feet. In this scale, a true zero point means complete absence of an attribute. In physical sciences and in all physical measurements, we use ratio scales. Measurement of physical dimensions such as height, weight and distance are examples of ratio scales.

1.2.5 Scope of Measurement

American educational psychologist Lee J. Cronbach (1949) classified all applications of mental measurement under the following three main functions:

(i) **Prognosis function:** The first function of measurement is prognosis function. Any test tells about some differences among people's performance at this movement. All decisions involve prediction when psychological test is mentioned, so-called IQ test administered to students in school, to predict their academic performance. The measurement provides the extent of a variable which has the specific purpose of predicting future behaviour.

The prognosis has the administrative function such as classification, selection, promotion and gradation of students. The college students ask a counsellor to help them choose the best curriculum or job. Thus, the guidance and counselling services are also based upon the prognosis function of measurement, i.e., the effectiveness of method, instruction and treatment evaluated on the basis of a student's achievement.

(ii) **Diagnosis function:** The second major function of measurement is diagnosis. The prognosis function reveals the level of a student with regard to certain characteristics, whereas the diagnosis function identifies the weakness of the student-learning. The remedial instruction can be prepared on the basis of diagnosis. It also implies the prediction, but there is considerable justification in listing diagnosis of a separate function of measurement. The diagnostic function establishes the cause—effect relationship, but prediction implies the simple relationship. The instructional procedure can be improved by this function of psychological measurement.

(iii) **Research function:** The third major function of measurement is verification of scientific hypotheses of research, The use of measurement for research purpose, however, is not much when compared to prediction and diagnosis. This is because a measurement is usually considered a completely valid measure of certain human characteristics. An investigator must treat test scores in this experiment as an accurate quantification of real and useful variable. Measurement provides a more objective and dependable basis for comparison than rough impressions. Thus, the valid generalizations are made on the basis of an accurate measurement.

1.2.6 Methods of Measurement

For both physical and mental measurement, some tools and methods are necessary. The variation of method may be due to the nature of variable and purpose of measurement. The methods of measurement are as follows:

(i) **Tests:** A test consists of a set of questions to be answered or tasks to be performed. Tests are used to assess the ability or trait in question. Psychological and educational tests are a standardized procedure to measure quantitatively or qualitatively one or more than one aspect or trait by means of a sample of verbal or non-verbal behaviours. Items of a test are placed in an increasing order of difficulty and its procedure of administration is standardized to ensure maximum objectivity. The psychological tests are used to know the ability of the students, to diagnose the weakness, to predict the future progress, and to provide educational and vocational guidance. The different types of tests are: achievement tests, intelligence tests, attitude tests, aptitude tests, personality tests, creativity tests, etc.

(ii) **Inventories:** Different inventories are used for different traits. Interest inventories are used to measure interest; personality inventories are used to measure certain traits on personality, etc.

(iii) **Observation:** There are certain traits like honesty, punctuality, persistence, truthfulness, etc., which can hardly be measured objectively via tests. So here, observation is an important technique of measurement. The observation may be participant observation or non-participant observation for accurate and scientific observation. One may use observation schedule and other instruments.

(iv) **Interview:** Interview is a face-to-face interaction between one interviewee and one interviewer or more than one interviewer. There are certain things which an individual does not want to express and they can only be assessed through interviews. The interview schedules may be used and the interviewer through a

better personal support, and in congenial atmosphere, can succeed to bring out the inner feelings of the interviewee through carefully planned interviews.

(v) **Checklist:** A checklist consists of a series of items which needs response from the respondent. The presence or absence of an item may be indicated by 'Yes' or 'No' (by a 'O' or 'X' against the items). Checklists are popularly employed for appraisal of studies, school buildings, textbooks, outcomes, instructional procedures, etc.

(vi) **Rating scales:** Psychological traits are relative concepts. So it is very difficult to make watertight compartments between them. Sometimes, the degree of a

trait is necessary on the part of the rater. Rating scale is used to evaluate the personal and social conduct of the learner. We take the opinion of teachers or parents or friends or judges on a particular quality or trait of a pupil along a scale. The rating scale may be of 5 points, 7 points, 9 points or 11 points. For example, to assess a particular trait, we can use a 5 point scale as: very good, good, average, below average and poor. The trait in question is marked by the judges in any one of the five categories. Rating scales can be used to evaluate: personality traits, tests, school courses, school practices and other school programmes.

(vii) **Attitude scales:** Attitude refers to the bent of mind or feelings of an individual towards an object, an idea, an institution, a belief, a subject or even a person. Attitude scales are used to measure this trait objectively with accuracy.

(viii) **Projective techniques:** Projective techniques are very ambiguous and subjective in nature. Through projective techniques, the sub-conscious and preconscious mind of an individual is reflected. For example, with the help of Thematic Apperception Test (TAT), we measure the personality of individuals.

Anecdotal record cards and cumulative record cards are also used for educational measurement and evaluation.

1.3 CONCEPT OF EVALUATION

Evaluation is an act or process that assigns 'value' to a measure. When we are evaluating, we are making a judgment as per the suitability, desirability or value of a thing. In the teaching-learning situation, evaluation is a continuous process and is concerned with more than the formal academic achievement of students. Evaluation refers to the assessment of a student's progress towards stated objectives, the efficiency of the teaching and the effectiveness of the curriculum. Evaluation is a broad concept dealing not only with the classroom examination system, but also evaluating the cognitive, affective and psychomotor domain of the students. The success and failure of teaching depends upon teaching strategies, tactics and aids. Thus, an evaluation approach improves the instructional procedure. Glaser's basic model of teaching refers to this step as a -feedback function'.

American educational psychologist Rober J. M. Bradfield defines evaluation as 'the assignment of symbols to phenomenon in order to characterize the worth or value of the phenomenon usually with reference to some social, cultural and scientific standards'. J.

M. Wright Stone stated, 'Evaluation is a relative new technical term introduced to designate a more comprehensive concept of measurement than is implied in conventional test and examination'. Professor of education I. James Quillea and Paul R. Hanna defined evaluation as 'The process of gathering and interpreting evidence on change in the behaviour of all students as they progress through school'.

Evaluation takes place with the help of test and measurement. In the classroom situation, the teachers first use classroom tests to evaluate the students according to their different traits. After getting the answer papers, the teachers provide some numerals to them. This step is known as measurement. So, measurement deals with only the quantitative description. After the measurement step, the teachers arrange the students in order, such as, first, second, third according to their achievements.

This step is evaluation. So evaluation is a philosophical and subjective concept. It includes both quantitative and qualitative descriptions, and value judgment.

Therefore, Evaluation = Quantitative Description (Measurement) and/or Qualitative Description (Non-Measurement) + Value Judgments.

1.3.1 Nature of Evaluation

The characteristics of evaluation are as follows:

- It is a systematic process.
- It measures the effectiveness of learning experiences provided.
- It measures how far the instructional objectives have been achieved.
- It uses certain tools like tests, observation, interview, etc.
- It is a continuous process.
- It is a subjective judgment.
- It is philosophical in nature.
- It includes quantitative description, qualitative description and value judgment.
- It gets data from measurement.
- It not only determines the magnitude, but also adds meaning to measurement.
- It involves values and purposes.

1.3.2 Classifications of Evaluation

In order to know how far the students have previous learning, to get the knowledge of the effectiveness of instructional objectives, and to diagnose the problems on the way of achievement of students, the teachers use four types of evaluation. These four classifications of evaluation are as follows:

(i) **Placement evaluation:** Through this evaluation, the entire behaviour of the student is assessed. It is like round peg in the round hole and square peg in the square hole. In this ease, the students are given admission to new courses according to their intelligence, attitude, motivation, aptitude, etc. This type of evaluation questions: Does the student possess the knowledge and skills needed

to begin the planned instruction? To what extent has the student already developed the understanding and skills that are the goals of the planned instruction? To what extent do the student's interests, work habits and personality characteristics indicate that one mode of instruction might he better than the other? The goals of placement assessment are to determine for each student the position in the instructional sequence and the mode of instruction that is most beneficial. For example, the B.Ed. entrance test is conducted to give admission to the students in B.Ed. course. This type of evaluation is called 'placement evaluation'.

(ii) **Formative evaluation:** It aims at the evaluation of a student's learning progress during the period of instruction. Formative evaluation views evaluation as a process, and thus, it is an integral part of the learning process. It is not terminal in character. Formative evaluation is concerned with both achievement of students during a course of instruction and its improvement. Formative

evaluation is a tool for providing feedback to the teaching-learning process. It is concerned with the teacher, content, instructional objectives and provision of learning experiences. Formative evaluation also helps the teacher to modify the instructional objectives and the methods of teaching, if necessary. Formative assessment depends heavily on specially prepared tests and assessments for each segment of instruction (e.g., unit, chapter). The unit tests, the weekly tests, monthly tests, etc., are examples of formative evaluation.

(iii) **Diagnostic evaluation:** It is concerned with the persistent learning difficulties that are left unresolved by the corrective prescriptions of formative assessment. It aims at identifying or diagnosing the weaknesses of students in a given course of instruction. Diagnostic evaluation involves the use of specially prepared diagnostic tests and various observational techniques. The aim of diagnostic assessment is to determine the causes of persistent learning problems of students and to formulate a plan for remedial action. When a teacher finds that in spite of the use of various alternative methods and techniques, the student still faces learning difficulties, he takes recourse to a detailed diagnosis. This type of evaluation includes vision tests, hearing tests and other tests used to determine how the student approaches to a reading assignment, such as, whether the student relies on pictures, sound out words, use context clues, skip over unfamiliar words, etc.

(iv) **Summative evaluation:** This evaluation comes at the end of a course of instruction. It is designed to determine the extent to which the instructional goals have been achieved, and is used primarily for assigning course grades or for certifying student mastery of the intended learning outcome. Summative evaluation's chief functions are: 'crediting' and 'certifying' the level of achievement of the students, and selecting the students for different courses. It is judgmental and terminal in character. Summative evaluation judges the achievement of students and the efficacy of school programmes, and guides whether the programme or the system is to be accepted or not. University annual examination is an example of summative evaluation.

1.3.3 Scope of Evaluation from Educational Angle

Anything to be evaluated has certain aims and objectives, and through evaluation, we assess how far these objectives have been fulfilled. From an educational angle, we can evaluate many aspects which is the part and parcel of an educational system, such as:

(i) Evaluation of a school site (with reference to its location, building, hygienic condition, strength of students and teachers, etc.)

(ii) Evaluation of a school programme (school syllabus, co-curricular activities, guidance programmes, etc.)

(iii) Evaluation of teaching methods (with reference to aims, purposes, suitability and efficacy)

(iv) Evaluation of total programme of instruction (with reference to cognitive, affective and psychomotor domain)

(v) Evaluation of school administration, discipline, control, management and organization

(vi) Evaluation of the textbooks

(vii) Evaluation of students' growth, etc.

The steps involved in an evaluation process take place in a hierarchy. These steps are as follows:

• Evaluating
• Planning appropriate learning experiences
• Selecting appropriate teaching points
• Specifying the desired student behaviour
• Identifying and defining specific objectives
• Identifying and defining general objectives

1.3.4 Techniques of Evaluation

The evaluation process produces the data for cognitive, affective and psychomotor objectives. The traditional examinations confine to cognitive objectives only. In this way, the evaluation is a more broad process. Various types of techniques are used in it. These are as follows:

(a) The oral, written, essay type, objective type, practical examination and observation techniques are used for the evaluation of cognitive objectives.

(b) The interest inventory, attitude scale, values test and observation techniques are employed for appraising the affective objectives.

(c) The performance test, practical examination and observation techniques are employed for assessing the psychomotor objectives.

The criterion test is different from the more traditional achievement test. This is because criterion test is concerned with objectives of teaching and instruction, whereas achievement test concerns with content coverage. There are three major characteristics of a criterion test or techniques of evaluation. These are as follows:

(i) *Appropriateness.* The criterion test must cover the terminal behaviour of teaching or instruction. Each item of criterion behaviour should represent the total terminal behaviours.

(ii) *Effectiveness:* The items of the criterion test must have difficulty value and discriminative power. The test should be reliable and valid.

(iii) *Practicability:* The test can be administered easily and the scoring procedure should be simple. The obtained data may be interpreted and may be made meaningful. It should be acceptable to the teacher and the students.

1.3.5 Classification of Evaluation Techniques

All techniques of evaluation can be broadly classified into two categories: (a) quantitative techniques of evaluation and (b) qualitative techniques of evaluation. A brief description of these techniques is as follows:

(a) **Quantitative techniques:** The quantitative techniques are mainly used in educational evaluation. These are highly reliable and valid. They possess all three characteristics and can be classified into three types: (i) Oral, (ii) Written and (iii) Practical.

(i) *Oral techniques:* This technique of evaluation is used at a lower level in organizing and leading teaching activities. The oral questions, debate and drama are used for this purpose.

(ii) *Written techniques:* In this type of evaluation, the written questions are asked and students have to write their answers. The written tests are more effective than oral, The written tests are usually of two types: essay type and objective type. The objective-type tests are now preferred for construction of the criterion test because they are highly objective and reliable.

(iii) *Practical techniques:* In this type of evaluation, some work is assigned to the student to accomplish it. Such techniques are based to assess the skill or psychomotor objectives. This technique is used in Science, Geography, Home Science, Agriculture, etc.

(b) **Qualitative techniques:** The qualitative techniques are used in schools and colleges for internal assessment. These techniques are subjective and less reliable, but they are used for assessing the affective objectives. These techniques are classified into five categories as given below:

(i) *Cumulative records:* Such records are prepared in the schools for each student. The cumulative record of the student includes the educational progress, results of monthly, half yearly test, attendance, participation in games, sports, co-curricular activities and physical health. The cumulative record of the students provides the awareness about their progress and weakness of the parents, teachers and principal.

(ii) *Anecdotal records:* This type of record maintains the description of significant event and work or performance of the students. A merit list is prepared in schools. The correct information is obtained regarding incident and performance of the students in anecdotal records. The interest and learning of the students

can be assessed with the help of such records. Such records can also be used to provide the guidance to students.

(iii) **Observation technique:** This technique is indispensable in school evaluation. It is used at all the stages of education, but it is very useful in evaluating the performance of small children at primary stage. It is used for evaluating cognitive, affective and psychomotor objectives. The student of higher class can make use of this technique for self-evaluation. The classroom interaction can be evaluated only by the observation technique.

(iv) *Checklist:* The technique is used for evaluating interest, attitudes and values of the students. It includes certain statements of 'yes' and 'no' type. The student has to check either of the two.

Each statement of the checklist evaluates a specific objective of learning. Some examples are as follows:

- Are you interested in the steps of teaching?
- Do you take interest in lesson planning?
- Do you really enjoy during classroom presentation?
- Are you interested in encouraging the student activities'?

The above checklist may be used for evaluating teaching interest of teachers.

(v) Raring scale: The rating technique is used for assessing the attitude of the students towards teaching and subjects. It consists of some statements which can be rated on three, five and seven points scale. It is used for higher classes because it requires the power of judgment of the students. A teacher can make use of this technique for evaluating the effectiveness of this instructional procedure, teaching strategy, tactics and aids. The statements of the scale are concerned with the specific objectives of learning.

1.3.6 Types of Evaluation Procedures with Reference to CRE

Tests are tools of measurements, and measurements guide us in evaluation. Different types of tests are in vogue to facilitate the realization of the different purposes of education in the varying contexts of use. They may be categorized as follows:

- Purpose-specific categorization of test-types
- Mode-specific categorization of test-types
- Process-specific categorization of test-types

The tests of one type form a variety absolutely distinct from those of every other type. Also, tests of different types may require the same kind of task, and sometimes even repeat the very same item. That is to say, it is quite possible to find an item recurring in tests which otherwise belong to different types, or even different categories. What distinguishes the test-types, then, is not what is obvious in them but what guided them into being what they are. It is not in the choice of test-tasks, nor in the realization of these tasks in the form of test-items that we are to find the difference. We are to notice the difference: in the overall design of the tests, the purpose that guided them in their construction and sometimes in the nature and extent of coverage of a given area of

learning. With this understanding, we can now discuss the features of design of the different test-types.

(a) Purpose-specific category

Purpose-specific category includes test designed to achieve a specific purpose of evaluation. Generally, four test-types are identified in this category. The features of each of these arc discussed below:

(i) **Diagnostic tests:** These help us identify the areas of learning in which the learner needs a remedial course. They give us a profile of what the learner knows and does not know in a given area of learning. To present such a profile, a diagnostic test has to be a battery of a number of sub-tests with each having a fairly thorough coverage.

(ii) **Aptitude tests:** These tests serve a predictive function. They help us identify potential talents. They identify the prerequisite characteristics which are essential for one to be competent to perform a given task. Presenting items on such subskills which may eventually be developed into expert complex skill, these tests identify those who can do well in a field of study or a profession and those who cannot. These tests are generally used for selecting people for special course/ careers.

(iii) **Achievement tests:** These tests aim to measure the extent to which the objectives of a course have been achieved. The scope of these tests is governed by the objectives of the given course, and they cover only the areas of learning demarcated by the given syllabus. You will learn more about these later in the book.

(iv) **Proficiency tests:** These tests aim to assess the general ability of a person at a given time. Its scope is governed by a reasonable expectation of what abilities learners of a given status (say, matriculates or graduates) should possess. It is not restricted by considerations of the areas covered in any specific course-objectives or syllabus as in the case of achievement tests. While the usual end of course-examination in a school or college may be taken as a typical example of an achievement test, a national level selection or admission test for candidates coming from different states and or university jurisdictions can be taken as a typical example of a proficiency test.

(b) Mode-specific category

Under mode-specific category, we identify test-types on the basis of the mode/attitude that governs the construction and use of a test. Under this category, we present six pairs of test-types along six dimensions.

(i) **Formal assessment vs. informal assessment:** Formal assessment is applicable to a situation where a body answerable to the public is holding a test for a selection or an award. In this kind of situation, there should be objectivity, credibility and relevance in the assessment created. In order to have that, the assessment will have to follow the preset norms/procedures followed in test construction, administration and interpretation.

An informal assessment serves best in situations where a person or entity is holding a test for obtaining certain specific information from the candidates. While an informal assessment should ideally be reliable and objective as well, the evaluator may or may not choose to justify the methodology to anyone. Therefore, an informal assessment may or may not adhere to the standard guidelines of evaluation.

(ii) **Formative assessment vs. summative assessment:** The objective of a formative assessment is to locate the gaps in the student's knowledge SO that the teacher can formulate the Lesson accordingly. On the other hand, the objective of the summative assessment is to certify and grade the student according to the level of learning he or she has attained after undergoing the lesson/training. While formative assessment takes place before instruction begins or during the instruction, summative assessment always takes place at the logical end of the instruction, such as a term or a semester or a year. For example, during a six- month long course, a test every fortnight will be considered formative assessment and the test that takes place at the end of six months will be considered summative assessment. Besides this, the complexity of level of generalization in questions in formative and summative assessments will be entirely different. While formative assessments cover a small part of the course and will have more recall-based questions, summative assessment covers the entire course material and will have recall-based as well as analytical questions, so as to lest the level of learning of the student at the end of the course.

Another kind of assessment used in educational situations is 'developmental assessment'. It is ideal for the time when the course is under development and is undertaken with a sample group of students which represents the general intelligence of the learner. It is considered a constituent of course development programme.

(iii) **Continuous assessment vs. terminal assessment:** While progress or achievement in learning is the concern of formative vs. summative mode of assessment, it is the purpose of grading learner's achievement which guides the continuous vs. terminal mode of assessment. Continuous assessment seeks to spread the basis of grading on a number of tests with regular even intervals, instead of placing it on one end-of-the course test (terminal test). Continuous assessment, thus, allows for more intense accommodation of the learning content in the test process than the terminal assessment normally does. Scores on a series of continuous assessment test, taken together can serve for summative assessment. Taken individually, a continuous assessment test may be used formatively at the time of its administration. In the same way, a terminal assessment may serve the purpose of formative assessment for follow-up courses.

(iv) **Course work vs. examination:** A learner's assessment can be based on course work(s) performed by him during or at the end of a course, or, it may be based on examination(s) taken by him during or at the end of the course. Evaluation of course work or examination at different points of time during a course can be compiled at the end of the course to serve the purpose of summative evaluation.

(v) **Process vs. product assessment:** The basis for evaluating may be either the final product or the result of a given task or the performance at different stages leading to the accomplishment of the task (as in a research work). While evaluating a learner, one may look for the correct solution to a given problem or may take into consideration the correctness of the successive stages followed to solve the given problem (as in problem-solving tasks). If we do the former, we are supposed to be engaged in product assessment; if we do the latter, we are supposed to make process assessment.

(vi) **Internal assessment vs. External assessment:** The mode of assessment is external when the evaluation of a learner's ability is made by an outsider—a person who is not related with the actual process of teaching. The evaluator and the learner are anonymous to each other in this ease. When the assessment is made by the person responsible for effecting the learning being measured, it becomes internal assessment. Formative and summative assessment of both scholastic and non-scholastic abilities is possible in the ease of internal assessment. External assessment serves only summative evaluation of scholastic abilities. These are actually different perspectives along which the assessment of learner ability can be thought of and planned accordingly. It is possible to practically combine two or more of these perspectives in one's approach to assessment. For example, one may include both course work and examination as the basis for learner's assessment, and these two may constitute the units of continuous assessment. One may opt for formative assessment of the process of course work, or may include both internal and external assessment to servo the purposes of formative and summative evaluation.

(c) Process-specific category

Sometimes, test-types are identified on the basis of the process of test-construction. We can talk of two pairs of contrastive test-types in this section.

(i) Teacher made test vs. standardized test: Standardized tests are commercially produced tests adhering meticulously to certain procedures to meet the demands of objectivity and accuracy. They are finalized through the construction procedures of formulating objectives, designing test-blueprints, employing item try-outs, item-analysis and item-revisions. The teacher-made tests, on the other hand, are not governed rigidly by such processes. The teacher, who makes the tests, uses his discretion in matters of the scope of test area and choice of task-types and items. Standardized tests derive their name by the fact that they ensure standardization of the procedures of administration, scoring and interpretation through elaborate specific instructions.

While a teacher-made test is designed to operate within the restricted situation of a given classroom (in terms of test purpose, construction and use), a standardized test is designed for a larger operational situation crossing the barriers of a classroom, an institution or even a region. A standardized test may be chosen for use by different teachers/institutions in different classrooms on different occasions and in different regions.

Criterion-referenced evaluation (CRE): Criteria are the basis for evaluating the progress of students. They identify, in specific terms, the critical aspects of a

performance or product that indicate how well a student is meeting the prescribed leaming outcomes. For example, weighted criteria, rating scales or scoring guides (reference sets) are the ways for evaluating student performance using criteria. It is important to give the student time and support required to respond to formative evaluation by increasing the quality of his work. Criterionreferenced evaluation (CRE) is based on prescribed learning outcomes. CRE is the assessment of performance by judging an individual's behaviour, performance or knowledge against specific criteria or standards.

(ii) **Norm-reference tests vs. CRE:** The purpose of CRE or criterion-reference-test (CRT) is to assess the objective. It is an objective-based test. The objective-type test is objective-centred, assessed in terms of behavioural changes among the students. It assesses the ability of a student against the criterion-behaviour of other learners.

In a norm-reference test (NRT), the purpose is to discriminate between the high-achievers and the low-achievers. Its focus is not on what or how much one has learnt. Its focus is on where one stands in relation to the others' calibre/ level. It assesses the ability of one against the standard 'norm' of achievement in a test.

1.3.7 Functions of Evaluation

The most common purposes and the related functions for which evaluation is used in an educational system are given below:

(i) **Placement function:** This may be the purpose of the evaluation when it helps us select suitable persons for a course or career. To study the entry behaviour of students and to provide suitable measures for individualization of instruction, placement assessment takes place.

(ii) **Diagnostic function:** Through measurement and evaluation, we diagnose the weakness of students and the difficulties in learning and suggest remedial programmes to overcome the weakness.

(iii) **Predictive function:** We also predict the future performance of the students through measurement and evaluation. This function also helps a student in selecting the right electives.

(iv) **Instructional function:** The instructional functions of educational evaluation are to:

- Plan appropriate learning strategies

- Assess different educational practices

- Improve instruction and quality of teachers

- Assess and foster students' growth

- Formulate suitable and realistic objectives of instruction

- Assess the effectiveness of different methods and techniques of instruction

- Provide a basis for modification of curriculum and syllabus

(v) **Administrative function:** The administrative functions are to:

- Appraise the different programmes undertaken

- Appraise the supervisory practices

- Classify the students

- Select students for higher courses of specialized courses or even for some jobs

- Promote students to next higher class

- Have appropriate placement

- Compare the performance of children

- Test the efficiency of teachers in providing suitable learning experiences

(vi) **Guidance function:** Having the knowledge of measurement and evaluation, the teacher can know the level of achievement, motivation, intelligence, aptitude, attitude, etc., of the students. According to the outcomes, the students should be provided educational, vocational and personal guidance.

(vii) Feedback and motivation function: After evaluation, the students are given appropriate feedback to educational progress, and by it, they get motivated towards teaching-learning process.

(viii) Communication function: Educational measurement and evaluation helps in communicating the results of progress to the students, to parents and to the school.

(ix) Grading and certification: This purpose is served when evaluation helps in ranking the learners of a given group. Usually, the terminal examination of a course serves this purpose.

1.3.8 Comparison of Grading System and Marking System

The recent movement in grading has been the result of an honest confession of our inability to measure with precision the human qualities with all their intricate levels of variation and mix-up of observable and non-observable attributes. It has reduced the wide 101 point range of scale to a viable short range. Though grading gives only rough estimates of learner ability, the estimates are more realistic and reliable than the false precision in terms of marks. In other words, grading reduces considerably the chances of error of judgment (however, it also suffers from inaccuracy, which even otherwise cannot be overcome).

The possibility of an error in judgment in awarding one grade-point in the place of another is relatively much less than possibility of awarding 59 marks instead of 60 marks and vice versa. Thus, the 'narrow-range' scale used in grading minimizes the inter-examiner differences in evaluation. Besides, when narrow-range scale is applied, chances of utilizing the whole range are more than when a wide-range scale is used. Irrespective of the subject-variation, the narrow-range scale may, thus, lessen the disparity in the value of scores in different subjects. Thus, the practice of grading tends to overcome the errors of the traditional marking system.

Different bases of grading

Whatever be the mode, all grading involves the use of a set standard. When we say something is 'good', the judgment implies an implicit or overt rating against a set standard— a physical or conceptual model. Similarly, when we grade scores, we describe them to be good, or average or poor when we compare the scores with some standards. These standards may be what the others in the class have scored or what is the expected ideal performance. Thus, basically, grading may be done along two lines, using two different kinds of standards: (i) relative grading standard (i.e., keeping the performance of the whole group of test-takers in the given task as a standard), and (ii) absolute grading standard (i.e., keeping an ideal conception of what a good performance on a given task should be as a standard).

(i) **Relative grading standard:** When a relative grading standard is applied, the learners get grades according to their ranks in relation to the other learners taking the same test. The classical form of relative grading proceeds from a pre-determined plan of learner distribution. It begins with a specification of what percentage of learners are to be grouped under each of the different grades of a chosen scale (the scale may comprise 5 or 7 or 9 grades).

Figure 1.1 represents some distribution modes.

You can notice how the distribution is even and balanced on either side of the central block, which constitutes the 'average group'. The extreme right hand side block gets the same percentage of learners as the extreme left hand side block in each scale. Similarly, the second block from the right end is equal to the second block from the left end. This design of distribution deliberately attempts to put equal number of learners on either side of a fixed average group of learners.

If the distribution is 'represented on a graph-sheet, we will get what is known as a 'normal distribution curve'. It will be a bell-shaped curve evenly shaped on either side of its vertical middle line.

5 Point				7	24	38	24	7		
Scale				A	B	C	D	E		
7 Point			3	7	22	36	22	7	3	
Scale			O	A	B	C	D	E	F	
9 Point		4	7	12	17	20	17	12	7	4
Scale		A	A+	A	B+	B	C+	C	D	E

Better than the —   Average Group   Poorer than the
average group —   average group

Fig. 1.1 Distribution of Learners of Different Scales of Relative Grading

If there are 300 learners taking a test, the distribution of learners in each percentage (i.e., grade) group will be as given in Table 1.1 (if a 7-point scale is used).

Table 1.1 Distribution of Test-takers on a 7-point Scale

Grade	Percentage	No. of Students out of 300

O	3	9
A	7	21 (High)
B	22	66
C	36	108 (Average)
D	22	66
E	7	21 (Low)
F	3	9

The learners are rank-ordered on the basis of their performance (score) in the test. The top nine learners in the rank-order are given the grade '0', the next 21 in the rank order are given the grade 'A', and so on.

Limitations of relative grading: This kind of normative distribution of grades will be useful to universities, boards of examinations, recruitment/selection boards, and other testing agencies who have to deal with a large number of candidates (say, tens of hundreds at a time) and have to make pass-fail/select-reject decisions, and do not have to find 'what' or 'how much' of a 'subject' has been mastered by the candidates. The latter considerations are the typical concerns of a class teacher, both in formative and summative evaluation, and applied in these situations, the relative grading following normative distribution is bound to have limitations.

(a) Relative grading of a learner's performance in a class test will be inadequate in its message to the learner, the teacher, the parents and whosoever concerned with progress in learning. The position of a learner on the normal curve is determined by the overall performance level of the class. So a grade without any information about the performance of a class as a whole will not convey much sense pedagogically. A learner who gets 'A' in one class may, when placed in another, get 'F'! So the A's and Fs are meaningless, unless they are accompanied by some information about performance ability of the group to which they belong.

(b) Under 'normative distribution' (another description of relative grading) scheme, each grade has to be assigned to a fixed number of learners in the class. This makes farce of the idea of a teacher attempting to help the 'poor' performers do better or the high-achievers 'improve' their performance. Besides, it also tends to create bad attitude among learners. When a leaner manages to come up from a lower grade, it naturally pushes down someone else to a lower grade. A learner may, therefore, develop negative attitude and be tempted to hope that his classmates do as poorly as possible so that his efforts will be better in comparison. This can inhibit learners from helping one another and can even hurt social relations among them.

(ii) **Absolute grading standards:** While applying relative grading standards, we fix for each grade the number of learners to be assigned. However, when we apply absolute grading standards, we fix for each grade a range of scores for a learner to credit with, in order to claim the grade mentioned against them:

Table 1.2 shows the distribution of a scorer on absolute grading standards.

Table 1.2 Distribution of Scorer on Absolute Grading Standards

	Grade	Range of Scores
	O	(90-100)%
	A	(80-89) %
	B	(70-79) %
Average	C	(60-69) % Reference point
	D	(50-59) %
	E	(40-49) %
	F	Less than 40%

In another form of absolute standards, one may fix the quantum of content and/ or the level of mastery of given skills in a relative hierarchy of the groups of relevant attributes (like creativity, intelligence, imagination, resourcefulness, simplicity, etc., besides relevance and appropriateness of the content and the form expected of the responses) for each grade to be awarded. The grading scheme following such absolute standards in terms of qualitative (in the place of percentage) values is sometimes called 'criterion-reference grading'.

While the percentage standards can be applied to grade the performance on objective (fixed-response) items, the qualitative standards facilitate grading the performance on subjective (free-response) items.

The absolute grading standards are ideal to evaluate and grade students. However, there are some difficulties in putting them into practice. The fixation of standards (quantitative and qualitative) has to be flexible and accommodative as the learners' scores depend on the facility level of the test administered. Formulating reasonable base for the standards is not easy at all levels and all types of courses. They may be done with relative case at higher levels and in vocational courses rather than at lower levels and in non-vocational courses.

Process of grading

The attempt to develop a narrow-range grading, as a consequence of realizing the weaknesses of marking on a 101-point scale, has led to the suggestion for grading on different scales and of adopting different approaches. The approach has to be suitable to the type of composite tests made up of different item-types that are most commonly in use today. It allows the use of different short-range scales to suit the types of responses expected of different item-types. It is more reliable also because the choice of scale for each item-type is governed more or less by the number of 'test-point' it presents.

For instance, in an objective type item, the response can either be right or wrong. There cannot be any range of quality within which the responses can vary. So responses to such questions can be graded only on a two-point scale (0-1), where '1' represents an acceptable response and '0' an unacceptable one. However, in the case of short-answer (free-response) type question, the range of coverage of the subject matter may be restricted to a few value-points, say, 3 or 4. In such a case, both a 7- point scale and a 2-point scale may tend to enhance imprecision. The levels of quality of the response may not vary so much as to warrant a 7-point scale, and at the same time, the responses may

not be totally devoid of variations so as to allow for a binary operation of a 2-point scale. So a scale in between has to be applied say, a 3-point scale (0-2), where '0' can mean an unsatisfactory response, '1' an answer of mediocre quality, and '2' an answer of appreciable quality.

While evaluating a long-answer (essay) type question, we may come across responses of varying qualities, and we may use a 5-point (7- or 9- point) scale. The grading can be done taking into account some relevant attributes, such as adequacy relevance of the content, clarity of the form, intelligence or imagination of the presentation.

1.4 NEW TRENDS IN EVALUATION

Higher education in India has seen many changes in the last 50 years. After the Kothari Commission report in 1966, a debate on college autonomy began and a few colleges became autonomous since 1978, and recently, a few of them have completed 25 years of this freedom. Academic freedom has led colleges to formulate new curricula, start relevant courses, design new syllabi and establish new evaluation techniques. These have made the Indian education system more comprehensive and modem. Let us learn more about the new methods in the following sections.

1.4.1 Choice Based Credit System (CBCS)

Research and experience has revealed that the learner-centric contextual curriculum recommended and the desired learner outcomes projected can be achieved mainly through Choice Based Credit System (CBCS). The system, which is implemented in most universities of North America and Europe, is considered a fundamental tool for transforming the orthodox and outdated system of higher education in India. CBCS functions on a modular pattern based on module/units called 'credits'. Here, credit defines the quantum of contents/syllabus set for a course/paper and decides the minimum number of teaching-learning hours required. One credit denotes 15 hours of instructions per semester. CBCS permits students to:

1. Learn at their own pace.

2. Select subjects from a broad range of electives offered by the university.

3. Opt for supplemental/value added courses and obtain more than the requisite credits, based on the learner's capabilities.

4. Implement an interdisciplinary approach in education.

5. Utilize the expertise of faculty across the university apart from the specific department faculty.

6. Acquire knowledge, skill and attitude of learning outcomes through participatory teaching and learning and continuous evaluation process.

1.4.2 Continuous and Comprehensive Evaluation

The term Continuous and Comprehensive Evaluation (CCE) refers to the two important and essential conditions of evaluation. These are: (i) continuous evaluation and (ii) comprehensive evaluation. The meaning of continuous evaluation is that the evaluation of learners at levels of schools should be done on continuous basis, i.e., throughout the year or session, rather than only at the end of a session, course or class. The evaluation

system for students' achievement in all domains of personality development should be scattered throughout the course, such as on weekly basis, monthly basis, quarterly basis or unit-end test, and chapter-end test. This type of evaluation is called continuous evaluation and it is done by the teachers of the school; therefore, it is also called internal assessment or evaluation. Internal evaluation mechanism puts teachers as well as students focused on teaching and learning and, therefore, gives better results as compared to one-time evaluation which is done at the end of the course. The whole academic session is utilized properly by the teachers and the students.

The other aspect is comprehensive evaluation which means that the evaluation process should cover all aspects of personality development of the students, i.e., scholastic as well as non-scholastic areas. There have been drawbacks in our evaluation system for a longtime. The main flaws have been that the evaluation system has been lopsided, i.e., evaluating only one or two domains, the cognitive and psychomotor. Even in these two domains, it only evaluates the cognitive domain in depth and psychomotor domain very little. It neglects the affective domain completely. Under comprehensive evaluation of learners, there is a consideration that their interests, attitudes, values, personal and social qualities, and proficiencies in co-curricular activities are taken proper care of along with the development of cognitive and psychomotor abilities, i.e., knowledge, understanding, application, analysis, synthesis, evaluation and skill in the area of study. Continuous and comprehensive evaluation is essential for promoting the all-round development of the learners. The all-round development of the children or learners at any level of education involves the development of both scholastic and non-scholastic areas of pupils' personality, that is development in all the three domains—cognitive, affective and psychomotor domains of personality.

H. S. Srivastava, ex-professor and head of NCERT's department of educational measurement and evaluation, has given a scheme of continuous and comprehensive evaluation in his book titled Comprehensive Evaluation in Schools published by NCERT (National Council of Educational Research and Training) in 1989. Table 1.3 depicts the scheme of continuous and comprehensive evaluation adapted from the scheme given in H. S. Srivastava's book.

Table 1.3 Scheme of Continuous and Comprehensive Evaluation

SCHOLASTIC AREAS

A. Curricular Areas
1. Knowledge
2. Understanding
3. Application
4. Analysis
5. Synthesis
Diagnostic
6. Evaluation
7. Skills

I. Evaluation Techniques:
(i) Written
(ii) Oral
(iii) Practical or Performance Test
II. Tools for Evaluation:
(i) Teacher-made test (Questionnaire,

Unit and Performance Test)
(ii) Standardized achievement test
(iii) Assignments or home-work
(iv) Quizzes
(v) Debate and group discussion
(vi) Extempore
(vii)   Seminar

III. Periodicity of Evaluation:
    (i) Weekly, bi-weekly, monthly, quarterly, etc.
    (ii) Terminal tests
    (iii) Random tests
IV. Coverage: All students

B. Intelligence

  I. Evaluation Techniques:
    (i) Performance tests
    (ii) Paper-pencil tests
  II. Tools of Evaluation: Standardized tests of
intelligence and   performance

  III. Periodicity of Evaluation: Yearly, at the
beginning of the academic   year.

  IV. Coverage: All students

NON-SCHOLASTIC AREAS

A. Personal and
   Social Qualities:
   1. Regularity
   2. Punctuality
   3. Discipline
   4. Cleanliness
   5. Emotional stability
   6. Initiative
   7. Cooperation and
      coordination
   8. Sense of responsibility
   9. Hard-working
   10. Civic sense
   11. Spirit of social service
   12. Tense or relaxed
   13. Patient or impatient
   14. Achievement orientation
   15. Domineering

I. Techniques of Evaluation: Observation
   (Direct and Indirect)
II. Tools of Evaluation:
   (i) Anecdotal records
   (ii) Rating scales
   (iii) Observation schedule
   (iv) Checklist
III. Periodicity of Evaluation:
   (i) Regularly (Continuous evaluation)
   (ii) Grading once in each term based on
      continuous evaluation
IV. Coverage: All students

NON- SCHOLASTIC AREAS

B Interests: These interests and attitudes should be graded on
5-,7-9 point scales.   1. Literary   If neither positive nor
negative aspect is observed in a student for any   2.   Scientific
   particular personal or social qualities, interests and attitudes, then the   3.
   Musical   student should be given the average grade or
middle value.
   4. Artistic
   5. Social service
C. Attitude :
   1. Towards teachers
   2. Towards parents and society
   3. Towards studies
   4. Towards class/schoolmates

5. Towards school activities
6. Towards school assets
7. Towards nature and environment

D. Physical Health :
1. Height
2. Weight well as physical
3. BMI (Body Mass Index)
4. Chest expansion
5. Physical defects (if any)
6. Eye defects
7. Hearing capacity medical specialist
8. Teeth
9. History of disease months or as fixed by (if any)

I. Techniques of Evaluation:
(i) Medical check up by doctor
(ii) Observation by teachers (Class teacher as education teacher)

II. Tools of Evaluation:
(i) Rating scales
(ii) Observation schedule
(iii) Medical instruments to be used by the doctor

III. Periodicity of Evaluation: Once every six the higher authority

IV. Coverage: All students

OTHER AREAS (Educational, Sports, Cultural Activities etc.)

A. Literary and Scientific
1. Library work
2. Debate
3. Recitation
4. Creative writing
5. Talks
6. Clubs month
7. Museums minimum

B. Cultural cultural activities
1. Drawing and painting
2. Drama
3. Music
4. Dance
5. Sculpture
6. Artistic embroidery

C. Outdoor Activities
1. Games and sports
2. Swimming
3. Gymnastics
4. N.C.C.
5. N.S.S. students and minimum

I. Techniques of Evaluation:
(i) Observation

II. Tools of Evaluation:
(i) Observation schedule
(ii) Rating scale
(iii) Anecdotal records

III. Periodicity of Evaluation: At least once every

IV. Coverage: Library work for all students and one of the remaining literary, scientific and for each student

I. Techniques of Evaluation:
(i) Proficiency test
(ii) Observation

II. Tools of Evaluation:
(i) Observation schedule
(ii) Anecdotal record
(iii) Rating scales
(iv) Performance test

III. Periodicity of Evaluation: Monthly

IV. Coverage: Games and sports for all

6. Scouting   one of the remaining activities for each student
7. First aid
8. Junior Red Cross
9. Community surveys and services
10. Community-based craft

Other elements that make up the system of continuous and comprehensive evaluation arc assignments, periodic tests, progress report and cumulative records. Let us study each of these in detail.

(a) Assignments

According to Chambers 21st Century Dictionary, 'assignment' means 'a task or duty that has been selected for someone to do' or an 'an exercise that is set for students'. It is used as a teaching method in all subjects at secondary and senior secondary levels. Considering the subject of study, nature of content and the time available, the subject teacher determines the number of assignments for the academic calendar. While designing assignments, the subject teacher takes care of the abilities, interests and levels of the class and the group who has to work on the assignment.

An assignment combines the merits of lecture-cum-demonstration method and the individual laboratory method. The success and the effectiveness of this method totally depends on proper development and planning of the assignments.

The assignment given to students to work at secondary and senior secondary levels may be classified into the following two categories:

1. Home assignment

This is the assignment given to the student to complete at home and includes writing of answers to questions, problem solution, reading text, etc. These assignments are given by teachers. Along with the assignment, the teacher also gives references from different sources concerning the topic so that the students can take help from them. Learners go through the textbooks, reading materials and other sources referred by the teacher and grasp the idea of the assignment. They may consult libraries, seniors and other sources available to them. They prepare the answers to the questions or solve the problem given to them by the teacher. They bring the assignment file to the school and hand it over to the concerned teacher for evaluation. The teacher evaluates the answers, gives grades or marks as per the provision and tries to find out discrepancy (if any) therein. The teacher also gives feedback to the students for improvement in future. The feedback may include:

- Further reading required by the student

- Any textbook which should have been consulted by the student

- Suggestion whether the answers were up to the mark or needed improvement

- Advice on what special matter could have been added in the answers

- Suggestion about the presentation of ideas/answers, etc.

- Any other suggestion that may have been considered by the teacher or the examiner

2. School assignment

It is the assignment which is prepared or planned at home and performed or conducted in the school or laboratory. Assignment of this category includes performing experiments in the laboratory and preparing answers of the questions that have been put up by the teacher. School assignments can be divided into the following two parts:

(i) **Preparatory part:** In this part, the teacher demonstrates experiments to the class. He conducts those experiments in front of students which are difficult, complex and carry some risk, or those experiments which require costly apparatus to be used with precautions. Students observe the demonstration and participate in the discussion with the teacher. Those experiments which are simple in nature, are assigned to the students. They are instructed to study about these experiments at home as well as in school (if time allows). The teacher gives references of useful textbooks and other sources for this purpose. Detailed instruction regarding the experiment is given in the assignment and the pupil works accordingly under the guidance of the teacher.

(ii) **Laboratory part:** Before entering into the laboratory, students hand over the notebooks to the teacher. They are supposed to answer a few questions before entering into the lab related to the experiment to be conducted by them. The teacher evaluates the notebooks. If the teacher gets satisfied with the work in the notebooks and with the oral answers of the students, he or she allows them to enter the lab and conduct a planned experiment. The teacher may not allow a student to do practical if his theory work or preparatory part of the school assignment is not completely satisfactory.

Objectives of assignment method of teaching

The assignment method of teaching has the following objectives:

- To develop habit of self-study among students
- To nurture the habit of inquiry, hard work and taking initiatives
- To develop self-reliance and self-confidence among the students
- To assure the proper utilization of students' leisure time
- To develop hidden abilities of the students
- To develop the attitude of self-help among students

When should an assignment be given?

An assignment is normally given to the students on the following three occasions:

(i) At the beginning of the unit with the aim to prepare students in advance for further studies.

(ii) During the teaching-learning of a unit of the content when various problems come up under one umbrella concept. A part of the problem is solved by the teacher in the class and rest are given to the students as assignments.

(iii) At the end of the section or unit to evaluate students' learning of various concepts through the assignment.

Key points for planning an assignment

The success of the assignment totally depends on how the teacher plans, executes, supervises and evaluates it. The teacher should keep in mind the following key points while preparing an assignment:

- It should be planned and allotted considering the age, intelligence, interests, abilities, aptitudes, etc.. of the students.
- It should be challenging to students to motivate them. It should be neither too tough nor too easy.
- It should be planned with the cooperation and discussion of the pupils concerned.
- The teacher should ensure the availability of related resources before finalizing an assignment. It is useless to give an assignment for which no resource is available.
- Complete instructions must be given to students regarding the work in the assignment. The instructions must not be vague, ambiguous or confusing.
- The laboratory assignment should clearly specify the precautions to be taken by the students, possible sources of errors in the assignment or experiment, method of collecting data or record, the tabular columns to be drawn for data feeding and the number of observations to be recorded.
7. Details of the apparatus required for the experiment should be mentioned clearly.
8. It is better to indicate the allotted time for every part of the assignment separately.
9. Questions for the students framed by the teacher must fulfil the following criteria:

    • They should be comprehensive in nature, i.e., They should cover all portion of the content.

    • They should be able to test the understanding and application ability of the pupil.

    • They should be objective in nature.

    • They should involve testing laboratory applications and precautions.

    • They should involve drawing and labelling skills (if possible).

Salient features of a good assignment

The following are the salient features of a good assignment:

- It should be appropriate to the unit under study.
- It should have clearly stated and realistic objectives.
- It should stimulate reflective thinking, imagination and creativity among students.

- It should take care of the individual differences among students, i.e., it should be designed considering the age, level, intelligence, abilities, available resources and interests of all the students.
- The language of the assignment must be simple, non-ambiguous and to the point.
- The instructions must be very clear.

Merits of assignment method

The assignment method has the following merits :

- It is based on 'learning by doing' approach.
- Students work at their own pace and time.
- This method takes care of the individual differences among pupil.
- It nurtures the habit of self-reliance, self-dependence and self-confidence among students.
- Opportunity of individual attention to the students is possible.
- Laboratory facilities can be utilized effectively.

Demerits of assignment method

The assignment method has the following demerits:

- Generally, it is taken very casually. Teachers do not work properly on assignments; only students do on their own. Teachers check the assignment seriously. The reason behind this flaw is that it takes much of the teachers' time.
- Chances of copying are much higher in this method of teaching.
- It hinders in timely completion of the syllabus.
- It requires well-equipped library and laboratory, and resourceful teachers, which are rarely available to the students.
- All content area cannot be taught through this method. It has its own limitation.

Practical suggestions for making assignment method a success

Through the following suggestions, the assignment method can be made successful:

- Apply this method in secondary and senior secondary classes because students at this level are mature enough to work on assignments.
- The teacher must plan well and select the relevant topics which may be completed through assignments. While selecting topics for assignments, students' interest, intelligence, age, aptitude, creativity and resource availability must be kept in mind.
- Teacher should be satisfied with the preparatory work of the students before giving them permission for carrying out practical in the lab or field work.
- Progress of the students in the assignments must be checked by the teacher on regular basis.

- Teacher should be available to the students at all time for any kind of help they may need. It makes students confident and ensures proper help and guidance in the hour of need.

(b) Periodical and annual tests

The tests which are conducted on regular intervals are referred to as periodical tests. These tests are very important as far as the continuous and comprehensive evaluation is concerned. Some of the important types of periodical tests are as follows:

1. **Weekly test:** It is conducted every week on a particular day fixed by the teacher.

2. **Unit test:** It is conducted at the end of every unit. In general practice, this test is conducted on any one day of the week on a regular basis. Students generally call it by the names of the day on which it is conducted like Monday Test', Tuesday Test', etc. Generally, weekly tests and unit tests are the same.

3. **Monthly test:** This test is conducted on monthly basis. The content of this test covers that part of the syllabus which has been covered during the month.

4. **Quarterly test:** This test is conducted in every three months. The syllabus of the test covers that part of the syllabus which has been completed during the three months' period.

5. **Half-yearly test:** This test is conducted in every six months and it constitutes that part of the syllabus prescribed for the six months.

6. **Annual test:** This is the final test or summative test for the whole academic year. It covers the entire syllabus and is conducted to judge the overall achievement of the students. Results are declared on the basis of this test.

The periodical tests are very important as far as the achievements of the students are concerned. Teachers as well as students work very hard throughout the year for achieving the objectives of education in that particular academic year. Periodical tests expect regular hard work on the students' part and provide them motivation and encouragement for learning. They reduce the loss of time and energy of the students.

(c) Progress reports

A progress report is very important as it shows the progress of a student in his or her areas of study and various activities related to the curriculum or course. It is important due to the following reasons:

- It provides the students with a kind of motivation, satisfaction and makes them aware about the level of their performance.

- It gives feedback to the students for their efforts in learning.

- It makes students aware about the weak areas so that they can work on them properly.

- Parents get an opportunity to be acquainted about the progress of their wards.

- With the help of a progress report, parents can arrange the necessary study support for their sons and daughters.

- The school administration gets to know about the performance of its students as well as its teachers in various subjects.
- Remedial classes may be arranged by the schools for weak students as per the records brought out by the progress report of the students.

A student's 'cumulative record' is the record of all the personal and social qualities, values, attitudes, co-curricular activities and his or her behaviour in the classroom. This kind of record helps in the overall development of a student.

(d) Cumulative records

The following are the important features of the cumulative records:

- A cumulative record helps in appreciating each student's academic and behavioural performance in the class.
- After constant observation and complete evaluation, every student can be categorized into special groups in the classroom on the basis of his or her interests and abilities.
- Cumulative records can contain a wealth of information for teachers and other staff to help the students in implementing various effective instructions and strategies to support them in their academic and social development in and beyond the classroom.
- The record includes the recording of all the activities, special interests, hobbies and personality traits of a student in a systematic order.
- It is important for teachers to have a complete understanding of the learning abilities of every student to maintain his or her cumulative record.
- It is an effective current and future guide in tracking the progress of a student as it serves as an additional information base for providing vocational and academic guidance to the student.
- Special areas of academic or behavioural weaknesses can be identified in the student's cumulative record and the teacher can design a programme to include corrective actions in order to minimize the various weaknesses.
- The cumulative record helps teachers to adopt different teaching methods or strategies according to the measured skill levels of students.
- The cumulative record also helps teachers to adopt suitable, remedial teaching measures and curriculum modification in addressing student's learning difficulties.
- A student's cumulative records can provide information for the effective execution of behaviour modification strategies in a class.

1.4.3 Question Bank

As the name suggests, a question bank is a collection of questions from a particular subject, covering each and every unit or chapter of the course. The questions are designed in such a manner that it covers all chapters and all content areas of the subject. It includes all types of questions like objective types and descriptive types.

Among the objective-types items, it may include multiple-type questions, fill in the blank type question, true and false type questions, matching-type question and some other types as the question bank developer thinks suitable. Among the descriptive type items, the main types of questions are very short-answer type questions, short- answer type questions, long-answer type question and very long-answer type questions and essay-type questions.

Question banks help students and teachers in many ways. Teachers can use questions from the question banks in evaluating student's performances and in diagnosing their weaknesses. Students may also use question banks for testing themselves. If they find any difficulty in solving any problems from the question bank, they consult their teacher and get the problems discussed and solved. Question banks also help students in developing confidence as they evaluate themselves by solving the problems given in it. They indicate the coverage of syllabus to the students. As it contains questions of varying difficulty levels, the students and the teachers can use a question bank in a better way for proper guidance to all students having the individual differences.

A question bank plays vital role for those students who are preparing for any competitive examinations like engineering entrance test, medical entrance tests, etc.

To prepare a question bank, you must:

1. **Ensure that there are all types of questions:** Objective-type questions, short-answer questions, long-answer questions, questions requiring illustrations, diagrams etc. The main thing is that the question bank must be exhaustive as far as the content coverage is concerned.

2. **Ensure that there are all levels of questions:** There must be all levels of questions in a question bank. These levels pertain to complexity, from very simple to very complex questions, in order to prepare the student fully for examination.

3. **Ensure there are questions from previous years' question papers:** Sometimes, examiners setting papers use cryptic language to form questions. Students should be used to interpreting complicated language in questions, and this can be achieved by allowing them to practice with questions from old exam papers.

4. **Provide answers at the end:** A good question bank always contains answers at the end. Otherwise there is no way for the student to confirm if he/she has answered correctly. For mathematical problems or Physics numerical questions, there need not be the entire solution, but just the final answer.

1.5 SUMMARY

In this unit, you have learnt that:

• Measurement answers the question of 'how much'. The application of the principles of measurement in the field of education is known as educational measurement.

- To measure the psychological traits with validity and reliability, the measuring instrument or tests should be far away from the aspects like-personal errors, variable errors, constant errors and interpretative errors.

- There are four types of scales of measurement: (i) nominal, (ii) ordinal, (iii) interval and (iv) ratio.

- The first function of measurement is prognosis function; the prognosis has the administrative function such as classification, selection, promotion and gradation of students.

- The prognosis function reveals the level of a student with regard to certain characteristics, whereas the diagnosis function identifies the weakness of the student-learning.

- The third major function of measurement is verification of scientific hypotheses of research.

- A checklist consists of a series of items which needs response from the respondent; these are popularly employed for appraisal of studies, school buildings, textbooks, outcomes, instructional procedures, etc.

- Evaluation refers to the assessment of students' progress towards stated objectives, the efficiency of teaching as well as the effectiveness of the curriculum.

- There are four types of evaluation: (i) placement, (ii) formative, (iii) summative and (iv) diagnostic.

- Evaluation Quantitative Description (Measurement) and/or Qualitative Description (Non-Measurement) + Value Judgments.

- Rating scale is used to evaluate the personal and social conduct of the learner.

- Formative evaluation views evaluation as a process, and thus, it is an integral part of the learning process.

- The aim of diagnostic assessment is to determine the causes of persistent learning problems of students and to formulate a plan for remedial action.

- All techniques of evaluation can be broadly classified into two categories:

  (a) quantitative techniques of evaluation and (b) qualitative techniques of evaluation.

- Different types of tests are in vogue to facilitate the realization of different purposes of education in varying contexts of use. These may be: (i) purpose-specific categorization of test-types, (ii) mode-specific categorization of test- types and (iii) process-specific categorization of test-types.

- Formative assessment is concerned with identifying a learner's weaknesses in attainment with a view to help the learner and the teacher overcome/remedy those, while summative assessment aims at certifying and grading the attainment of the learner at the end of a given course.

- Continuous assessment seeks to spread the basis of grading on a number of tests with regular even intervals, instead of placing it on one end-of-the course test (terminal test).

- Formative and summative assessment of both scholastic and non-scholastic abilities is possible in the case of internal assessment.

- Though grading gives only rough estimates of a learner's ability, the estimates are more realistic and reliable than the false precision in terms of marks.

- When a relative grading standard is applied, the learners get grades according to their ranks in relation to the other learners taking the same test.

- A grade without any information about the performance of a class as a whole will not convey much sense pedagogically.

- Comprehensive evaluation is concerned with scholastic and non-scholastic areas.

- Home assignment is the assignment that is given to a student to be completed at home and includes writing of answers to questions, problem solution, reading text, etc.

- The success of the assignment totally depends on how the teacher plans, executes, supervises and evaluates it.

- Periodical tests are very significant as far as the continuous and comprehensive evaluations are concerned.

- A student's 'cumulative record' is the record of all the personal and social qualities, values, attitudes, co-curricular activities and his or her behaviour in the classroom.

1.6 KEY TERMS

- **Measurement:** It is the process by which the attributes or dimensions of some physical object are determined.

- **Quantification:** It indicates to what extent a particular attribute is present in a particular object.

- **Ordinal scale:** It corresponds to quantitative classification of a set of objects with reference to some attribute or rule.

- **Evaluation:** It is the assessment of a student's progress towards stated objectives, the efficiency of the teaching and the effectiveness of the curriculum.

- **Achievement test:** It is a test that measures the outcomes of learning after the teaching-learning process is over.

- **Periodical tests:** These are the tests that are conducted on regular intervals in the classes or labs.

- **Cumulative record:** It is the record of a student's personal and social qualities, values, attitudes, co-curricular activities and his or her behaviour in the classroom.

1.7 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The two basic types of measurement are: (i) physical measurement and (ii) mental measurement.

2. The important characteristics of a good measuring tool are as follows:

   (i) It should be valid.

(ii) It should be reliable.

(iii) It should be objective.

(iv) It should be usable and practicable.

(v) It should be comprehensive and precise.

(vi) It should be easy in administering.

(vii) It should be economical from the viewpoint of time, energy and money.

(viii)   It should be easily available

(ix) Its get up or appearance must be good and attractive.

3. Stanley Smith Stevens' levels of measurement are: (i) nominal scale, (ii) ordinal scale, (iii) interval scale and (iv) ratio scale.

4. Lee J. Cronbach (1949) classified all applications of mental measurement under the following three main functions: (i) prognosis function, (ii) diagnosis functions and (iii) research function.

5. The methods of measurement are as follows: (i) tests, (ii) inventories, (iii) observation, (iv) interview, (v) checklists, (vi) rating scales, (vii) attitude scales and (viii) projective techniques.

6. The four types of evaluation are: (i) placement, (ii) formative, (iii) diagnostic and (iv) summative.

7. 'Errors of measurement' can be classified into four types: (i) personal errors, (ii) variable errors, (iii) constant errors and (iv) interpretive errors.

8. Basically, grading may be done along two lines, using two different kinds of standards:

(i)   Relative grading standard (i.e., keeping the performance of the whole group of test-takers in the

given task as a standard).

(ii)   Absolute grading standard (i.e., keeping an ideal conception of what a good performance on a   given task should be as the standard).

9. Criterion-referenced evaluation is the assessment of performance by judging an individual's behaviour and performance or knowledge against specific criteria or standards.

10. The meaning of 'continuous evaluation' is that the evaluation of learners at levels of schools should be done on a continuous basis, i.e., throughout the year or session, rather than only at the end of a session, course or class.

11. H. S. Srivastava, ex-professor and head of NCERT's department of educational measurement and evaluation, is the author of Comprehensive Evaluation in Schools.

12. Home assignment is the assignment given to a student to complete at home and includes writing of answer to question, problem solution, reading text, etc. These assignments are given by teachers.

13. The two features of cumulative records are as follows:

    (i) A 'cumulative record' helps in appreciating each student's academic and behavioural performance   in the class.

    (ii) The record includes the recording of all the activities, special interests, hobbies and personality   traits of a student in a systematic order.

1.8 QUESTIONS AND EXERCISES

Short-Answer Questions

1. List the characteristics of good measurement.
2. Enumerate the types of scale for measurement.
3. List the techniques of evaluation.
4. Write a note on criterion-referenced evaluation.
5. Differentiate between physical and mental measurement.
6. How should continuous and comprehensive evaluation be implemented in schools?
7. Differentiate between home and school assignments.
8. List the salient features of a good assignment.
9. List the importance of a progress report.
10. What is a question bank?

Long-Answer Questions

1. Write a detailed note on educational measurement.
2. Describe the various scales of measurement.
3. Elaborate on the different tools of measurement.
4. Explain the functions of evaluation.
5. Describe the standard scheme of continuous and comprehensive evaluation as given by H. S. Srivastava.

1.9 FURTHER READING

Annastasi, A. 1976. Psychological Testing, 4th Ed. New York: McMillan Publishing Co.

Block, J. H, L. W. Anderson. 1975. Mastery Learning in Classroom instruction. New York: McMillan Camp.

Bloom, B. S. 1956. Taxonomy of Objectives of Cognitive Domains. New York: Longmans Green &Co.

Bloom, B. S. 1956. Taxonomy of Objectives (Affective Doming). New York: Longmans Green & Co.

Chronbach, L. J. 1970. Essentials of Psychological Testing, 3rd Ed. New York: Harper.

Catel, R. B. Personality: A Systematic Theoretical and Factual Study. New York: McGraw Hill.

Ebel, R.L, 1972. Essentials of Educational Measurement. New Jersey: Englewood Cliff.

Edwards, A.L. 1957. Techniques of Attitude Scale Construction. New York: Application Century.

Frank S. Freeman. Theory & Practice of Psychological Testing. Mumbai: IBII Publishing Co.

Grunlund, N. E. 1970. Stating Behavioural Objectives for Classroom Instruction. New York: Mcmillan.

Grunlund, N. E. 1973. Preparing Criterion Referenced Tests for Classroom Instruction. New York: Mcmillan.

Grunlund, N. E. 1976. Measurement and Evaluation in Teaching (3rd Ed). New York: Mcmillan Publishing Co.

Goslin, D. A. 1967. Teachers and Testing. New York: Sage Foundation.

Grunlford, J. P. 1967. The Nature of Human Intelligence. New York: McGraw Hill.

Grunlund, N. E. 1959. Sociometry in the Classroom. New York: Harpt & Row.

Julian, C., Stanley and Hopkins. 1972. Educational and Psychological Measurement and Evaluation. New Delhi: Prentice Hall of India Pvt. Ltd.

Lindermon, Richard, H, Educational Measurement. Mumbai: B. S. Taraporevala Sons & Co.

Lynum, H. B. 1971. Test Scores and What They Mean. New Jersey: Prentice Hall.

Ten Brink, T. D. 1974. Evaluation: A Practical Guide for Teachers. New York: McGraw Hill.

Bernard, H. W., D.W Fuilmer. Principles of Guidance: A Basic Text (Indian Edn.). New Delhi: Allied Pub.

Crow Lester, D., A. Crow. An Introduction to Guidance Principles and Practices. New York: American Books Co.

Downing, Lester, N. 1964. Guidance and Counselling Services. New York: McGraw Hill.

Fusterm J. N. Psychological Counselling in India. Mumbai: McMillan.

Kochhar, S. K. 1987. Educational Vocational Guidance in Secondary Schools. New Delhi: Sterling Publisher Pvt. Ltd.

Hoose, William, N. Van. (ed). 1979. Counselling and Guidance in 20th Century. Boston: Noughton Muffin Co.

Khorshed, A. W. Guidance Movement in India, Guidance Services. New Delhi: NCERT.

Pandey, K. P. Educational and Vocational Guidance in India. Varanasi: Vishwavidyalaya Prakashan.

Pasricha, Pren. 1976. Guidance and Counselling in India Education. New Delhi: NCERT.

Swamy, R. V. (ed). 1971. Guidance Services in Colleges and Universities. Bangalore: University and Directorate of Employment and Training.

Traxier, R. E., R. D. Norla. Techniques of Guidance. New York: Harper and Road.

UNIT 2 TAXONOMY OF EDUCATIONAL OBJECTIVES

Structure

2.0 Introduction
2.1 Unit Objectives
2.2 Educational Objectives and their Classification
2.3 Cognitive Domain
    2.3.1 Classification of Cognitive Domain
    2.3.2 Assessment of Cognitive Learning
    2.3.3 Thinking Skills Required in Cognitive Learning
2.4 Affective Domain and Formulation of Specific Objectives
    2.4.1 Classification of Affective Domain
    2.4.2 Assessment of Affective Learning
    2.4.3 Attitude and Values
2.5 Psychomotor Domain
    2.5.1 Classification
    2.5.2 Assessment of Performance or Psychomotor Learning
2.6 Changing Concepts of Educational Objectives and Recent Changes in Instructional Objectives
    2.6.1 Functions of Educational Objectives/Usefulness of Taxonomical Classification
    2.6.2 Guidance Functions of Evaluation
2.7 Summary
2.8 Key Terms
2.9 Answers to 'Check Your Progress'
2.10 Questions and Exercises
2.11 Further Reading

2.0 INTRODUCTION

Learning constitutes one of the most significant achievements of a person's life. However, learning is dependent on a lot of factors and one of them is the process of assessment. Assessment can contribute to learning in a variety of ways and is of paramount importance for teachers, students and academicians. Assessment of learning focuses on the ways students can be assessed so as to improve the learning process.

In this unit, the three dimensions of learning-cognitive, affective and psychomotor-are discussed. Besides this, you will learn about how these three types of learning are assessed.

2.1 UNIT OBJECTIVES

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

- Explain the cognitive domain and how its objectives are formulated and assessed
- Discuss the affective domain and how its objectives are formulated and assessed
- Describe the psychomotor domain

2.2 EDUCATIONAL OBJECTIVES AND THEIR CLASSIFICATION

The study of classification is known as taxonomy. You have various objectives of education related to personal needs, social needs, community needs, national needs, global needs, and various aspects to your personality. The first and the most successful effort in this direction was made by American educational psychologist Benjamin Bloom and his co-worker, who began their work in 1948. Bloom was teaching educational psychology at the University of Chicago. He developed and proposed the classification of educational objectives in 1956, which is popularly known as Bloom's Taxonomy of Educational Objectives. This is the classification of the different aims and objectives, which are required to be achieved through curriculum of any course and involves knowledge, skills, attitude values and abilities that a teacher, teacher educator or curriculum developer sets for students as teaching-learning objectives. This classification fits properly for all subjects and courses whether academic or technical. In all the areas of study, you try to develop content knowledge and understanding, skills, and values related to it, You will find these in the classification of educational objectives by Bloom.

Bloom's taxonomy of educational objectives also fits properly with Mahatma Gandhi's views on education. Mahatma Gandhi defines education as, 'By education, I mean the all-round drawing out the best in child and man-body, mind and soul'. Gandhi's definition remarkably reflects the idea that education should develop the whole personality of the child by developing the three most important dimensions of personality, i.e., body, mind and soul. Mind indicates intellectual aspects; soul indicates the emotional, spiritual and moral aspects; and body indicates the development of skills which requires the use of body parts like fingers, hands, legs, eyes, vocal cord etc., for performing various skills.

Bloom classified educational objectives into three 'domains', namely:
(i) cognitive domain, (ii) affective domain (iii) psychomotor domain, as shown in Figure 2.1.

Educational Objectives

Cognitive Domain	Affective Domain	Psychomotor Domain
Objectives concerning	Objectives concerning emotional, spiritual and	Objectives concerning skill aspects of

intellectual or mental aspects of personality	moral aspects of personality	personality

Fig. 2.1 Domains of Educational Objectives

All the three domains are not divided entirely by a water tight compartment. Each of the three domains is interrelated and achievements in one domain influence the achievement by other domain. However, all domains develop in a hierarchical order and constitute the whole personality of an individual. If an individual develops his or her three domains in any area of study, he or she is having a good personality, and if any one domain is underdeveloped, then he or she has not nurtured himself or herself as a groomed personality. Hierarchical means of learning at the higher level in any domain is entirely dependent on having acquired prerequisite knowledge and skills at lower level. The three domains may be represented in short as:

1. Cognitive domain (about knowing-intellectual aspects of personality of an individual)

2. Affective domain (about attitudes, feelings, interests, values and beliefs of an individual)

3. Psychomotor domain (about doing-skill aspects of the personality of an individual)

All the three domains have their own taxonomy or classification. In all domains, the levels of expertise or abilities are arranged in order of increasing complexity, i.e., in hierarchical order of difficulty levels. Learning outcomes that require higher levels of expertise require more effective teaching and more sophisticated classroom techniques and method of teaching.

Dimensions of learning basically form a framework of learning focused on preparing instructional planning, keeping in view cognition (the awareness part) and learning in practical classroom situations. This framework serves three major purposes. These are as follows:

1. It provides a process for planning and delivering curriculum and instruction that integrates much of the research on effective teaching and learning.

2. It offers a way of integrating the major instructional models by showing how they are connected and where the overlaps occur.

3. It provides a framework for organizing, describing and developing research- based teaching strategies that engage students in the types of thinking that can lead to meaningful learning.

The following five aspects of learning should be considered while finalizing curriculum, instruction and assessment:

• Attitudes and perceptions about learning

• Using knowledge meaningfully

• Identifying productive habits of the mind

• Acquiring and integrating knowledge

- Extending and refining knowledge

2.3 COGNITIVE DOMAIN

Cognitive domain includes those objectives of education which attempt to develop mental faculties or intellectual abilities, i.e., the ability of knowing, understanding, thinking and problem solving. It develops our factual/knowledge, conceptual understanding and all levels (lower, middle and higher) of thinking. It covers whole of mental abilities and mental operations.

2.3.1 Classification of Cognitive Domain

The classification of cognitive domain was done in 1956 by Bloom and is commonly known as Bloom's Taxonomy of Cognitive Domain (Bloom et al, 1956). Cognitive taxonomy or classification of cognitive domain has knowledge-based goals. This domain has been classified into six major categories which are arranged in a hierarchical order based on the levels of complexities of cognitive or mental or intellectual tasks or operations. These are arranged from simple to complex and concrete to abstract starting from category one to category six, respectively.

1. **Knowledge:** This is the first and the lowest level of cognitive domain. It represents memory and constitutes recall and recognition of various facts, concepts, principles, theories and laws of physical science. No addition or deletion is done in this category; we simply recall and recognize things. In the revised Bloom's taxonomy, this category is given new name 'remembering'.

Example: The symbol of iron is 'Fe'.

The chemical formula of sulphuric acid is 'H_2SO_4'.

Laws of motion were given by 'Isac Newton'.

A magnet has two poles-N (north) and S (south)

2. **Understanding:** This is the second level of cognitive domain and develops only after the development of the first category, i.e., knowledge or remembering in any particular area of study, also in physical science. Learners are expected to go beyond the level of recall and recognition. After having developed this level of understanding on any topic, learners become capable of doing the following major tasks, which in turn indicates that the learners have acquired the level of understanding in a given topic:

(i) Translate, summarize or define the acquired knowledge in their own words.

(ii) Describe, elaborate, extrapolate and explain natural phenomena or events or process or method, etc.

(iii) Interpret the acquired information or knowledge in their own way and give their own examples. They can discriminate or differentiate between two or many objects or concepts. Classify and categorize various objects into groups on the basis of some criteria. Verify and generalize facts and concepts.

Example: After having understood the structure of atom, learners not only recall protons, electrons and neutrons but also describe the structure of an atom. Now learners can also explain why an atom is neutral with the help of charges acquired by every fundamental particle and numbers of all three particles in an atom.

3. **Application:** After having acquired knowledge and understanding levels of any topic (may be a fact, concept, principle, theory or law), learners should be able to apply them in their day-to-day lives. Application of any concept, principle, theory or law in daily life and solving problems of varied nature is impossible without its knowledge and understanding. Unless the learner is able to apply whatever knowledge and understanding he or she has acquired, it has no meaning at all and indicates that the learners have not understood the content properly by applying or implementing the gained knowledge and understanding of various contents, you can solve many problems of daily life, under concrete and abstract situations.

   Example: If learners know and understand the importance of natural resources, underground water crisis, electricity supply and demand relationship and other such problems of daily lives, they will take care of these things in their day-today life, and by applying this understanding, they will try to minimize wastage of water and electricity in their homes, schools and society by proper and judicial use of these things.

4. **Analysis:** This is the fourth higher level category of cognitive abilities. At this stage, learners develop the potential to analyse and breakdown the whole into its various components or constituents and detect the relationship and organization of its various components. Learners develop the ability to break a law and theory into its various inherent facts, concepts and principles on the basis of which that theory or law has been created or proposed.

   Example: Learners are taught about the laws of motion. Suppose they know and understand the third law of motion which states, 'to every action there is an equal and opposite reaction'. They have also developed the ability to apply this knowledge and understanding in their daily lives. If the analytical ability has developed, they would be able to analyse this law in the event of some likely situations. They would also be able to describe its every concept like action and reaction. One can analyse anything if he or she has knowledge and understanding of that thing and also has the potential to apply it. In the process of analysis, three tasks are performed in general. These are as follows:

   (a) Analysis of elements or constituents making the whole

   (b) Analysis of relationship among various constituents

   (c) Analysis of the organizational patterns of the constituents

5. **Synthesis:** This is the process of putting together of various constituents to make a whole. This is a higher level thinking ability and is complex in nature, which involves the creation of a new pattern or structure by manipulating various constituents. It has the elements of creativity attached with it. Development of creative personality requires this level of cognition to be achieved by the learners. All creative people have this ability in common. Synthesis involves the following three things:

- Development of a unique communication
- Development of a plan, procedure or proposed set of operation
- Development of a set of abstract relations

6. **Evaluation:** This is the process of judgment about the worth or value of a process ora product. It includes all the content, i.e., facts, concepts, principles, theories and laws of physical sciences. It is the highest and the most complex level of cognitive ability and involves all the five categories discussed earlier, It is a quantitative as well as a qualitative process. It leads to the development of decision-making ability among the learners and involves judgment in terms of internal as well as external criteria.

2.3.2 Assessment of Cognitive Learning

Cognitive learning is defined as the acquisition of knowledge and skills by the mental process. Cognition includes representation of physical objects and events and other information processing.

The cognitive learning is a result of listening, watching, touching or experiencing. Cognitive learning is mainly awareness towards the environment—the environment that provides some meaning to acquired knowledge.

After these activities, it is important to process and remember the information. No motor movement is needed in cognitive learning and the learner is active only in a cognitive way. Cognitive learning includes symbols, values, beliefs and norms.

When we say the word 'learning', we usually mean 'to think using the brain'. This basic concept of learning is the main viewpoint in the Cognitive Learning Theory (CLT). The theory has been used to explain mental processes as they are influenced by both intrinsic and extrinsic factors, which eventually bring about learning in an individual.

CLT implies that the different processes concerning learning can be explained by analysing the mental processes first. It posits that with effective cognitive processes, learning is easier and new information can be stored in the memory for a long time. On the other hand, ineffective cognitive processes result in learning difficulties that can be seen anytime during the lifetime of an individual.

Insight and motivation

In CLT, motivation and purpose are much the same. The learner's goal is the end result he anticipates and desires. The goal controls the behaviour of the learner. The teacher's most important responsibility is to help the learner find worthwhile goals which may be clear and realistic. They should recognize and use the prominence of social and achievement motives in school learning. The teacher must know about what is familiar to students and then he must introduce elements of novelty but not too rapidly. A good teacher must pace his presentation to maintain the interest and attention in learning.

The teachers' management of conflicting motives may be an important factor in student's success.

Guiding exploration and action

The cognitive theorists in teaching of reading begin with thoughts interesting and understandable to the learner. In every type of instruction, we start with meaningful wholes.

- An attempt is made to focus attention on elements and relationship that determine the correct response.

- Teachers' guidance must match the students' level of thought or ways of working. If a student has not advanced above the level of concrete thinking, information presented symbolically will not help him.

- The teacher can help students to find purpose or order in learning.

The formation of concepts may be regarded as the organization of experience. The teacher's role is to use appropriate means to clarify the critical features of both old and new experiences.

2.3.3 Thinking Skills Required in Cognitive Learning

The following critical thinking skills are required in cognitive learning:

1. Convergent thinking

Convergent thinking is a term introduced by American Psychologist Joy Paul Guilford. It generally means the ability to give the 'correct' answer to standard questions that do not require the utilization of much creative faculty. Divergent and convergent thinking skills are both important aspects of intelligence, problem solving and critical thinking. Bringing facts and data together from various sources and then applying logic and knowledge to solve the problem to achieve an objective is known as convergent thinking.

Assessing convergent thinking: Convergent thinking can be assessed by administering standard IQ (Intelligent Quotient) tests, by various recognition or knowledge tests, logical discussions and by giving some problems to the students.

2. Divergent thinking

Divergent thinking is thinking outwardly instead of inwardly. It is the ability to develop original and unique ideas and then come up with a solution to a problem or achieve an objective.

The goal of divergent thinking is to generate many different ideas about a topic in a short period of time. It involves breaking a topic down into its various component parts in order to gain insight into the various aspects of the topic. Divergent thinking typically occurs in a spontaneous, free-flowing manner, such that the ideas are generated in a random, unorganized fashion.

Self analysis

The following questions may help to find out the potential of a learner:

- What are my activities during a normal day?

- What do I know about?

- How do I spend my time?

- What are my areas of expertise?

- What would I like to change in my world or life?
- What are my strongest beliefs and values?
- What am I studying in school?
- What bothers me?
- What do I like? What are my hobbies? What are my interests?

Topic analysis

The following questions may help in refining a large topic into a specific, focused one:

- What are the most important aspects of some specific things?
- What are the effects of a thing?
- How has an object changed? Why are those changes important?
- What are the different aspects of anything you can think of?
- What are the smaller parts that comprise an object?
- What do I know about a thing?
- What suggestions or recommendations can be made about that thing?
- Is something good or bad? Why?

Techniques to stimulate divergent thinking

The following are some of the techniques used to stimulate divergent thinking:

- **Brainstorming:** This is a technique which involves generating a list of ideas in a creative, but unstructured manner. The goal of brainstorming is to generate as many ideas as possible in a short period of time. All ideas are recorded during the brainstorming process.

- **Recording ideas:** By recording ideas, one can create a collection of thoughts on various subjects that later becomes a source book of ideas.

- **Free writing:** The idea is to write down whatever comes to mind about the topic, without stopping. This can help generate a variety of thoughts about a topic in a short period of time.

- **Mind or subject mapping:** This involves translating brainstormed ideas into the form of a visual map or picture.

3. Critical thinking

Critical thinking is described as reasonable reflective thinking focused on what to believe and what not to believe. It is that mode of thinking — about any subject, content, or problem — in which the thinker improves the quality of his or her thinking by skillfully taking charge of the structures inherent in thinking and imposing intellectual standards upon them. Michael Scriven, British polymath and academic, and Richard Paul, an internationally recognized authority on critical thinking, believe that critical thinking is the intellectually disciplined process of:

- Analysing the situation
- Synthesizing two or more pieces of information

- Applying knowledge

- Active and skilful conceptualization

- Evaluating gathered information based on certain universal intellectual values

Characteristics of critical thinking: It comprises various modes of thinking, such as scientific thinking, mathematical thinking, historical thinking, anthropological thinking, economic thinking, moral thinking and philosophical thinking. The following are its four characteristics:

1. It is self-guided, self-disciplined thinking which attempts to reason at the highest level of quality.

2. Critical thinking of any kind is never universal in any individual; everyone is subject to episodes of irrational thought.

3. It is affected by the motivation.

4. Critical thinking can be seen as having two components: (a) a set of information and belief generating and processing skills, and (b) the habit, based on intellectual commitment to use those skills to guide behaviour.

Importance of critical thinking skills: Critical thinking skills are essential to:

- Learn how to approach problems logically and confidently

- Balance using both the right and left sides of the brain

- Make wise decisions in life

- Put oneself on the path to knowledge

The list of core critical thinking skills includes observation, interpretation, analysis, inference, evaluation, explanation and meta-cognition. There is a reasonable level of consensus among experts that an individual or group engaged in strong critical thinking gives due consideration to the procedure involved.

Critical thinking calls for the ability to:

- Recognize problems

- Find workable means to solve those problems

- Comprehend and use the text with accuracy and clarity

- Interpret data to evaluate arguments

- Recognize unsaid assumptions

- Understand the importance of steps in problem solving

- Collect relevant information

- Recognize the existence (or non-existence) of logical relationships between propositions

- Draw conclusions

- Cross-check conclusions

4. Problem solving

Problem solving is a mental process that involves discovering, analysing and solving problems. The ultimate goal of problem-solving is to overcome obstacles and find a solution that best resolves the issue.

Formalized learning theory developed in late 1930, when proponents of various approaches attempted to build their own theory to explain the problems of learning. A theory of learning cannot be defined to satisfy all interested persons. We can quote the definition of a theory as 'a provisional explanatory proposition or set of propositions, concerning some natural phenomena and consisting of symbolic representations of:

(a) the observed relationships among independent and dependent variables, (b) the mechanisms or structures presumed to underlie such relationships, or (c) inferred relationships and underlying mechanisms intended to account for observed data in the absence of any direct empirical manifestations of the relationships' (Learning Theories edited by Melvin H. Marx).

Approaches to problem solving

Traditionally, two different approaches have been mentioned by psychologists, adhering to two families of learning theories: (a) cognitive field theory, and (b) stimulus-response theory.

Cognitive field theory emphasizes the importance of perception of total situation and relationship among its components, and restructuring the cognitive field. German Psychologist and Phenomenologist Wolfgang Kohler conducted his classical experiments on a chimpanzee named Sultan to study the process of problem solving in animals. He, from his study on problem solving, proposed that solution of a problem is arrived at, all of a sudden, after some initial efforts by the individual. Many studies have been conducted on children and adults which confirm that solution of a problem is reached, all of a sudden, through insight into the situation.

The second point of view has been advanced by stimulus-response theorists who emphasize the importance of trial and error. They hold that a problem is solved through a gradual process of elimination of errors and putting together correct responses. There has been considerable controversy as regards the superiority of one approach over the other as an interpretation of problem solving. Some psychologists are of the opinion that cognitive field theory, approach is most effective for solving problems which require higher mental processes, while stimulus-response approach is effective for solving simple problems.

To do away with the controversy of cognitive and stimulus response theorists' approach, American psychologist Harry Harlow (1959) proposed a third explanation. His approach is more realistic and rational in nature. He conducted a series of experiments on monkeys and human subjects of low mental abilities. He presented his human subjects with simple problems of discrimination. He observed that in the beginning, his subjects showed trial and error behaviour to solve a series of problems, but he noticed that when similar problems were presented to the subjects in future for the first time, they made correct discrimination. The later stage appears to be insightful learning, that is, suddenly getting the problem solved. According to Harlow, the underlying assumption is that in the previous trial and error learning, the subjects have learned 'how to learn'. They acquired what he called a learning set. They acquired a method of learning that transferred positively to other problem situations of similar type.

Harlow says, 'Generalizing broadly to human behaviour, we hold that original learning within an area is difficult and frustrating, but after mastery of the basic facts, learning within the same area becomes simple and effortless.'

The steps in problem solving

In order to correctly solve a problem, it is important to follow a series of steps. Many researchers refer to this as the problem-solving cycle, which includes developing strategies and organizing knowledge. The following are the steps required in the process of problem solving:

1. **Identifying the problem:** While it sounds like the simplest thing to do, it actually is not. It is very common for people to identify the reason incorrectly, which obviously renders all following efforts useless.

2. **Defining the problem:** Once the problem has been correctly discovered, the next step is to define it. Defining the problem carefully offers solutions and insights into the problem.

3. **Forming a strategy:** Next, a strategy must be devised which would be as per the individual's mindset, attitude, experience and available resources.

4. **Organizing information:** The strategy should ideally lead to a solution, but first, the information needs to be organized in a coherent manner, such that it leads naturally to the best possible solution.

5. Allocating resources: it is good to allocate more resources to solve a problem, so that the available resources can be used to find out a solution.

Problem solving techniques

These techniques are usually called problem-solving strategies. These are as follows:

- **Brainstorming:** Suggesting a large number of solutions or ideas and combining and developing them until an effective solution is found

- **Abstraction:** Solving the problem in a model of the system before applying it to the real system

- Assessing the output and interactions of an entire system

- **Lateral thinking:** Approaching solutions indirectly and creatively

- **Dividing and conquer:** Breaking down a large, complex problem into smaller, solvable problems

- Employing existing ideas or adapting existing solutions to similar problems

- **Analogy:** Using a solution that solved an analogous problem

- **Hypothesis testing:** Assuming a possible explanation to the problem and trying to prove one's perspective

- Synthesizing seemingly non-matching characteristics of different objects into something new

- Transforming the problem into another problem for which solutions exist

- Eliminating the cause of the problem

- Testing possible solutions until the right one is found

5. Decision making

Decision making can be regarded as the mental processes resulting in the selection of a course of action among several alternative scenarios. The end result of each decision making process is a final selection. The output can be an action or a suggestion.

Steps in decision making

The following are the steps that are to be followed in the decision-making process:

- Objectives are to be established first.
- Objectives must be classified and placed in the order of importance.
- Alternative actions must be developed.
- The alternative must be evaluated.
- A tentative decision can be made.
- The tentative decision is evaluated and analysed.
- Few steps are followed that result in a decision model. This can be used to determine an optimal plan.
- In a situation of conflict, role-playing can be used.

Each step in the decision-making process may include social, cognitive and cultural problems. It has been suggested that becoming more aware of these obstacles allow one to better anticipate and overcome them. The following few steps can help in establishing good decision making:

- Creating and nurturing the relationships, norms and procedures that will influence bow problems are understood and communicated.
- Recognizing that a problem exists.
- Identifying an explanation for the problem and evaluating it.
- Finding more suitable justice among many responses.
- Following through with action that supports the more justified decision. Integrity is supported by the ability to overcome distractions and obstacles, developing implementing skills and ego strength.

Decision-making stages

There are four stages that are involved in all group decision making. These stages, or sometimes called phases, are important for the decision-making process to begin. These were developed by communication researcher B. Aubrey Fisher. The four stages are as follows:

1. **Orientation stage:** This is introductory stage, when members meet for the first time and get to know each other.

2. **Conflict stage:** Once group members become familiar with each other, disputes, little fights and arguments take place. Nevertheless, group members eventually work it out.

3. **Emergence stage:** The group begins to clear up vague opinions by talking about them.

4. **Reinforcement stage:** Members finally make a decision, while justifying to themselves that it was the right decision.

2.4 AFFECTIVE DOMAIN AND FORMULATION OF SPECIFIC OBJECTIVES

The affective domain of Bloom's taxonomy of educational objectives is related to the development of emotions, values, attitudes, and the development of those aspects of personality which are more influenced by heart than the mind. It also includes the development of interests, appreciation, feelings, likes and dislikes towards something.

2.4.1 Classification of Affective Domain

The classification of this domain was done by American educational psychologists D. R. Krathwohl, B. S. Bloom and B. B. Masia in 1964. The categories in this domain are also arranged hierarchically from the lowest to the highest level of complexity.

1. **Receiving:** This is the ability, inclination and readiness of learners to receive information. It requires attention, awareness, listening, seeing and willingness on the part of the learners. These are preconditions of learning, personality development, and imbibing culture and values. It needs sensitization of learners to stimuli, phenomena or environment. On the whole, learners should be made receptive in their habit and attitude, Whatever you want learners to learn, you should make them receptive toward those things.

Examples:

• Reading newspapers, magazines, journals, books, reports, etc., of interest to the learner

• Watching news, shows, reports, programmes as per interest

• Listening patiently and attentively to teachers, parents, seniors, friends and more experienced persons

• Having curiosity to learn from various sources

2. **Responding:** This is the second level objective under affective domain. Learners are required to be responsive along with being receptive; otherwise it will not serve the purpose. Responding behaviour reflects that the learners are receiving or trying to receive. Continuity in attention and motivation behaviour (receiving) leads to the development of responding behaviour.

This category of ability is represented by interest, which is the tendency to respond to a particular object or event or situation. This creates the way for two-way communication and facilitates the process of teaching and learning. Students 'listen' to the teachers attentively and 'respond' to them to give their reflection and share their experiences.

Example:

- Response of students in class
- Interaction of students with teachers, friends and seniors on various issues or problems
- Visit to clubs, libraries, museums and other knowledge resource centres
- Participation in various activities, competition, seminars, conferences, cross word and such other programmes

3. **Valuing:** During the cyclic process of receiving and responding, learners are automatically inclined towards taking value judgment regarding the things they are concerned with. These things may be an object, an event, an idea, a rule, any ritual, a set norm or any traditional or modern aspects of our culture. Through the process of valuing, individuals set guidelines for regulating their own behaviour. Character formation or value inculcation in the growing generation is done through the following three sequential steps:

(a) Value acceptance

(b) Value preference

(c) Value commitment

Example: A class is taught by several teachers. All teachers practice various values in which some are common and some are unique for individual teachers. Students attend their classes and interact with them. They observe and analyse various values being practiced by their teachers. Through regular observation and analysis, students develop their own value based on their preference, acceptance and commitment.

4. **Organization:** Through the process discussed above, students absorb various values from their teachers, parents and society. They analyse various values absorbed from different sources and finally construct relatively enduring value system through the process of synthesis and organization of values for a balanced conduct and behaviour pattern. This leads to the development of a set value structure or philosophy of life for every individual. It assists individuals in decision-making process about conduct in real life situations and in forming opinions on major issues of social and personal concern.

5. **Characterization of values or value complex:** This is the highest level category of objectives under affective domain. At this level, individuals develop a set of values, attitudes and beliefs for themselves that build their character and give shape to their philosophy and personality. This process goes on continuously throughout life resulting into the shift of preference of various values, depending upon situation, age and experience.

2.4.2 Assessment of Affective Learning

The affective domain given by Krathwohl, Bloom and Masia includes the manner in which we deal with things emotionally, such as feelings, values, appreciation, enthusiasm, motivations and attitudes. The five major categories are listed from the simplest behaviour to the most complex in Table 2.2.

Table 2.2 Five Major Categories Describing the Affective Domain

Category	Key Words (Verbs)
1. **Receiving phenomena**	Asks, choose, describe, follow, give, identify,
(Awareness, willingness to hear)	locate, name
2. **Responding to phenomena**	
(Active participation on the part of the learners.	Answer, assist, aid, compile, conform, discuss,
Attends and reacts to a particular phenomenon)	greet, help, perform, practice, present, read, recite,
	report, select, tell, write
3. **Valuing** (Valuing is based on the internalization	Complete, demonstrate, differentiate, explain,
of a set of specified values, while clues to these	follow, form, initiate, invite, propose, read, report,
values are expressed in the learner's overt	select, share, study, work
behaviour and are often identifiable)	
4. **Organization** (Organizes values into priorities	Adhere, alter, arrange, combine, compare,
by contrasting different values, resolving conflicts	complete, defend, explain, formulate, generalize,
between them, and creating a unique value	identify, integrate, modify, order
system)	
5. **Internalizing values**	Act, discriminate, display, influence, listen,
(characterization): Has a value system that	modify, perform, practice, solve
controls their behaviour	

2.4.3 Attitude and Values

'An attitude can be defined as a positive or negative evaluation of people, objects, events, activities, ideas, or just about anything in your environment.'

—Philip George Zimbardo, 1999

In the opinion of American sociologist Read Bain, attitude is 'the relatively stable overt behaviour of a person which affects his status'. An attitude is a state of mind or a feeling or disposition. It is important to have a positive attitude about work.

Values

Affective learning also involves internalizing a set of values expressed in behaviour. Teachers affect the values and behaviour in a student by setting examples by their own behaviour.

Attitude formation

Attitudes are expected to change as a function of experience, environment and education. American psychologist Abraham Tesser has argued that hereditary variables may also affect attitudes.

Measurement of attitude

A number of techniques for measuring attitudes are in use. However, they all suffer from different kinds of limitations. Largely, the different types of techniques focus on the components of attitudes, namely the cognitive, the affective and the behavioural components. The two basic categories that attitude measurement methods can be divided into are as follows:

1. Direct measurement, such as Likert scale
2. Indirect measurement, such as projective techniques

Direct observation

This is a simple and logical method which records the behaviour patterns of people under study. This method is widely used for various purposes. However, even if the individuals to be studied are easily accessible, observing the behaviour of a large sample of individuals is not practically feasible.

Direct questioning

This method involves asking pre-set questions on certain topics on which the individual's behaviours are to be evaluated. While it seems like the most straightforward approach to simply ask questions to test attitude, the results may not be accurate because an individual may try to hide his or her real opinions and attitudes.

Some other approaches

In projective techniques, attitude gauging objects are hidden and results are interpreted on the basis of pre-set criteria. While this technique overcomes some limitations of the direct observation technique, the projective technique falls short when it comes to objective and reliable interpretation of data.

Thurstone scale

The first official method for measuring attitude was formulated by Louis Leon Thurston, a US pioneer in the fields of psychometrics and psychophysics, in 1928. His objective was to measure people's attitudes regarding religion. The Thurstone scale contains statements regarding the issue in question and every statement is assigned a numerical value based on the value the evaluator considers it to have. When people have selected the statements for each question, the values are added up and the average is calculated, which corresponds to a particular attitude.

Likert scale

Liken introduced the Liken scale of attitudes in *A Technique for the Measurement of Attitudes*. The scale covered a range of attitudes, from extremely positive to neutral to extremely negative. This scale also consists of statements and the subjects are asked to express their opinion regarding the statement through a 0—5 point scale. Once each statement has been assigned a numerical value, again they are added up and the mean is calculated. This kind of scale is often used in career assessment programmes to gauge a learner's interests and tendencies, so that they can help themselves select the right career path.

The Likert scale is applied in the form of questionnaires. A Liken scale questionnaire would contain a statement, which would need to be evaluated by the individual on the basis of the following kind of responses:

(a) Strongly disagree

(b) Disagree

(c) Neither disagree nor agree

(d) Agree

(e) Strongly agree

The individual will tick or circle one response for each statement or question. Each statement/question and its possible responses are together known as a Liken item. A Liken scale is, in turn, a sum of responses to multiple Liken items. A Liken scale is considered a 'balanced' form of attitude testing because it contains an equal number of positive and negative options.

In some cases of Liken scale questionnaires, the neutral response, such as 'neither disagree nor agree', is removed. Such an evaluation method is known as the forced choice method, because the individual cannot remain neutral on any statement. The forced choice method is used when it is thought that the individual might select the neutral response to avoid controversy or hide ignorance.

Generally, three kinds of common biases may render Liken scale questionnaires unreliable. These are as follows:

- **Central tendency bias:** In a professional setting especially, the individual may be reluctant to admit strong feelings about an issue.

- **Acquiescence bias:** The individual may feel obliged to agree with the presented statement.

- **Social desirability bias:** The individual may select an option to preserve the self-image or to be popular.

Once the respondent gives in the complete questionnaire, the responses may be evaluated individually or in a grouped form, depending on the pattern being studied. When studied in a grouped form, the items provide a score, which can be categorized and evaluated. The Likert scale is also sometimes known as the Summative Scale.

Interest and its measurement

The reason for finding out the interest of the students is to help them in getting their dream careers. This inventory helps the teachers in knowing the interest areas of their students so that they can encourage and get various opportunities to grow in a particular field.

Thus, an interest inventory is often used in career assessment. The goal of this assessment is to give insight into the students' interests, so that they may face less difficulty in deciding on an appropriate career choice for themselves. It is also frequently used for educational guidance as one of the most popular career assessment tools. The test was developed in 1927 by psychologist E. K. Strong to help people who were quitting the military jobs to find suitable jobs.

Supporting students

Prior to selecting a career, students need to identify the right path for themselves. This can be done through an assessment, which would help them get an insight into their own

interests, preferences and personal styles. Analysing these aspects will direct them into identifying the right courses, jobs, internships and activities that are suitable for them.

Self-concept and its assessment

Self-concept defines how we assess ourselves as individuals or what we think of ourselves. Self-concept is a commonly used term. There are two aspects to the development of self-concept. These are as follows:

1. **The existential self:** This aspect of the self can be defined as, 'the sense of being separate and distinct from others and the awareness of the constancy of the self'. (Bee 1992)

2. **The categorical self:** Once a child realizes that he has a distinct identity, he gradually become aware of the world around him and his own place in the world. He starts relating to the world and starts thinking of himself as more than his physical characteristics, such as hair colour or height, etc. Finally, he becomes aware of the fact that others perceive him in a certain way, which may or may not be similar to how he perceives himself.

According to American psychologist Carl Rogers, self-concept has three different components. These are as follows:

1. The view you have of yourself (self-image)

2. How much value you place on yourself (self-esteem or self-worth)

3. What you wish you were really like (ideal self)

Factors affecting self-concept

The following factors affect self-concept:

1. **How others react to us:** People's approval and recognition and affection helps to develop a positive self-image.

2. **Comparison with others:** Comparing oneself with people who seem to be doing better financially or more popular socially than oneself can lead to a negative self-image. However, if the comparison is made with people who are less successful, it would lead to a positive self-image.

3. **Social roles:** There are some roles that are associated with prestige and positive self-image, such as that of a doctor.

4. **Identification:** Self-concept is also influenced by the role we play in the group that we belong to.

2.5 PSYCHOMOTOR DOMAIN

Psychomotor domain is concerned with those objectives which are intended to develop various skills. For example: typing, painting, drawing, dissecting, preparing food, beautification, carpentry, bookbinding, sculpture, photography, operation of computer or any other machine, working with any tools to produce something. It includes all manipulative skills. For any motor activity or skill work, psychological readiness is an

essential condition. If a person is psychologically ready, he or she will also be mentally ready and will act towards the desired, skilled work.

2.5.1 Classification

Psychomotor domain is classified by many psychologists. Some of them are Ragole (1950), E. J. Simpson(1966), J. P. Guilford (1958), R. H. Dave (1969), A. J. Harrow (1972), Allyn & Bacon (1994). In the following paragraphs, we are going to discuss the classification given by R. H. Dave (1969) in an adapted form. There are five categories under this domain, arranged from 1 to 5 in the order of increasing complexity, difficulty level and fineness in the skill being developed.

1. **Initiation-observation or observation-initiation:** For learning any skill (simple or complex), learners need psychological readiness. It is common to most of the learners that they hesitate of any skilled work at the time of beginning. They generally hesitate to take initiative. Contrary to this, there are some learners who are highly motivated to start as they have observed somebody doing it and they are highly motivated to perform that skill.

   If learners are not interested or motivated, they need to become motivated by means of promise of some reward, discussion, fulfillment of some desires and aspirations of the learners, etc. In the above case, 'initiation' is the first step towards the development of skill. On the other hand, if a learner has observed some person performing some skill and is highly motivated and encouraged to do that skill, then he takes initiative automatically. In this case, observation is followed by initiation.

2. **Manipulation:** When the learner is ready to take initiative, he observes others performing that skill and sees how performers are manipulating tools required in that skill and he also starts manipulation of those tools to produce or copy that skill. Manipulation and observation work together continuously for quite some time. As a result, improvement in performance is achieved and it leads towards perfection. During this process, learners perform the following three tasks:

   (i)   Perform selected steps

   (ii) Follow directions

   (iii) Fix their performance through necessary practice

3. **Precision:** Repeated observation of the expert performers and continuous practice leads learners to the performance of the skill with a desired level of precision, i.e., accuracy and exactness. They reach at a higher level of refinement. They achieve this level through the following points of consideration:

   • Controlling faults

   • Eliminating errors

   • Reproducing the desired skill with precision

4. **Articulation:** This is the level at which learners bring some novel attributes or features to their skill performance in addition to the general attributes.

5. **Naturalization:** This is the highest level of performance in skill development. The act of the performer becomes automatic or natural. Achiever of this level of

proficiency, which is rare, performs with the highest degree of refinement and convenience as natural as possible. For performer as well as audience or observer, it looks like an effortless performance.

2.5.2 Assessment of Performance or Psychomotor Learning

Performance assessments are designed to judge a student's abilities to use specific knowledge and research skills. Most performance assessments require the student to manipulate equipment to solve a problem or make an analysis. Rich performance assessments reveal a variety of problem-solving approaches, thus providing insight into a student's level of conceptual knowledge.

Performance assessment is a form of testing that requires students to perform a task rather than select an answer from a ready-made list. Experienced teachers or other trained staff then judge the quality of the student's work based on a pre-decided set of criteria. This new form of assessment is most widely used to directly assess writing ability based on the text produced by students under test instructions.

How does it work?

The student's work is assessed by teachers or other staff on the basis of pre-set criteria. The teachers assess the student's learning abilities when given specific instructions. It is important to understand how the assessment works. The following methods have been used successfully to do so:

- **Open-ended answers:** The student has to come up with the answer independently after some degree of analysis, either orally or in writing. No options are given and the students are usually asked to discuss their observations about some material given to them.

- **Extended tasks:** In this kind of assessment, the student is required to research, draft and revise and come up with a well-though out analysis of the given problem. It is like an assignment or a project. It may take the student many hours to complete.

- **Portfolios:** They are pictorial descriptions of the student's past work and the teacher's evaluation of it. Usually, portfolios include a person's best work and some work in progress which showcases the student's strengths.

What is common among all these methods is that they are all specific tasks completed by the student according to the instructions from the teachers and the students are made well-aware of the standards expected from them. They are also aware that the teachers will be grading them based on their work. This aspect renders performance assessment distinct from other testing methods.

Purpose of assessment

Performance assessments evaluate a student's work and assess how well the student is able to apply the learned concepts. Assessments may be of different types—some may require the student to replicate a procedure learnt in class, some may require simple recall or even intricate analysis.

A limitation of performance assessments is that such assessments may not turn out to be suitable for simply testing the knowledge of facts.

Teaching goals of performance assessment

Performance assessment helps the student to:

- Apply systematic procedures
- Make use of resource texts, laboratory equipment and computers
- Use scientific methodology
- Evaluate and apply several approaches
- Work out difficult problems

The three main purposes of performance assessments are as follows:

1. **Diagnostic purposes:** The assessment is expected to diagnose the students' knowledge about solving specific categories of problems.

2. **Instructional purposes:** A well-crafted assessment can also be used for instructional purposes, except for the fact that it is standardized and graded.

3. **Monitoring purposes:** Since assessments are undertaken at regular intervals, it is very useful for tracking a student's levels of learning, especially in subjects like science and mathematics, which signifies a student's skills of problem solving.

This methodology is revolutionary, in the sense that it is not just about testing, but also about the learning process. In this methodology, students are encouraged to think and create the finished product, during the course of which they are automatically testing their skills and knowledge.

Unlike the old model of teaching where students were tested after each lesson or unit was covered in the classroom, here the approach is to use the book knowledge merely as a resource or means to an end, and not the end itself.

Let us study a relevant example to understand performance assessment better. If a coach assembles 11 players and tells them all the rules of the game of cricket, the spirit in which the game must be played and the dangers of playing without protective padding, will not make them world champions. The players need to go and play an actual game. The performance assessment method involves students 'playing the game'. This way, what goes on in the class is actual preparation for life after school.

Typically, a performance assessment activity can be broken down into the use of the following skills:

1. **Obtaining information:** Students would need to research, compile and organize information as per the specified goal.

2. **Employing the information:** After they have organized the information, they will need to work with it or infer, sequence, analyse and synthesize it to create something new.

3. **Use the information:** When they have the information in the required form, for example, a flowchart, they can use it to convince a target group, or simply present ideas and concepts.

4. **Communication:** Making presentations, oral and visual, doing projects and demonstrating skills is all a part of communicating with others. When a student or

a group of students presents their ideas in any form to others, they are automatically honing their communication skills.

Here are the key benefits of employing the performance assessment methodology:

1. **The variety of tasks call for use of various learning styles and skills:** Each student has a core competency, which simply put, means a skill they have more refined than others. Someone may be good at writing descriptive book reports and someone else may be proficient at making Powerpoint presentations. Some learners are exceptional at organizing information and still others may be brilliant speakers, and so on. Not only does performance assessment allow students to display their particular skills, it also encourages them to acquire new ones. For example, a student who likes writing detailed, informational pamphlets should be encouraged to make a persuasive speech to a group of adults describing the pamphlet he or she has made.

2. **Performance assessment tasks encourage team work:** Working in teams is a skill that will help a student throughout life, especially if he or she chooses to work in a corporate set up. Therefore, it is a useful skill to acquire. Making group presentations, working on projects in a team, and so on, helps students to learn the value of collaboration at an early age. Division of tasks, cooperation, empathy are all valuable lessons that each student must learn.

3. **There is focus on personal habits and attitudes:** In performance assessment tasks, success depends on the learner's attitude almost as much as book learning. Inculcating habits such as flexibility, diligence, cooperation, tolerance, persistence, problem solving, planning and many more are all inherent to the performance assessment methodology.

It is important to have predetermined criteria to evaluate the students' performance. Students should not be scored or graded against their classmates and should be provided with the criteria before the assessment.

Student's feedback reflects levels of competency, rather than comparative scores. It is always useful to try to find in students' performance patterns of appropriate and inappropriate responses.

2.6 CHANGING CONCEPTS OF EDUCATIONAL OBJECTIVES AND RECENT CHANGES IN INSTRUCTIONAL OBJECTIVES

Bloom's taxonomy of educational objectives was revised in 2001 under the title of *A Taxonomy for Learning, Teaching and Assessing*. The revision work was done by authors Lorin. W. Anderson, David R. Krathwohl, Peter W. namely, Airasian, Kathleen A. Cruikshank and others. Due to the changes brought in the cognitive domain, the instructional objectives for this domain also need changes accordingly.

The following changes were made in the classification of educational objectives:

The nomenclatures of the categories under cognitive domain were changed.

• Knowledge was replaced by remembering.

- Comprehension was replaced by understanding.
- Application was replaced by applying.
- Analysis was replaced by analysing.
- Synthesis was replaced by creating.
- Evaluation was replaced by evaluating.

Rationale behind the changing concepts

It is evident that the six categories in the original classification of Bloom are in the form of 'noun', which has been changed to 'verb'. The rationale behind this change is simple, logical and easy to digest. Taxonomy of educational objectives, i.e., various categories reflect different forms of thinking. Thinking, being an active process, represents the forms of verbs more accurately than the noun. Therefore, some of the sub categories were reorganized and renamed. Knowledge, the first category, is a product of thinking and was considered inappropriate to describe a category of thinking and, therefore, it was replaced with the word 'remembering', the lowest level of cognitive domain activities.

Knowledge is a product of thinking (lower as well as higher order products) and was not clarifying its level of category; hence change was needed. Remember is the right word because it indicates the product of lower order thinking. We can cram, recall, recognize and remember something without understanding it. Comprehension is replaced by understanding, synthesis is replaced by creating, and so on. All these changes reflect the nature of the thinking process and the outcomes described by each category in a better way.

2.6.1 Functions of Educational Objectives/Usefulness of Taxonomical Classification

Educational objectives perform the following functions:

- Bloom's taxonomy motivates educators to focus on all three domains of educational objectives.
- Creates a more holistic form of education to develop the whole personality of the learners.
- Whole personality development makes an individual more useful for his ownself and for the society.
- Allows teachers to select appropriate classroom assessment techniques for student evaluation and for the performance or skills to be evaluated.
- Understanding Bloom's taxonomy develops the competencies in the teachers through which a teacher can design effective learning experiences.
- Helps in devising and organizing fruitful and interesting co-curricular activities.
- Takes care of the all-round development of the personality of the child.
- Helps teachers in framing instructional objectives in behavioural terms.

2.6.2 Guidance Functions of Evaluation

Guidance and counselling of students have become one of the most important areas of school education. Due to diverse kinds of problems like stress and suicide committed by

students, this area has become very important for students, parents and educational administration. Because of these reasons, counsellors are being appointed in schools and several guidance and counselling clinics and centres are coming up at different places to support students and parents. Some of the important guidance functions of evaluation are listed as follows:

- Evaluation results are very much useful for the guidance and counselling of the needy students.

- Proper educational and vocational guidance can be given to students on the basis of their evaluation on various abilities, interests, skills, attitudes, aptitude, values, and after knowing the strength and weakness of the needy students.

- On the basis of evaluation results of the test of anxiety and stress among students, proper guidance may be given to students to reduce their level of stress.

- Various types of personality disorders may be diagnosed by using variety of tests and the subject may be given proper guidance to cure or control them properly.

- Students may be helped in choosing the right kind of vocation or profession or career in which they may succeed and grow.

- Students may be helped in choosing suitable courses of study which suit them and in which they may excel.

- Based on the findings of evaluation, students may be guided and counselled properly to nurture and develop their life skills.

2.7 SUMMARY

In this unit, you have learnt that:

- The study of classification is known as taxonomy.

- Bloom developed and proposed the classification of educational objectives in 1956 which is popularly known as Bloom's Taxonomy of Educational Objectives.

- Bloom classified educational objectives into three 'domains' which are as follows:

  o Cognitive domain

  o Affective domain

  o Psychomotor domain

- Cognitive domain includes those objectives of education which attempt to develop mental faculties or intellectual abilities, that is, the ability of knowing, understanding, thinking and problem solving.

- Bloom has further classified cognitive domain into six major categories which are knowledge, understanding, application, analysis, synthesis and evaluation.

- Affective domain of Bloom's taxonomy of educational objectives is related to the development of emotions, values, attitudes and the development of those aspects of personality which are more influenced by heart than the mind.

- Affective domain has been further classified by D. R. Krathwohl, B. S. Bloom and B. B. Masia into five categories which are receiving, responding, valuing, organization, and characterization of values or value complex.

- Psychomotor domain deals with the development of motor abilities which in turn are responsible for skill development for performing any task.

- Psychomotor domain has been further classified by R. H. Dave into five categories which are initiation-observation, manipulation, precision, articulation and naturalization.

- Instructional objectives are desired behavioural changes in the form of statements formulated by the teacher for describing the observable outcomes of learning at the completion of classroom instructions.

- The following critical thinking skills are required in cognitive learning: convergent thinking, divergent thinking, critical thinking, problem solving and decision making.

- Bringing facts and data together from various sources and then applying logic and knowledge to solve the problem to achieve an objective is known as convergent thinking. Convergent thinking can be assessed by administering standard IQ tests, by various recognition or knowledge tests, logical discussions, etc.

- Divergent thinking is thinking outwardly instead of inwardly. It is the ability to develop original and unique ideas and then conic up with a problem solution or achieve an objective. Divergent thinking typically occurs in a spontaneous, free-flowing maimer, such that the ideas arc generated in a random, unorganized fashion.

- The following are some of the techniques used to stimulate divergent thinking: brainstorming, recording ideas, free writing, and mind or subject mapping.

- Critical thinking comprises various modes of thinking, such as scientific thinking, mathematical thinking, historical thinking, anthropological thinking, economic thinking, moral thinking and philosophical thinking.

- Problem solving is a mental process that involves discovering, analysing and solving problems. The ultimate goal of problem solving is to overcome obstacles and find a solution that best resolves the issue.

- Decision making can be regarded as the mental processes resulting in the selection of a course of action among several alternative scenarios. The end result of each decision-making process is a final selection.

- There are four stages those are involved in all group decision making: orientation stage, conflict stage, emergence stage and reinforcement stage.

- The assessments used in affective assessment are attitudes and values, interest, and self-concept.

- An attitude is a state of mind or a feeling or disposition. It is important to have a positive attitude about work.

- Various methods of measuring attitudes have been developed. Attitude measurement can be divided into two basic categories: direct measurement (A liken scale) and indirect measurement (projective techniques).

- Direct measurement involves recording the actual behaviour of people whose attitude is to be studied or asking them direct questions about their feelings. A Liked scale is a bipolar scale running from one extreme through a neutral point to the opposite extreme.

- Another tool for affective learning is assessing the interest of the student. This inventory helps the teachers in knowing the interest areas of their students so that they can encourage and get various opportunities to grow in a particular field.

- The self-concept is the way in which we think about and evaluate ourselves. The development of self-concept has two aspects: the existential self and the categorical self

- Performance assessments are designed to judge a student's abilities to use specific knowledge and research skills. Performance assessment is a form of testing that requires students to perform a task rather than select an answer from a ready-made list.

2.8 KEY TERMS

- **Hierarchical:** It means learning at the higher level in any domain is entirely dependent on having acquired prerequisite knowledge and skills at lower level.

- **Cognitive domain:** It includes those objectives of education which attempt to develop mental faculties or intellectual abilities, that is, the ability of knowing, understanding, thinking and problem solving.

- **Affective domain:** It is related to the development of emotions, values, attitudes and the development of those aspects of personality which are more influenced by heart than the mind.

- **Psychomotor domain:** It deals with the development of motor abilities which in turn are responsible for skill development for performing any task.

- **Events of instruction:** They constitute a set of communications of the teacher to the students which have the aim of aiding the learning process.

- **Curriculum:** It is a course of study; the linked series of academic courses leading to mastery of a discipline.

- **Likert scale:** It is a tool used to determine opinions or attitudes; it contains a list of declarative statements, each followed by a scale on which the subject is to indicate degrees of intensity of a given feeling.

- **Self-concept:** It is the composite of ideas, feelings and attitudes that a person has about his or her own identity, worth, capabilities and limitations.

2.9 ANSWERS TO 'CHECK YOUR PROGRESS'

1. American educational psychologist Benjamin Bloom developed and proposed the classification of educational objectives in 1956, which is popularly known as Bloom's Taxonomy of Educational Objectives.

2. The guidance functions of evaluation are as follows:

   (i) Evaluation results are very much useful for the guidance and counselling of the needy students.

   (ii) On the basis of evaluation results of the test of anxiety and stress among students, proper guidance may be given to students to reduce their level of stress.

3. The various categories of cognitive domain are knowledge, understanding, application, analysis, synthesis and evaluation.

4. Convergent thinking means the ability to give the 'correct' answer to standard questions that do not require the utilization of much creative faculty.

5. Problem solving is a mental process that involves discovering, analysing and identifying solutions to problems.

6. The first two steps in decision making are as follows:

   (a) Objectives are to be established first.

   (b) Objectives must be classified and placed in the order of importance.

7. The four decision-making stages are orientation stage, reinforcement stage, emergence stage and conflict stage.

8. Affective domain is related to the development of emotions, values, attitudes and the development of those aspects of personality which are more influenced by the heart than the mind. Et also includes the development of interests, appreciation, feelings, likes and dislikes towards something.

9. Zimbardo defined attitude as, 'An attitude can be defined as a positive or negative evaluation of people, objects, events, activities, ideas, or just about anything in your environment'.

10. The two ways in which attitude can be measured are by either using direct measurement or indirect measurement.

11. The two scales that can be used to measure attitude are Thurstone scale and Likert scale.

12. It is important to measure the interest of the students in order to help them achieve their dream careers.

13. E. K. Strong devised a way to measure interest to help people who were quitting their military jobs to get new jobs.

14. Self-concept can be defined as the way in which we think about and evaluate ourselves.

15. Performance assessments are designed to judge a student's abilities to use specific knowledge and research skills.

16. Portfolios describe a student's variety of performance-based work. A portfolio might include a student's 'best work' and the student's evaluation of the strengths and weaknesses of several tasks.

17. Bloom's taxonomy of educational objectives was revised in 2001 under the title of A Taxonomy for Learning, Teaching and Assessing.

18. Course or programme evaluation is a systematic and organized process of collecting and analysing data on some specific course or programme being run in any educational institution including schools.

2.10 QUESTIONS AND EXERCISES

Short-Answer Questions

1. Describe 'knowledge' as the first level of cognitive domain.
2. Describe the reasons behind the changing concepts educational objectives.
3. Define instruction'.
4. What is programme evaluation?
5. What are the factors one has to keep in mind in order to be an affective learner?
6. How is the ability of critical thinking significant?
7. State few ways which can ensure that one makes good decisions.
8. What are the two aspects of the development of self-concept?
9. What are the purposes of performance assessment?

Long-Answer Questions

1. Discuss the classification of cognitive domain under educational objectives.
2. Discuss the classification of affective domain under educational objectives.
3. Explain the classification of psychomotor domain under educational objectives.
4. State the functions of educational objectives.
5. Explain the three dimensions of learning.
6. Describe the techniques that help to stimulate divergent thinking.
7. What are the characteristics of critical thinking?
8. List and describe the steps involved in problem solving.
9. What do you understand by the term self-concept? How can it be used in the process of assessment?

2.11 FURTHER READING

Annastasi, A. 1976. Psychological Testing, 4th Ed. New York: McMillan Publishing Co.

Block, J. I-I, L. W. Anderson. 1975. Mastery Learning in Classroom Instruction. New York: McMillan Camp.

Bloom, B. S. 1956. Taxonomy of Objectives of Cognitive Domains. New York: Longmans Green &Co.

Bloom, B. S. 1956. Taxonomy of Objectives (Affective Doming). New York: Longmans Green & Co.

Chronbach, L. J. 1970. Essentials of Psychological Testing, 3rd Ed. New York: Harper.

Catel, R. B. Personality: A Systematic Theoretical and Factual Study. New York: McGraw Hill.

Ebel, R.L. 1972. Essentials of Educational Measurement. New Jersey: Englewood Cliff.

Edwards, A.L. 1957. Techniques of Attitude Scale Construction. New York: Application Century.

Frank S. Freeman. Theory & Practice of Psychological Testing. Mumbai: IBII Publishing Co.

Grunlund, N. E. 1970. Stating Behavioural Objectives for Classroom Instruction. New York: McMillan.

Grunlund, N. E. 1973. Preparing Criterion Referenced Tests for Classroom Instruction. New York: McMillan.

Grunlund, N. E. 1976. Measurement and Evaluation in Teaching (3rd Ed). New York: McMillan Publishing Co.

Goslin, D. A. 1967. Teachers and Testing. New York: Sage Foundation.

Grunlford, J. P. 1967. The Nature of Human Intelligence. New York: McGraw Hill.

Grunlund, N. E. 1959. Sociometry in the Classroom. New York: Harpt & Row.

Julian, C., Stanley and Hopkins. 1972. Educational and Psychological Measurement and Evaluation. New Delhi: Prentice Hall of India Pvt. Ltd.

Lindermon, Richard, H. Educational Measurement. Mumbai: B. S. Taraporevala Sons & Co.

Lynum, H. B. 1971. Test Scores and What They Mean. New Jersey: Prentice Hall.

Ten Brink, T. D. 1974. Evaluation: A Practical Guide for Teachers. New York: McGraw Hill.

Bernard, H. W., D.W Fullmer. Principles of Guidance: A Basic Text (Indian Edn.). New Delhi: Allied Pub.

Crow Lester, D., A. Crow. An Introduction to Guidance Principles and Practices. New York: American Books Co.

Downing, Lester, N. 1964. Guidance and Counselling Services. New York: McGraw Hill.

Fusterm J. N. Psychological Counselling in India. Mumbai: McMillan.

Kochhar, S. K. 1987. Educational Vocational Guidance in Secondary Schools. New Delhi: Sterling Publisher Pvt. Ltd.

Hoose, William, N. Van. (ed). 1979. Counselling and Guidance in 20th Century. Boston: Noughton Muffin Co.

Khorshed, A. W. Guidance Movement in India, Guidance Services. New Delhi: NCERT.

Pandey, K. P. Educational and Vocational Guidance in India. Varanasi: Vishwavidyalaya Prakashan.

Pasricha, Pren. 1976. Guidance and Counselling in Indian Education. New Delhi: NCERT.

Swamy, R. V. (ed). 1971. Guidance Services in Colleges and Universities. Bangalore:

University and Directorate of Employment and Training.

Traxler, R. B., R. D. Norla. Techniques of Guidance. New York: Harper and Road.

UNIT 3 CHARACTERISTICS OF A GOOD TEST

Structure

3.0 INTRODUCTION

A 'test' is a particular type of assessment that typically consists of a set of questions to be answered and administered under reasonably comparable conditions for all students. It is an advice or a systematic procedure for assessing a sample of behaviour by asking a set of questions in a uniform manner. Because a test is a form of assessment, tests also answer the question, 'how well does the individual perform either in comparison with others or in comparison with a performance tasks?' Psychological and educational tests are a standardized procedure to measure quantitatively or qualitatively one or more than one aspect or trait by means of a sample of verbal or non-verbal behaviour. A test

consists of a standard set of questions to be answered or tasks to be performed. Items of a test are placed in increasing order of difficulty and its procedure of administration is standardized to ensure maximum objectivity. In the teaching-learning situation, psychological and educational tests are used to know about the level of achievement of the students.

The scientific phase of testing and measuring starts with the work of American psychologist E. L. Thorndike. Thorndike constructed and published a number of pioneer psychological tests. Ink-Blot test developed by psychologist Hermann Rorschach, intelligence test developed by French psychologist Alfred Binet and Theodore Simon, intelligence test on adults developed by American psychologist David Wechsler, attitude scale developed by Bhatia, etc., are the well-known examples of psychological tests. Psychological tests nowadays have become a very popular feature in nearly all fields of contemporary education and guidance.

This unit covers the significance of the reliability and validity in qualifying a test as good. It also examines the methods of computing reliability, factors affecting reliability and validity, types of validity and factors affecting validity.

3.1 UNIT OBJECTIVES

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

- Explain the concepts of validity and reliability
- Define reliability
- Identify the methods of computing reliability
- List the factors that affect reliability
- Identify the types of validity
- List the factors that affect validity
- Describe the relationship between validity and reliability

3.2 RELIABILITY AND VALIDITY

An educational test is not just a test that measures achievement in subjects of study, but is also a psychological test that leads to the over-all development of a student. According to American psychologist Anne Anastasi, 'Psychological test is essentially an objective and standardized measure of a sample of behaviours.' For American psychologist F. S. Freeman, it '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'.

A test is a stimulus selected and organized to elicit responses which can reveal certain psychological traits in the person who deals with them. The diagnostic or predictive value of psychological test depends upon the degree to which it serves as an indicator of a relatively broad and significant area of response. It is obvious that a psychological test is the quantitative and qualitative measurement of the various aspects of behaviour of the individual for making generalized statements about his total performances.

The aspects which affect the characteristics of a good test are as follows:

- Validity of the test

- Reliability of the test

- Objectivity of the test

- Usability of the test

- Comprehensive and preciseness of the test

- Administration of the test

- Test from economic viewpoint

- Availability of the test

- Appearance of the test

- Standardization of the test

- Norms of the test

Some of the important characteristics of a test are analysed below.

Validity: Validity of a test refers to its truthfulness; it refers to the extent to which a test measures what it intends to measure. Standardization of a test requires the important characteristic viz., validity. If the objectives of a test are fulfilled, we can say that the test is a valid one. Validity of a test measures the truthfulness of a test. The validity of a test is determined by measuring the extent to which it matches with a given criterion. Let us take an example. Suppose we want to know whether an 'achievement test in mathematics' is valid. If it really measures the achievement of students in mathematics, the test is said to be valid, or else not. So 'validity' refers to the very important purpose of a test, and hence, it is the most important characteristic of a good test. A test may have other merits, but if it lacks validity, it is valueless.

Freeman states, 'An index of validity shows the degree to which a test measures what is purports to measure when compared with the accepted criteria'. American educational psychologist Lee J. Cronbaeh held the view that validity 'is the extent to which a test measures what it purports to measure'.

Reliability: Reliability refers to consistency of scores obtained by some individuals when re-tested with the test on different sets of equivalent items or under other variable examining conditions. It refers to the consistency of scores obtained by the same persons when they are re-examined with the same test on different occasions or with different sets of equivalent items or under different examining conditions. Reliability paves way for consistency that makes validity possible and identifies the degree to which various kinds of generalizations are justifiable. It refers to the consistency of measurement, i.e., how stable test scores or other assessment results are from one measurement to another.

Reliability refers to the extent to which a measuring device yields consistent results upon testing and retesting. If a measuring device measures consistently, it is reliable. The reliability of a test refers to the degree to which the test result obtained is free from error of measurement or chance errors. For instance, we administer an achievement test in mathematics for students of class IX. In this test, Paresh scores 52. After a few days, we administer the same test. If Paresh scores 52 marks again, we consider the test to be reliable, because we feel that this test accurately measures Paresh's ability in

mathematics. American psychologist and segregationist H. E. Garrett stated, 'the reliability of test or any measuring instrument depends upon the consistency with which it gauges the ability to whom it is applied'. The reliability of a test can also be defined as 'the correlation between two or more sets of scores on equivalent tests from the same group of individuals'.

Objectivity: Objectivity is an important characteristic of a good test. Without objectivity, the reliability and validity of a test is a matter of question. It is a prerequisite for both validity and reliability. Objectivity of a test indicates two things: item objectivity and scoring objectivity.

Item objectivity refers to the item that must call for a definite single answer. In an objective-type question, a definite answer is expected from the test-takers. While framing the questions, some points to be kept in mind are: ambiguous questions, lack of proper direction, double barreled questions, questions with double negatives, etc. These concepts affect the objectivity of a test. Let us take an example of an objective item. Suppose we ask students to write about Gandhi. This question does not have objectivity, because here the answers will have different perceptions for different individuals and also the evaluation. If we ask the students 'what was Gandhi's father's name', this obviously will have only one answer and even the biasness of the evaluator will not affect the scoring. So, all the items of a test should be objective.

Objectivity of scoring means whosoever checks the test paper, it would fetch the same score. It implies that the subjectivity or personal judgment or biasness of the scorer should not affect the scores. The essay-type questions are subjective and the scores are affected by a number of factors like mood of the examiner, his language, his biasness, etc. Essay-type questions can have objectivity if scoring key and proper direction for scoring are provided.

Usability: Usability of a test refers to the practicability of a test. It refers to the degree to which the test can be successfully used by the teachers/evaluators. Usability of a test depends on certain aspects which are expressed in the following manner:

(a) *Comprehensibility:* The test items should be free from ambiguity and the direction to the test items and other directions to the test must be clear and understandable. The directions for scoring and the interpretation of scores must be within the comprehension of the user.

(b) *Ease of administration:* If the directions for administration are complicated or if they need more time and labour, the users may lag behind to use such tests. The directions for administration must be clear and concise. The test paper should be constructed according to the availability of time. Lengthy tests involving more time may not be preferred for use.

(c) *Availability:* If a test is not available at the time of necessity, it lacks its usability. Most of the standardized tests are of high validity and reliability, but their availability is very less. So it is desirable that in order to be reliable, the tests must be readily and easily available.

(d) *Cost of the test:* The cost of the test must be cheap, so that the schools and teachers can afford to purchase and use them. If it is costly, then every school cannot avail it. So a good test should be of reasonable price.

(e) *Ease of interpretation:* A test is considered to be good if the test scores obtained can be easily interpreted. For this, the test manual should provide age norms, grade norms, percentile norms and standard score norms like standard scores, T-scores, Z-scores, etc. So 'interpretability' of test refers to how readily the raw scores of test can be derived and understood.

(f) *Ease of scoring:* A test in order to be usable must ensure ease of scoring. The scoring procedure must be a simple one.

All the directions for scoring and the scoring key should be available, to make the scoring an objective one. The examiner's biasness or the handwriting of the examinee should not affect the scoring of a test.

Evaluation is the part and parcel of teaching learning situation. How far the instructional objectives have been achieved, how far the curriculum and methods of teaching is suitable for the students, and everything can be known through the process of evaluation. Test is a tool for evaluation. In the classroom situations, the teachers develop tests to measure and evaluate students' achievement. Different subject teachers construct different tests for their students, Teacher-made tests serve the purposes, viz., measure the pupils' achievement, know how far the specific objectives have been fulfilled, diagnose the learning difficulties, arrange specific remedial measures, award grades, etc. Teacher-made tests are based on content and objective specific to teacher's own class or school. They may deal either with a specific topic or skill or with larger segments of knowledge and skill. Generally, the teacher-made tests are less reliable in comparison to the standardized test. They have a limited use, to a particular class or a single school. Teacher-made tests are well adapted to outcomes and content of local curriculum.

Flexibility affords continuous adaptation of measurement to new materials and changes in procedure. It too often neglects complex learning outcomes.

The quality of items of the teacher-made tests is lower than standardized tests because of a teacher's limited time and lack of opportunity to pretest items. The score comparisons and interpretations of teacher-made tests are limited to local school situation. Teacher-made tests evaluate student's day-to-day progress and their achievement on work units of varying size.

Standardized tests follow four steps for its construction, but teacher-made tests follow only two steps for its construction. Such tests do not need rigorous process of standardization. They are not out, not subjected to rigorous item analysis, and not evaluated for their validity and reliability. Simply, teacher-made tests go through two stages, viz., 'planning' of the test and 'preparation' of the test.

Mostly, teacher-made tests are used for formative evaluation. A teacher-made test may be a full-fledged achievement test covering the whole course of a particular subject or a unit test covering one unit of study. The results of teacher-made tests are useful for action researches. The teacher-made tests are evaluated by the same teacher, who makes the test papers. These tests if skillfully prepared can even serve the purpose of standardized tests. Table 3.1 differentiates between teacher-made tests and standardized tests.

Table 3.1 Comparison between Teacher-Made Tests and Standardized Tests

Teacher-made tests	Standardized tests
Teacher-made tests are prepared by the classroom subject teachers.	Standardized tests are prepared by the subject experts.
Teacher-made tests follow two steps: planning and preparation.	Standardized tests follow four steps: planning, preparation, try-out and evaluation.
Teacher-made tests are useful for the concerned classroom only.	Standardized tests are useful all over the world.
Teacher-made tests are useful for action research.	Standardized tests are useful fur basic or fundamental research.
The items are not arranged according to their difficulty level	The items are arranged according to their difficulty level.
The scores of the teacher-made tests have limited comparison value and their interpretations are limited to local school situation.	The scores of standardized tests can be compared with those of norm groups.
The reliability of teacher-made tests are usually unknown, but can be high if carefully constructed.	The reliability of standardized tests is high, commonly between 80 and 95, frequently around 90.
The quality of items is unknown unless test item file is used.	The quality of items is very high.
Scoring key is not used.	Scoring key is very necessary and hence used.
In case of teacher-made tests, there may or may not be objective type tests.	In case of standardized test usually objective type tests are used.
Take less time for construction and experts may not be involved in it.	Tests involve so many experts, use so many sophisticated procedures and thus take along time for construction
Based on content and objectives specific to teacher's own class or school.	Based on content and objectives common to many schools throughout the country.

3.3 MEANING OF RELIABILITY

Reliability paves way for consistency that makes validity possible and identifies the degree to which various kinds of generalizations are justifiable. It refers to the consistency of measurement, which is how stable test scores or other assessment results are from one measurement to another. Reliability refers to the extent to which a measuring device yields consistent results upon testing and retesting. If a measuring device measures consistently, it is reliable. The reliability of a test refers to the degree to which the test results obtained are free from error of measurement or chance errors.

3.3.1 Characteristics

The characteristics of reliability are as follows:

- It refers to the degree to which a measuring tool yields consistent results upon testing and retesting.
- It indicates the level to which a test is internally consistent, i.e., how accurately the test is measuring.
- It refers to the results obtained with measuring instrument and not to the instrument itself.
- An estimate of reliability refers to a particular type of stability with the test result.
- Reliability is necessary but not a sufficient condition for validity.
- Reliability is a statistical concept.
- It refers to the preciseness of a measuring instrument.
- It is the coefficient of internal consistency and stability
- It is the function of the length of a test.

3.4 METHODS OF COMPUTING RELIABILITY

When examining the reliability coefficient of standardized tests, it is important to consider the methods used to obtain the reliability estimates. American Psychological Association (APA) introduced several methods of estimating reliability. The methods are similar in that all of them involve correlating two sets of scores, obtained either from the same assessment procedure or from equivalent forms of the same procedure. The chief methods of estimating reliability are shown here. The reliability coefficient resulting from each method must be interpreted according to the type of consistency being investigated.

We will consider each of these methods of estimating reliability in detail in Table 3.2.

Table 3.2 *Methods of Estimating Reliability*

Method administration	Types of reliability measure	Procedure of
Test-Retest method the same group between tests, from years.	Measures of stability and precision	Use the same test twice to with anytime interval several minutes to several
Equivalent forms method Apply two equal forms of the test to the same		Measure of equivalence group in close time gap.
Split-Half method two equivalent and even items), halves to fit whole	Measure of internal consistency	Apply the test once. Score halves of test (eg , odd items correct correlation between test by Spearman—Brown

formula to		measure reliability
of the test.		
Kuder-Richardson arid apply Kuder	Measure of Internal consistency	Give test once, score total
method know the degree		—Richardson formula to of reliability.
Inter-rater method response requiring to two or more raters and independently score the responses	Measure of consistency	Use a set of student judgmental scoring have them

(i) Test-retest method

To estimate reliability by means of the test-retest method, the same assessment is administered twice on the same group of pupils with a given time interval between the two administrations. The resulting assessment scores are correlated and this correlation coefficient provides a measure of stability, i.e., it indicates how stable the assessment results are over a given period of time. If the result shows that the students who were at good position at the first administration, and also they are at good position in the second administration, the test has stability. The closer the agreement between the two administrations of the test, the greater is the reliability co-efficient or coefficient of stability.

A high test-retest reliability or coefficient of stability shows that there is low variable error in the sets of obtained scores. A low coefficient of stability indicates that there is high variable error in the obtained scores. The error variance contributes inversely to the coefficient of stability. Such stability is indicated by a large correlation coefficient, that a perfect positive relationship is indicated by 1.00, perfect negative correlation by — 1 and no correlationship by 0.00. Measures of stability in the 0.80 range are generally reported for standardized tests. The Pearson's method of correlation is mostly used for obtaining the coefficient of reliability.

When we are dealing with the measures of stability, it is more important to keep in mind about the time gap between assessments. If the time gap is short, say a day or two, the consistency of the results will be inflated because a student remembers the tasks and responses accordingly. If the time interval is a long one, the results will be influenced not only by the instability of the assessment procedure but also by actual changes in the student over that period of time. The best time interval between assessment administrations depends largely on the use to be made of the results.

Limitations of test—retest method

The limitations of test-retest method are as follows:

- Testing conditions during test and retest may vary which result in instability in the scores.
- The individual's health, emotional conditions, mental health, motivational conditions and mental set do not remain the same in both the administrations at two occasions.

- If the test is repeated immediately, the test-taker may recall the first answer. This may increase the scores, Besides the memory effects, practice and confidence induced by familiarity with the test will almost affect the scores the second time. This is known as 'the carry-over effect' or 'transfer effect' or 'memory effect' or 'practice effect'.

- If the time gap between two administrations is a lengthy one, additional learning or changes in the characteristics of the individual will affect the scores at the later administration, and it may decline the reliability of a test.

(ii) Equivalent forms method

This method of reliability is also known as parallel forms reliability/equivalent forms reliability/comparable forms reliability. Two parallel forms of a test can be constructed by selecting the samples in each form from the same universe of contents. By parallel forms, we mean that the forms of tests are equivalent as far as the content, objectives, format, difficulty level and discriminating value of items, length of the test, etc., are concerned. Parallel tests have equal mean scores, variance and equal inter-correlation among items. So the two parallel forms of a test must be similar in all respects, but there should not be a duplication of test items. The equivalent forms method of establishing reliability is widely used in standardized testing because for most standardized tests, two or more forms are available when the two forms are virtually alike, reliability is too high; when they are not sufficiently alike, reliability will be too low. Pearson's method of correlation is used for calculating the coefficient of correlation between two sets of scores obtained by administering the two forms of the test. This coefficient of correlation is termed as the 'coefficient of equivalence'.

Limitations of equivalent forms method

The limitations of equivalent forms method are as follows:

- Construction of equivalent test forms is too difficult in comparison to others.

- There is a chance of memory effect or practice effect to operate at the administration of the second form.

- The testing conditions while administering the two forms may be different and the test-takers may not be in a same physical, mental or emotional state at both the time of administration.

(iii) Split-half method

Reliability can also be calculated from a single administration of a single test. The test is administered to a group of pupils and then is divided into half for scoring. To divide the test into halves that are equivalent, the procedure is to score the even-numbered and the odd-numbered items separately. This results two scores for every student which when correlated, gives a measure of internal consistency. This coefficient indicates the degree to which consistent results are obtained from the two halves of this test and may be thought of as the half-length test reliability estimate.

Among the different methods of dividing the test, the odd and even split is used mostly. Here, items no.1, 3, 5, 7, 9, 11, etc., (the odd-numbered items) constitute one half, and items no. 2,4, 6, 8, 10, etc., (the even-numbered items) constitute the other half of the test. Scores of individuals on these two halves are correlated to know the degree

of reliability between the two. The coefficient of correlation is usually denoted by the symbol ($_{rhh}$). The reliability coefficient of the whole test (r_{xx}) is estimated from the value of rhh using the following:

Spearman-Brown formula:

Where,

r_{xx} = Reliability of whole test

r_{hh} = Reliability of half test

Limitations of split-half method

The limitations of split-half method are as follows:

- The whole test can be divided into two parts in a number of ways. As such, reliability coefficient obtained through this method may vary from method to method.

- As the test is once administered, the factor chance error may affect the scores on the two halves, and thus, tend to make the reliability coefficient too high.

- It is very difficult to divide the test into two halves in such, way that both halves are of equivalent difficulty levels and discriminating power.

- This method cannot be used in power test.

(iv) Kuder-Richardson method

This method of reliability was developed by G. E Kuder and M. N. Richardson. This is also known as rational equivalence method of reliability. It is most useful for homogenous test reliability. Like the split-half method, the Kuder—Richardson method provides an internal consistency but it does not require splitting the assessment in half for scoring purposes. This method enables us to calculate the inter-correlation of the items of the test and the correlations of the items with the test as a whole. All the items in the test measure the same ability—the correlation between the items are equal, that all the items are of same difficulty and that all the items are highly homogeneous. When we use Kuder—Richardson formula, it is required that all the items of the test should be psychologically homogeneous and that every item in the test has a high correlation with every other item. This is referred to as inter-item consistency. This coefficient is called the 'coefficient of rational equivalence'. This coefficient provides some signals of how internally consistent or homogeneous the items of the test are.

Kuder-Richardson formula used to compute the internal consistency coefficient of test items is:

Where,

r_{tt} = Reliability index

n = Number of items in the test

p = Proportion of right responses

q = Proportion of wrong responses

Example: A test consisting of 50 items and standard deviation of test score is 7.5 and the sum of the product of proportion of right and wrong responses on the item is 10.43. Calculate the reliability.

Solution:

The reliability coefficient of the test is 0.84.

The KR-21 method is also a simple one for calculating reliability of a test. The test is administered once on a group to determine quickly the reliability. The mean and variance are calculated and then the following formula KR-2 1 is used.

Where,

 = Reliability of the whole test

 n = Number of items in the test

M = Mean of the test scores

x = Standard deviation

Example: An objective test of 100 multiple items have been administered to a small group of students. The mean of test score is 50 and standard deviation is 10. Calculate the reliability coefficient of the test.

Solution:

= 0.76

The reliability coefficient of the test is 0.76.

Limitations of Kuder—Richardson method

The limitations of Kuder—Richardson method are as follows:

- Kuder—Richardson formulae arc not suitable for speed assessments—assessments with time limits that prevent students from attempting all the items.

- The formulae indicate the consistency of student response from one day to the other.

- It cannot be used for power test and heterogeneous tests.

- The different Kuder—Richardson formula results differ in reliability coefficient.

- In case all the items of the tests are not highly homogeneous, this method will produce lower reliability coefficient.

(v) Inter-rater method

This method assesses reliability through scoring/evaluating, done by two or more independent judges for every test. The various scores given by the judges are then compared to determine the consistency of estimations. The way the comparison is carried out is: each rater assigns each test item a score, which would be on a scale from 1 to 1 0. Then the correlations between any two ratings are calculated. There is another method for testing inter-rater reliability. In this method, raters identify a category for each observation and then compute the percentage of agreement among the raters. For instance, if the raters are in agreement 7 times out of 10, the test will be said to have a 70 per cent inter-rater reliability rate.

In case the raters seem to be in disagreement, it would imply that either the raters need to be trained again or the scale is defective. Sometimes, it so happens that various raters would have different opinions about measurement results emerging from the same object, such as a scientific experiment or a test, wherein first the test is carried out, then its results are interpreted, recorded and presented. At any of these stages, the rater may become affected by rater's bias (the tendency to rate in the direction of what the rater expects). There may also be discrepancy during interpretation and presentation of results, for instance, the round-off may be different in terms of higher or lower digit next to the decimal.

Limitations of inter-ratermethod

The limitations of inter-rater method are as follows:

- The method can be tedious because inter-rater reliability statistics need to be calculated separately for every item and every pair of raters.

- It is a lengthy and difficult task to train the raters such that they are able to reach an exact agreement.

- Even when they are trained, the forced consensus might render the ratings inaccurate and this would be a threat to the validity of the student's scores.

- The resulting estimates might turn out to be too conservative if two raters show differences in the method used on the rating scale.

3.5 FACTORS AFFECTING RELIABILITY

The reliability of a test is affected by a couple of factors which are explained in the following manner:

 (i) **Length of the test:** There is positive correlation between the number of items in a test and the reliability of a test. The more the number of items the test contains, the greater is its reliability. In several tests, the scores of sub-tests and whole tests are calculated separately and their reliability is also calculated separately. The reliability of the whole test is always more than the sub-test, because whole test means more items, which is better representation of the content.

 (ii) **Construction of the test:** The nature of items, their difficulty level, objectivity of scoring, item interdependence and alternative responses are factors which

affect the reliability. More alternative responses will increase the reliability of the test.

(iii) **Nature of the group:** Reliability of a test will be more if the test is administered to a heterogeneous group. The more the variability, the higher the reliability coefficient.

(iv) **Testing conditions:** If the testing conditions are not similar at all the places, then differences in scores are obtained. The physical conditions of the tests and the environmental factors around the test-taker affect the reliability of a test.

(v) **Guessing and chance errors:** Guessing paves the way to increase error variances and it reduces reliability. If there are more opportunities for guessing in the test, the test will yield less reliable results.

(vi) **Test instructions:** If instructions in the test are complicated or difficult to understand, there will be less consistency in the scores. If the test-taker will not understand the instruction properly, his way of response will be wrong and this will hamper the reliability of test.

(vii)   **Too easy or too difficult items:** Too easy or too difficult items fail to distinguish between good and bad students which otherwise affects the reliability of a test.

The other factors which affect the reliability of tests are: subjectivity of the examiner, clerical error, interval between testing, effect of practice, etc.

3.6 VALIDITY: MEANING AND TYPES

The validity of a test is determined by measuring the extent to which it matches with a given criterion. It refers to the very important purpose of a test, and it is the most important characteristic of a good test. A test may have other merits, but if it lacks validity, it is valueless.

3.6.1 Characteristics of Validity

The characteristics of validity are as follows:

- Validity is a unitary concept.

- It refers to the truthfulness of the test result.

- In the field of education and psychology, no test is perfectly valid because mental measurement is not absolute but relative.

- If a test is valid, it is reliable; but if a test is reliable, it may or may not be valid.

- It is an evaluative judgment on a test. It measures the degree to which a test measures what it intends to measure.

- It refers to the appropriateness of the interpretation of the result, and not to the procedure itself.

- It refers to degree means high validity, moderate validity and low validity.

- No assessment is valid for all the purpose. A test is valid for a particular purpose only

3.6.2 Types

Validity is a specific characteristic of a test. There are different ways of calculating the validity of a test based on the objectives with which we validate the test. The six types of validity are discussed below.

(i) **Face validity:** When a test appears to measure what the test user wishes or intends to measure, it is said to possess face validity. Thus, face validity refers not to what the test measures, but what the test appears to measure, i.e., whether it seems to be relevant to its various objectives. Test contents should not appear to be inappropriate or irrelevant. If a test measures what the test author desires to measure, we say that the test has face validity. Face validity does not require any statistical technique. It is based on subjective judgment. Whenever a test is prepared, it is submitted to experts to assess if it measures what it intends to measure. It is the first step in validating the test. Once the test is validated at face, we may proceed further to compute validity coefficient. For example, suppose we prepare a test to measure 'skill in division'. If all the questions on the test are related to division, we can say that the test has face validity.

(ii) **Content validity:** An analysis of the content of an assessment evaluates the appropriateness of the content and determines the extent to which the assessment tasks provide a relevant and representative sample of the content under consideration. Content considerations are especially very important when validating achievement testing or constructing classroom tests. Content validity refers to the extent to which a test contains items representing the behaviour that we are going to assess. It is generally assessed by a critical scrutiny of the test contents to determine whether they cover a representative sample of behaviour or not. As it is measured from the content, such type of validity is referred to as content validity Content validity of a test is estimated by assessing the presence of the content to be assessed in the test paper. Suppose we want to construct an achievement test on English. So all the questions of the test should be related to prose, poetry, novel, grammar, etc., and all the items must measure the different behavioural objectives like knowledge, understanding, application, analysis, synthesis and evaluation.

Content validity rests upon an expert analysis on the items included in the test. Content validity is also known as 'curricular validity' or 'rational validity' or 'logical validity'.

(iii) **Concurrent validity:** The term 'concurrent validity' is used to refer the process of validating a new test by correlating it with some available source of information which might have been obtained shortly before or shortly after the new test is given. Concurrent validity indicates to which extent the test scores correspond to already accepted measures of performance (or status made at the same time). Suppose we want to administer an intelligence test upon an individual. Now the test is administered upon the individual and the intelligence scale is administered upon the same individual. If the coefficient of correlation is high, the intelligence test is said to have high concurrent validity. Concurrent validity is relevant to the tests employed for diagnosis. When new tests are validated against previous tests, these previous or established tests are known as criteria for the new tests.

(iv) **Construct validity:** A 'construct' is an individual characteristic that we assume exists in order to explain some aspect of behaviour. Whenever we wish to interpret assessment results in terms of some individual characteristics (e.g., reasoning, problem-solving activity), we are concerned with a construct. The construct validity of a test is the extent to which the test may be said to measure a theoretical construct or trait. Examples of such construct are scholastic aptitude, mechanical comprehension, anxiety, neuroticism, etc. Construct validation requires the gradual accumulation of information from a variety of sources. When we interpret assessment results as a measure of a particular construct, we are implying that there is such a construct that differs from other constructs, and that the results provide a measure of the construct, i.e., little influenced by extraneous factors. Verifying such implications is the task of construct validation. Although construct validation has been commonly associated with theory building and theory testing, it also has usefulness for the practical use of assessment results. It takes place primarily during the development and try-out of a test or an assessment, and is based on an accumulation of evidence from many different sources. When selecting a published test that presumably measures a particular construct, such as logical reasoning or writing comprehension, the test manual should be tested to determine what evidence is represented to support the validity of the proposed interpretations.

(v) **Predictive validity:** Predictive validity of a test refers to the predictive capacity of a test. It refers to the effectiveness of a test in predicting future outcomes in a particular area. The word 'prediction' may be used in more ways than one. In the broader sense, it would mean prediction from the test to a criterion situation. In the limited sense, however, it would imply prediction over a time period. When used in the limited context, it may be referred to in the sense of 'predictive validity'. Predictive validation provides the most relevant information for tests used in the selection and classification of personnel. Some other uses include hiring job applicants, selecting students for admission to college or professional schools, and assigning military personnel to occupational training programmes. Take an example where we have prepared an entrance test for admission into B.Ed. course, and based on the scores, we have admitted the candidates. These candidates completed the B.Ed. course and appeared for the final B.Ed. examination. The scores of the entrance test and the scores of the final B.Ed. examination are correlated. If the coefficient of correlation is high, we say that the entrance test has high predictive validity.

(vi) **Criterion validity:** Criterion validity evidence tells us just how well a test corresponds with particular criteria. Criterion validity includes both predictive validity and concurrent validity. A test may be validated against as many criteria as there are specific uses for it. Whenever test scores are to be used to predict future performance or to estimate current performance on some valued measure other than the test itself (called a criterion), we are especially concerned with evaluating the relationship between the test and the criterion. For example, speaking-readiness test scores might be used to predict a student's future a achievement in speaking, or a test of dictionary skills might be used to estimate

the student's current skills in the actual use of the dictionary. The first example is of predictive validity and the second example is of concurrent validity.

3.6.3 Methods of Validation

Different types of validity needs different methods for assessment. The methods used for assessing the validity of a test are discussed below.

(i) **Correlation method:** Correlation method is used in most of the cases for calculating the validity of a test. Multiple correlation is used where more than two measures are involved in which English mathematician and biometrician Karl Pearson's 'r' is used. Several methods are used for this, but the following method is a popular one.

Where,

r = Validity index

N = Size of sample

X = Raw scores in the test X

Y = Raw scores in the test Y

XY = Sum of the products of each X score multiplied with its corresponding Y score

In order to make the calculation an easy one, we can use the above formula in this way:

Where,

r = Validity index

N = Size of the sample

dx = X – M

dy = Y – M

X = Raw score of X group

Y = Raw score of Y group

M = Mean

Sometimes, we need to predict the future result of somebody with reference to the present result. The following regression equation is used for this purpose:

Where,

Y = Predicted value

My = Mean of predicted score (Y)

Mx = Mean of test score (X)

 = Standard deviation of predicted score (Y)

 = Standard deviation of test score (X)

X = Test scores (basis of prediction)

Y = Predicted value

(ii) **Cross validation:** Cross validation indicates a process of validating a test by using a population sample that is different from the sample on which it was originally standardized. It is necessary because the validity data may be high or low due to chance factors peculiar to the standardization sample. When test is administered to various samples in a variety of situations, it is being cross-validated. The different types of cross validation are: validity extension, validity generalization and psychometric signs.

3.7 FACTORS AFFECTING VALIDITY

Numerous factors tend to make assessment results invalid for their intended use. Let us have a discussion regarding the factors which affect the validity of a test.

(i) **Lack of clarity in directions:** Instructions that do not clearly indicate to the student how to respond to the tasks and how to record the responses decreases validity of a test. If the directions are not clear, the students will misunderstand the purpose of the test, and this in turn will hamper the validity of the test.

(ii) **Ambiguity:** Ambiguous statements lead to confusion and misinterpretation. Ambiguity sometimes confuses the good students more than it does the poor students. So no question of the test should be ambiguous.

(iii) **Length of the test:** Lengthening of a test not only increases its reliability but also its validity. If a test is too short to provide a representative sample of the performance we are interested in, its validity will suffer accordingly.

(iv) **Nature of the group:** A test may be valid for one group, but may not be valid for another. In heterogeneous groups, the scores are widely distributed and the validity coefficients are likely to be higher.

(v) **Difficult reading vocabulary and sentence structure:** Vocabulary and sentence structure that are too complex for the students taking the assessment, result in the assessment's measuring reading comprehension and aspects of intelligence, which will lessen the validity of a test.

(vi) **Inadequate time:** Sometimes, the importance is given upon speed test instead of power test. In the field of achievement test, if sufficient time will not be given to the students, it reduces the validity of interpretations of results. However, assessments of achievement should minimize the effects of speed on student performance.

(vii) **Poorly constructed test items:** Sometimes, the test items are very subjective, vague, unclear, not objective, etc., and this affects the validity of a test.

(viii) **Improper arrangement of items:** The items in a test should be arranged according to the difficulty order. It means the items should be arranged from easy to difficult. If difficult items are placed first, it will take time and makes the student confused. So the items of a test should be arranged properly in order to develop the validity of a test.

(ix) **Identifiable pattern of answers:** Correct answers in some systematic order enables a student to guess the right answers more easily, and it affects the validity of a test.

(x) **Factors in administration and scoring:** In the case of teacher-made tests, the factors like insufficient time, unfair aid to individual students who ask for help, cheating and unreliable scoring by students tend to lower validity. In case of a standardized test, failure to follow the standard directions and time limits, or giving students unauthorized assistance or errors in scoring similarly contribute to lower validity. So, all these factors should be checked for ensuring validity of a test.

(xi) **Cultural influences:** Cultural influence, socio-economic status, social class structure, etc., affect the test scores as well as validity of a test.

(xii) **Criterion correlated to the test:** The criterion, for which the validity is assessed, should be a reliable one and free from bias, or else it will affect the validity.

3.7.1 Relationship between Validity and Reliability

Reliability and validity are closely related, even though they cannot be interchanged. An assessment that has very low reliability will also have low validity; quite obviously, a measurement that has low levels of accuracy or consistency is not likely to suitably fullfil its objective. However, at the same time, the factors necessary for achieving a considerably high degree of reliability can affect validity negatively. For instance, consistent assessment settings lead to better degree of reliability because they minimize the 'noise' (variations) in the results. Alternatively, something that can enhance validity is flexibility in assessment exercises and settings. Such flexibility enables assessment to be customized to the learner's context and to be made specific and suitable for certain groups of students. Insistence on entirely consistent assessment settings for achieving high reliability will lead to minimum flexibility, and may, in turn, bring down validity.

For every dimension of interest and specific query or set of queries, there are multiple ways to devise questions. Even though the focal point should always remain the predetermined purposes of the research, there are good or poor questions for any particular operationalization. How can you evaluate the measures?

Two chief criteria of evaluation in any measurement or observation are as follows:

1. If we are measuring what we set out to measure

2. If the same measurement process always shows the same results These two criteria are validity and reliability.

The key concerns of 'reliability' are stability and consistency. This means that if you are using a certain tool to make measurements, is that tool giving you the same result each time you use it? To understand the concept more clearly, consider measurement processes in different contexts, such as for woodwork or construction work, a tape measure is an extremely reliable measuring tool.

Suppose you have a piece of wood that is 2½ feet long. You measure it once with the tape measure and find that it is of 2½ feet. Suppose you measure it again and you get the same result—2½ feet. If you measure it over and over again and you get the

measurement of 2½ feet each time, this means that the tape measure is capable of giving you reliable results.

Validity is concerned with the extent of measurement of what we hope to measure (and what we think we are measuring). To continue with our previous example of measuring the wood block, a tape measure that shows accurately spaced inches, feet, etc., should show up valid results as well. Measuring the wooden block with a 'reliable' tape measure should give an accurate measurement of the block's length.

When applying these concepts to educational research, it is better to use measurement tools that are both reliable and valid. So, this means that a test should have questions that receive consistent answers even when asked a number of times— this is reliability. Similarly, a test should contain questions that get correct responses from respondents — this is validity.

3.8 SUMMARY

In this unit, you have learnt that:

- Standardization of a test requires the important characteristics, i.e., validity. If the objectives of a test are fulfilled, then we can say that the test is a valid one.

- If a test is valid, obviously it is reliable; but if a test is reliable, it may or may not be valid.

- Content considerations are especially very important when validating achievement testing or constructing classroom tests.

- Predictive validity of a test refers to the predictive capacity of a test.

- Criterion validity evidence tells us just how well a test corresponds with a particular criterion.

- Correlation method is used in most of the cases for calculating the validity of a test.

- Cross validation indicates a process of validating a test by using a population sample which is different from the sample on which it was originally standardized.

- Reliability refers to consistency of scores obtained by some individuals when re-tested with the test on different sets of equivalent items or under other variable examining conditions.

- To estimate reliability by means of the test-retest method, the same assessment is administered twice on the same group of pupils with a given time interval between the two administrations.

- Reliability can also be calculated from a single administration of a single test. The test is administered to a group of pupils and then is divided into half for scoring.

- The standard error of measurement is the standard deviation of the errors of measurement.

- Face validity refers not to what the test measures, but what the test appears to measure, i.e., whether it seems to be relevant to its various objectives.

- Criterion validity includes both predictive validity and concurrent validity.

- A high test-retest reliability or coefficient of stability shows that there is low variable error in the sets of obtained scores and vice versa.

3.9 KEY TERMS

- Validity: It is the extent to which a test measures what it intends to measure.

- Item objectivity: The item must call for a definite single answer.

- Concurrent validity: It is the extent to which the test scores correspond to already accepted measures of performance.

3.10 ANSWERS TO 'CHECK YOUR PROGRESS'

1. A test is a stimulus selected and organized to elicit responses which can reveal certain psychological traits in the person who deals with them.

2. Objectivity of a test indicates two things—item objectivity and Scoring objectivity.

3. Reliability refers to the extent to which a measuring device yields consistent results upon testing and retesting.

4. The characteristics of reliability are as follows

   (i)   It refers to the preciseness of a measuring instrument.

   (ii) It is the coefficient of internal consistency and stability.

5. Two parallel forms of a test can be constructed by selecting the samples in each form from the same universe of contents.

6. The limitations of Kuder—Richardson method are as follows:

   (i)   The formulae indicate the consistency of student response from one day to the other.

   (ii) It cannot be used for power test and heterogeneous tests.

7. One of the method of calculating reliability is the split-half method.

8. The limitations of test-retest method are as follows:

   - Testing conditions during test and retest may vary which results in instability in the scores.

   - The individual's health, emotional conditions, mental health, motivational conditions and mental set do not remain the same in both the administrations at two occasions.

   - If the test is repeated immediately, the test-taker may recall the first answer. This may increase the scores. Besides the memory effects, practice and confidence induced by familiarity with the test will almost affect the scores the second time. This is known as 'the carry-over effect' or 'transfer effect' or 'memory effect' or 'practice effect'.

   - If the time gap between two administrations is a lengthy one, additional learning or changes in the characteristics of the individual will affect the scores at the later administration, and it may decline the reliability of a test.

9. Item objectivity refers to the item that must call for a definite single answer. In an objective-type question, a definite answer is expected from the test-takers. While framing the questions, some points to be kept in mind are: ambiguous questions, lack of proper direction, double barreled questions, questions with double negatives, etc. These concepts affect the objectivity of a test.

10. The validity of a test is determined by measuring the extent to which it matches with a given criterion.

11. The methods for assessing the validity of a test are as follows:

    (i) Correlation method

    (ii) Cross validation

12. 'Validity' refers to the degree to which a measuring tool yields consistent results upon testing and retesting.

13. The types of validity are: (i) face validity, (ii) content validity, (iii) concurrent validity, (iv) construct validity, (v) predictive validity and (vi) criterion validity.

14. Content validity refers to the extent to which a test contains items representing the behaviour that we are going to assess.

15. Concurrent validity refers to the extent to which the test scores correspond to already accepted measures of performance (or status made at the same time).

16. The construct validity of a test is the extent to which the test may be said to measure a theoretical construct or trait.

17. The different types of cross validation are: validity extension, validity generalization and psychometric signs.

3.11 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What are the characteristics of a good and standard test?

2. What do you understand by the validity of a test? Write a short note on the different types of validity.

3. List the similarities and differences between standardized test and teacher- made test.

4. Write a note on the relationship between validity and reliability.

Long-Answer Questions

1. Explain any two methods of computing reliability in detail.

2. Explain the various methods of validation of a test.

3. Give a detailed account on the factors that affect reliability.

4. Give a detailed account on the factors that affect validity.

3.12 FURTHER READING

Annastasi, A. 1976. Psychological Testing, 4th Ed. New York: McMillan Publishing Co.

Block, J. H, L. W. Anderson. 1975. Mastery Learning in Classroom Instruction. New York: McMillan Camp.

Bloom, B. S. 1956. Taxonomy of Objectives of Cognitive Domains. New York: Longmans Green &Co.

Bloom, B. S. 1956. Taxonomy of Objectives (Affective Doming). New York: Longmans Green & Co.

Chronbach, L. J. 1970. Essentials of Psychological Testing, 3rd Ed. New York: Harper.

Catel, R. B. Personality: A Systematic Theoretical and Factual Study. New York: McGraw Hill.

Ebel, R.L. 1972. Essentials of Educational Measurement. New Jersey: Englewood Cliff.

Edwards, A.L. 1957. Techniques of Attitude Scale Construction. New York: Application Century.

Frank S. Freeman. Theory & Practice of Psychological Testing. Mumbai: IBII Publishing Co.

Grunlund, N. E. 1970. Stating Behavioural Objectives for Classroom Instruction. New York: McMillan.

Grunlund, N. E. 1973. Preparing Criterion Referenced Tests for Classroom Instruction. New York: McMillan.

Grunlund, N. E. 1976. Measurement and Evaluation in Teaching (3rd Ed). New York: McMillan Publishing Co.

Goslin, D. A. 1967. Teachers and Testing. New York: Sage Foundation.

Grunlford, J. P. 1967. The Nature of Human Intelligence. New York: McGraw Hill.

Grunlund, N. E. 1959. Sociometry in the Classroom. New York: Harpt & Row.

Julian, C., Stanley and Hopkins. 1972. Educational and Psychological Measurement and Evaluation. New Delhi: Prentice Hall of India Pvt. Ltd.

Lindermon, Richard, H. Educational Measurement. Mumbai: B. S. Taraporevala Sons & Co.

Lynum, H. B. 1971. Test Scores and What They Mean. New Jersey: Prentice Hall.

Ten Brink, T. D. 1974, Evaluation: A Practical Guide for Teachers. New York: McGraw Hill.

Bernard, H. W., D.W Fullmer. Principles of Guidance: A Basic Text (Indian Edn.). New Delhi: Allied Pub.

Crow Lester, D., A. Crow. An Introduction to Guidance Principles and Practices. New York: American Books Co.

Downing, Lester, N. 1964. Guidance and Counselling Services. New York: McGraw Hill.

Fusterm J. N. Psychological counselling in India. Mumbai: McMillan.

Kochhar, S. K. 1987. Educational Vocational Guidance in Secondary Schools. New Delhi: Sterling Publisher Pvt. Ltd.

Hoose, William, N. Van. (ed). 1979. counselling and Guidance in 20th Century. Boston: Noughton Miffin Co.

Khorshed, A. W. Guidance Movement in India, Guidance Services. New Delhi: NCERT.

Pandey, K. P. Educational and Vocational Guidance in India. Varanasi:

Vishwavidyalaya Prakashan.

Pasricha, Pren. 1976. Guidance and Counselling in India Education. New Delhi: NCERT.

Swamy, R. V. (ed). 1971. Guidance Services in Colleges and Universities. Bangalore: University and Directorate of Employment and Training.

Traxler, R. E., R. D. Norla. Techniques of Guidance. New York: Harper and Road.

UNIT 4 CONSTRUCTION AND STANDARDIZATION OF ACHIEVEMENT TEST

Structure

4.0 INTRODUCTION

Tests and other means for measuring progress of the learners serve as a basis for instructional decisions of educators. Precisely for this, the teacher is supposed to prepare good quality tests to bring out the understanding and achievement capacity of the learners. A good test should have qualities such as validity, reliability and usability. The assessment instrument should include psychometric characteristics (difficulty, discrimination and destructor indices). It has been understood through various studies that there is a need for constructing standardized achievement test for learners. One frequent use of standardized achievement tests is to identify' students who perform

below, at the same level, or above their peers through uniform procedures in scoring, administering and interpreting results.

Norm-referenced testing is considered an important part of psychological and educational testing. This method is largely statistical and works on the assumption that human traits and characteristics, such as intelligence, academic achievement and behaviour, are distributed along a normal probability or bell-shaped curve or normal curve. A normal curve is used to represent the average or standard performance level of a sample population and also the exceptions to the norm (over or below average). The norms for a test are percentile ranks, standard scores and other statistics for the norm group on the basis of which the test standards are formulated. A certain percentage of the norm group falls within various ranges along the normal curve. Based on the range within which test scores fall, scores may be rated as ranging from poor to superior.

In this unit, you will learn about norms in measurement and classification of tests, such as standardized tests and teacher-made tests.

4.1 UNIT OBJECTIVES

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

- Define norm
- Discuss different types of norms
- Explain the different types of achievement tests
- Discuss the steps of constructing achievement tests

4.2 NORMS IN MEASUREMENT: DEFINITION AND SIGNIFICANCE

In words of American psychologists Thorndike and Hagen, 'Norms are defined as the average performance on a particular test made by a standardization sample'. Frank S. Freeman defines a norm in measurement as 'the average or standard score on a particular test made by a specified population'.

Norms are representations of average performance based upon the results of testing a specified group of students. It is a device of transforming raw scores into standard scores in a group. Educational measurement is not a case of physical measurement, but a case of mental measure. Therefore, educational measurement is relative. In case of this, we have to go for the evaluation of an individual in relation to other in a class or a group. Let us take an example. Suppose Prakash is a student of Class X. In the classroom examination, he scored 60 in English and 75 in Mathematics. If a layman analyses the result of Prakash, he will say that Prakash is better in Mathematics than in English. Prakash scored the highest marks in English test but the lowest marks in Mathematics in the class. So the interpretation made by the layman was wrong because the two scores are raw scores.

In order to get a valid and reliable result, we should go for a norm, through which we can get the perfect place of an individual in a group. For this purpose of interpretations, the raw scores are to be converted into derived scores. Norms tell us where a student or a class stands in relation to a reference group. Norms are representation of average performance based upon the results of testing a specified group of students. It is a statistical procedure to minimize the interpretive error of a test score. The norms of any educational test represent the average test performance of the standardized group or

sample selected from a specified population. In this, an individual score is compared with the standardized sample as a reference group. The importance of norms in the field of measurement and evaluation are explained as follows:

- Norms are the basis of interpreting raw scores into derived scores.
- Norms place an individual in the exact place within a group.
- Norms are helpful for selection and classification of students.
- Norms are helpful in provision of guidance and counselling the students.
- Norms speak about the attainment of instructional objectives by the students.
- Norms help in minimizing the interpretive error of a measuring instrument.

4.2.1 Types of Norms

In the field of educational and mental measurement, we use four types of norms, which are as follows:

1. Age norm

The concept of age norm was developed by French psychologist Alfred Binet in 1908. It basically deals with mental age. This age norm is also known as 'mental age norms' or 'age equivalent norms'. The 'age norm' is defined as the average performance of a representative sample of certain age group on the measure of intelligence or ability. Let us consider a suitable example to have clarity about age norms.

Suppose the average score of students of age 15 years 2 months on an achievement test is 80. So the age norm for the score of 80 will be 15 years 2 months. Suppose Mohan is 12 years old and he scores 80 in the achievement test. Here, though his chronological age is 12, Mohan's mental age is 15 years 2 months.

So, when that age norm is fixed, standardized test is given to a representative sample of students of a particular age level and the average score is calculated, and this score is considered as the norm for the group. The students who achieve that score are considered within that age norm.

Limitations of age norm

The limitations of age norm are as follows:

- It is very difficult to get a true representative sample of individuals of a selected age group.
- In case of very high and very low scores, it is difficult to interpret it with age norms.
- Mental age units are not fixed in case of different tests; it may vary.
- It has the limited scope to be used in some psychological and educational tests.
- Age norms lack a standard and uniform unit throughout the period of growth of physical and psychological traits.
- It is the difficult and a time consuming task to develop the age norms and mental age.

- The mental age of a particular age group may differ from locality to locality and test to test.

2. Grade norm

Grade norms are also like age norm. However, here, measurement is based upon class or grade level, not on age level. Grade norms have been widely used with standardized achievement tests, especially at the elementary school level. The grade equivalent that corresponds to a particular raw score is identified as the grade level at which the typical student obtains that raw score. A grade norm corresponding to a raw score is the grade level of those pupils whose average raw score is the raw score in question.

Suppose we conducted a test on the VIIIth grade students. After getting the result, we get 60 to be the average of that test. Therefore, 60 will be the grade norm for the students of VIIIth grade. Grade norms are based on the average performance of students at various grade levels. Grade norms are most useful for reporting growth in the basic skills during the elementary school period. They are least useful for comparing a student's performances on different tests.

Limitations of grade norms

The limitations of grade norms are as follows:

- The rate of growth from grade to grade is not uniform throughout.
- Grade norms lack a comparability of scores on different tests.
- When a student of VIIth grade gets a credit of IXth grade, it does not mean that the student has the ability to be a student of IXth grade.
- Fractional grades do not have any meaning.
- Grade norms are affected by quality of schools, quality of teachers and quality of students.
- The interpretation of grade norm is very confusing because it provides only level of performance with respect to a subject rather than the educational level of the students.

3. Percentile norm

Percentile norms are about the position of an individual in relation to the norming group, 'Percentile norm' is a point on the scale of measurement determined by the percentage of individuals in given populations that lie below this point. It describes a student's performance in terms of the percentage of other students in some clearly defined group that earn a lower score. This might be a grade or age group, or any other group that provides a meaning comparison. If the percentile norm of a score 60 is 65, we mean that 65 per cent of the students of the normative group lie below a score of 60.

Percentiles should not be confused with the common 'percentage score'. The percentage scores are raw scores, whereas percentiles are transformed scores. It provides a basis for interpreting an individual's score on a test in terms of his own standing in a particular standardization sample. It should be based upon a sample which has been made homogeneous with respect to age group, sex group grade level, socio-economic status, etc. This is applicable for all types of tests: intelligence, attitude, aptitude and achievement.

Limitations of percentile norm

- The percentile units are not equal on all parts of the scale. The percentile difference of 5 near the middle of the scale (e.g., 45 to 50) represents a much smaller difference in test performance than the same percentile difference at the ends (e.g., 85 to 90), because a large number of students receive scores near the middle, whereas relatively few students have extremely high or low scores.

- Percentile norms are generally confused with percentage scores which affects the interpretation.

- Percentile norm indicates only the relative position of an examinee in the standardization sample. It conveys nothing regarding the amount of the actual difference between the scores.

- Percentile rank of one group cannot be compared with percentile rank of another group.

- Conversion of raw scores to percentile scores give more differences in the middle than at the extremes.

  We can use the following formula to compute percentile rank of raw scores:

  where PP = Percentile point of a raw score

        L = Lower limit of the raw score falls in the particular class interval

         = Percentage of the frequency

         = Frequency below the class interval

         = Actual frequency of the class interval

        i = Size of the class interval

4. Standard score norms

The most important method used to indicate an individual's relative position in a group by showing how far his achieved score is above or below average. This is the approach of standard score and standard scores express performance of the individuals in terms of standard deviation units from the mean. There are numerous types of standard scores used in testing. They are as follows:

(a) Z-Scores: Z-score represents test performance directly as the number of standard deviation units a raw score is above or below the mean. The Z-scores are the units of normal probability curve, ranges from − 3 to + 3, with mean value zero and standard deviation one.

  The formula for computing Z-score is

Where,

X = Raw score

M = Arithmetic mean of raw scores

SD = Standard deviation of raw score

A Z-score is always negative when the raw score is smaller than the mean.

(b) T-Scores: T-scores are also the standard scores but the mean value is 50 and standard deviation is 10, T-scores can be obtained by multiplying the Z-score by 10 and adding the product to 50. Thus,

$$\text{T-score} = 50 + (10\ Z)$$

One reason that T-scores are preferred to Z-scores for reporting test results is that only positive integers are produced in T-scores.

(c) **Stanines:** The stanine norm is developed by the technique of normalized standard scores. It was developed by the US Air Force during the World War II. Stanines are single digit scores ranging from 1 to 9. This system of scores is so-called because the distribution of raw scores is divided into nine equal parts. Stanine 5 is in the centre of the distribution and includes all cases within one-fourth of a standard deviation of either side of the mean. Here, the mean score is 5 and standard deviation is 1.96 or 2. When raw scores are transformed into stanine scores, the distributions of scores take the shape of normal curve.

In the Stanine system, a 9-point scale is used, in which 9 is high, 1 is low and 5 is average. Stanines are normalized standard scores that make it possible to compare a student's performance on different tests.

The stanines are distributed on normal curve. The area of normal probability curve has been divided into nine standards with a fixed percentage. The first stanine includes 4 per cent, second stanine includes 7 per cent, third stanine includes 12 per cent, fourth stanine includes 17 per cent, fifth stanine includes 20 per cent, sixth stanine includes 17 per cent, seventh stanine includes 12 per cent, eighth stanine includes 7 per cent and the ninth stanine includes 4 per cent of the total cases.

Stanine position	Description	Percentage	Stanine's
1, 9	Bottom and top	4	(1st) (9th)
2, 8	Above bottom and below top	7	(2nd) (8th)
3, 7	Near to second or eighth	12	(3rd) (7th)
4, 6	Above or below mean	17	(4th) (6th)
5	Middle or mean	20	(5th)

Judging the adequacy of norms

- Test norms should be relevant to the reference group.
- Test norms should be comparable with other norms.
- A clear description is necessary for every test norm.

- Test norms should be changed from time to time.
- Test norm should represent the whole group.

Limitations of standard score norms

The limitations of standard score norms are as follows:

- The interpretation of the standard scores becomes problematic when the distribution is not normal.
- It needs expertise knowledge to deal with the standard scores.
- The minus values arc very confusing in case of educational and mental measurement.
- Sometimes the raw scores are not normally distributed: they arc positively skewed or negatively skewed. This linear transformation is based on the assumption of normal distribution.

4.3 CLASSIFICATION OF TESTS

Tests are divided into different types taking into consideration their content, objective, administration system, scoring style, etc. According to mode of administration, tests are of two types:

(i) **Individual test:** When a psychological test is administered on an individual at a particular time, it is known as 'individual test'.

(ii) **Group test:** When a test is administered on a group of individuals at a particular time, it is known as 'group test'. It is mostly applicable on adult literates.

According to the ability of the student, tests are of two types:

(i) **Speed test:** This type of test is applicable on the individuals to know the mental speed. Here, the time is limited and the number of questions is more and all the questions are equal in difficulty level. Railway examinations, banking examinations are examples of speed test.

(ii) **Power test:** This type of test is applicable on individuals to know their mental power or their ability. Here, there is no time limit and the individuals are expected to answer the questions within as much time they like. All the questions of this test are arranged according to difficulty level and discriminating power. The essay competitions by the media are the bright examples of power test.

According to the type of items involved in the test, it can be of three types:

(i) **Essay-type test:** Essay-type tests are otherwise known as open-ended tests. The essay question is especially useful for measuring those aspects of complex achievement that cannot be measured well by more objective means. These include: (a) the ability to supply rather than merely identify interpretations and application of data and (b) the ability to organize, integrate and express ideas in a general attack on a problem. Outcomes of the first type are measured by restricted-response questions and outcomes of the second type by extended-

response questions. For example, 'discuss the educational philosophy of M. K. Gandhi'.

(ii) **Short-answer type test:** This type of test requires to be written in a short-cut manner regarding a concept. It is suitable for measuring a wide variety of relatively simple learning outcomes, and it is used almost exclusively to measure the recall of memorized information. For example: 'What is measurement? Write within 50 words.'

(iii) **Objective-type test:** In objective-type questions, the individual is expected to answer the question with the help of a word, a phrase, a number or a symbol. The test with multiple-choice items, true-false items, matching type items, fill-in-the blanks items, one-word substitution are the examples of objective type test.

According to the method of scoring, tests are of two types:

(i) **Machine-scored test:** The tests which are scored or assessed by the machines like computer are known as 'machine-scored test'. The Bank P.O. examination is an example of machine-scored test.

(ii) **Hand-scored test:** The tests which are assessed by the human beings are known as 'hand-scored tests'. The classroom achievement tests is an example of hand-scored tests.

According to the principle of test construction, tests are of two types:

(i) **Teacher-made test:** Generally 'teacher-made tests' are prepared by classroom teachers to assess pupils' growth. It is related to action research. Teacher-made tests serve different purposes, viz., to measure pupil's achievement, to know how far the specific objectives have been fulfilled, to diagnose the learning, difficulties, and to arrange specific remedial measures to award grades, etc. This type of test only follows two steps—planning and preparation.

(ii) **Standardized test:** Standardized tests measure the common objectives of a wide variety of schools. They have standard procedures for administration and scoring, and provide norms for interpreting the scores. A test manual and other necessary material are typically provided to aid in the administration of the test and the interpretation and use of the results. The test items are generally of high quality because they have been prepared by specialists, subject experts, pre-tested and selected on the basis of their effectiveness and their relevance to a rigid set of specification. They are especially useful for measuring general educational development, determining student's progress from one year to the next, grouping students, analysing learning difficulties, and comparing achievement with learning ability.

Standardized test and teacher-made test have been discussed later in this unit in detail.

According to the nature of the test, they are classified as follows:

(i) **Oral test:** It is a kind of verbal test. In oral test, the individual is expected to answer orally. This type of tests is mostly applicable to illiterates or small children. In the public survey, the people are asked to speak something regarding

the issue. In the interview board, the interviewers ask questions to the interviewee and the interviewee answers orally. This type of test is known as 'oral test'.

(ii) **Written test:** Here, the individual has to respond the questions in writing form. So the respondent should have the writing ability. It is only applicable upon the literates. All the written examinations are the examples of written test. It is a kind of verbal test.

(iii) **Performance test:** This type of test is also known as 'non-verbal test'. The respondent is not expected to respond verbally. He has to perform the task. The running competition, jumping competition held by physical examination are the examples of performance test.

According to Robert Glaser, tests are of two types:

(i) **Norm-referenced test:** It is a test, or type of assessment, designed to provide a measure of performance interpretable in terms of an individual's relative standing in some known group. For example, Ram types better than 80 per cent of the class members. Norm-referenced interpretations depend on a comparison of a student's performance to that of other students whose performance defines the norms. The norms might be based on a local, state or national group, depending on the use to be made of the result. The norm-reference interpretation indicates the student's relative standing in a norm group by noting the percentage of students in the group who obtain the same or a lower score (called percentile score), An identifying feature of norm-referenced test is the selection of items of average difficulty and the elimination of items that all students are likely to answer correctly. This is useful for decisions based on relative achievement such as selection, grouping and relative grading.

(ii) **Criterion-referenced test:** It is a test, or type of assessment, designed to provide a measure of performance interpretable in terms of a clearly defined and delineated domain of leaning tasks. For example, Prakash types 80 words per minute without error. When interpretations are confined to the attainment of a specific objective, they are called objective referenced or criterion-referenced interpretation. Current-standard based assessments provide major example of criterion-referenced interpretations. The criterion-referenced interpretation focuses on the percentage of items answered correctly (called percentage on the correct score). Criterion-referenced tests include items that are directly relevant to the leaning outcomes to be measured, without regard to whether the items can be used to discriminate among students. No attempt is made to eliminate easy items or alter their difficulty. The goal of the criterion-referenced test is to obtain a description of the specific knowledge and skills each student can demonstrate.

4.3.1 Standardized Achievement Tests

When the concept of evaluation takes its place in a wider view, we talk about standardized tests. Standardized tests may be standardized achievement tests, standardized intelligence test, etc. Mostly, we talk about standardized achievement tests in the teaching-leaning situation.

Standardized achievement tests are typically norm-referenced tests that measure the students' level of achievement in various content and skill areas by comparing their test performance with the performance of other students in some general reference group (i.e., a nationwide sample of students at the same grade level). Standardized tests are prepared by the subject experts and they have more validity and reliability in comparison to teacher-made tests.

The quality of items in standardized tests is very high. The items are formulated by specialists, pre-tested and selected on the basis of effectiveness. The reliability of standardized tests is very high commonly between 80 and 95, frequently around 90. During the procedure of standardization of a test, some specific instructions are followed by the subject experts. The scores obtained by the test can be compared with those of norm groups.

Characteristics

The characteristics of standardized tests are as follows:

- The test items are of a high technical quality with validity and reliability.
- The items are developed by subject experts.
- The items are pre-tested and selected on the basis of difficulty and discriminating power
- Directions for administration are given in these tests.
- Scoring key is provided to the examiners during scoring of papers.
- Norms based on national samples of students in the grades, where the test is intended for use is provided as aid in interpreting the test scores.
- A test manual and other necessary accessory materials are included as guides for administering and scoring the test, evaluating its technical qualities, interpreting, and using the results.
- The reliability of standardized tests is generally around 90.
- Derivation of norms is apart of the process of standardization.
- Standardized tests generally follow four steps, viz., planning, preparation, tryout and evaluation.
- In case of standardized tests, usually objective type items are used.
- All the items of a standardized test are arranged from lower difficulty level to higher difficulty level.
- Standardized tests are useful for basic or fundamental research.

In case of standardized tests, the process of standardization includes all the following elements:

- Standardization of the content
- Standardization of methods of administration
- Standardization of methods of scoring
- Standardization of process of interpretation

Steps

A teacher-made test follows only two steps (planning and preparation) in its construction, but standardized test follows four steps for its construction as discussed in this section.

1. **Planning of the test:** At the planning stage, the subject experts plan about the whole test. This contains the following elements:

   • Number of units to be covered by the test

   • Types of questions to be included in the test

   • Instructional objectives that are going to be tested

   • Number of items to be included in the test

   • Total marks to be given to the test

   • Instructions that are to be given in the test

   • Preparation of a scoring key

Everything regarding the test is planned during this stage. A blueprint is prepared for the test which contains contents, objectives, the type of questions, and the weightage to be given to different content areas according to the type of items and objectives. If a test is to be successful, a careful planning must precede its construction.

(ii) Preparation of the test: The second step in the construction of a standardized test is preparation. At this stage, the following points are taken into consideration:

   • Types of test items

   • Directions to test items

   • Directions for administration and scoring

   • Question-wise analysis chart

Planning stage is the theoretical concept of a standardized test, but preparation stage is a practical concept. In this stage, all the items and instructions are prepared by the subject experts.

(iii) **Try-out of the test:** During the try-out step, some samples are collected to administer the same test upon them. Sometimes, both preliminary try-out and final try-out are followed to make the test items more valid and reliable. This try-out step helps in finalizing the number of items to be included in the final test.

   Let us take an example to make a clear-cut understanding about this step. Suppose a standardized test is to be prepared for the class X students of Assam. After preparing all the items, the test should be administered on a sample of Xth class students of Assam. This try-out test helps in rejecting the poor items improving the items which need improvement, finalizing the number of items to be included in the final draft of the test. After the try-out form of the test is administered, the answer sheets are scored as per the scoring key. Now the scores are ready for item analysis and evaluation of the test.

(iv) **Evaluating the test:** At this stage, the scores of the items are ready for evaluation. Whether the items are valid or not, good or poor, reliable or not,

everything is determined in this stage. For the purpose of evaluation, the following points are followed:

- Analysing the items to consider their value of inclusion in the final test
- Determining the validity and reliability of the test
- Assessing the objectivity and usability of the test

After all these steps, the final draft of the test is prepared. In the final draft, the items of good validity, objectivity usability and reliability, appropriate difficulty level and satisfactory discriminating power are retained.

4.3.2 Teacher-Made Test

Evaluation is the part and parcel of teaching-learning situation. How far the instructional objectives have been achieved, how far the curriculum and methods of teaching is suitable for the students, and other such things can be known through the process of evaluation. Test is a tool for evaluation. In the classroom situations, the teachers develop tests to measure and evaluate students' achievement. Different subject teachers construct different tests for their students. Teacher-made tests serve the purposes, viz., measure achievement of the students, know how far the specific objectives have been fulfilled, diagnose the learning difficulties, arrange specific remedial measures, award grades, etc. Teacher-made tests are based on content and objective specific to teacher's own class or school. It may deal either with a specific limited topic or skill or with larger segments of knowledge and skill. Generally, the teacher-made tests are less reliable in comparison to the standardized test. It has a limited use, to a particular class or a single school. Teacher-made tests are well adapted to outcomes and content of local curriculum.

Flexibility affords continuous adaptation of measurement to new materials and changes in procedure. It too often neglects complex learning outcomes.

The quality of items of the teacher-made tests is lower than standardized tests because of a teacher's limited time and lack of opportunity to pretest items. The score comparisons and interpretations of teacher-made tests are limited to local school situation. Teacher-made tests evaluate student's day-to-day progress and their achievement on work units of varying size.

Standardized tests follow four steps for its construction, but teacher-made tests follow only two steps for its construction. Such tests do not need rigorous process of standardization. They are not subjected to rigorous item analysis, not evaluated for their validity and reliability. Simply, teacher-made tests go through two stages, viz., 'planning' of the test and 'preparation' of the test.

Mostly, teacher-made tests are used for formative evaluation. A teacher-made test may be a full-fledged achievement test covering the whole course of a particular subject or a unit test covering one unit of study. The results of teacher-made tests are useful for action researches. The teacher-made tests are evaluated by the same teacher, who makes the test papers. These tests if skillfully prepared can even serve the purpose of standardized tests. Table 4.1 differentiates between teacher-made tests and standardized tests.

Table 4.1 Comparison of Teacher-Made Tests and Standardized Tests

Teacher-made tests	*Standardized tests*

Teacher-made tests	Standardized tests
Teacher-made tests are prepared by the classroom subject teachers.	Standardized tests are prepared by the subject experts.
Teacher-made tests follow two steps: planning and preparation.	Standardized tests follow four steps: planning, preparation, try-out and evaluation.
Teacher-made tests are useful for the concerned classroom only.	Standardized tests are useful all over the world.
Teacher-made tests are useful for action research.	Standardized tests are useful for basic or fundamental research.
The items are not arranged according to their difficulty level,	The items are arranged according to their difficulty level.
The scores of the teacher-made tests have limited comparison value and their interpretations are limited to local school situation.	The scores of standardized tests can be compared with those of norm groups.
The reliability of teacher-made tests are usually unknown, but can be high if carefully constructed, frequently around 90.	The reliability of standardized tests is high, commonly between 80 and 95,
The quality of items is unknown unless test item file is used.	The quality of items is very high.
Scoring key is not used.	Scoring key is very necessary and hence used.
In case of teacher-made tests, there may or may not be objective type tests.	In case of standardized test, usually objective type tests are used.
Take less time for construction and experts may not be involved in it.	Tests involve so many experts, use so many sophisticated procedures and thus take a long time for construction.
Based on content and objectives specific to teacher's own class or school.	Based on content and objectives common to many schools throughout the country.

4.4 SUMMARY

In this unit, you have learnt:

- Standardization of a test requires the important characteristics, i.e., validity. If the objectives of a test are fulfilled, then we can say that the test is a valid one.

- Usability refers to the degree to which the test can be successfully used by the teachers/evaluators.

- Interpretability of test refers to how readily the raw scores of the test can be derived and understood.

- The norms of any educational test represent the average test performance of the standardized group or sample selected from a specified population. In this, an individual score is compared with the standardized sample as a reference group.

- Percentile norm is a point on the scale of measurement determined by the percentage of individuals in a given population that lies below this point.

- The percentage scores are raw scores, whereas percentiles are transformed scores.

- An identifying feature of norm-referenced test is the selection of items of avenge difficulty and the elimination of items that all students are likely to answer correctly.

- The goal of the criterion-referenced test is to obtain a description of the specific knowledge and skills each student can demonstrate.

4.5 KEY TERMS

- **Age norm:** It is the average performance of a representative sample of certain age group on the measure of intelligence or ability.

- **Norms:** Defined as the average performance on a particular test made by a standardization sample.

- **Percentile norm:** It is a point on the scale of measurement determined by the percentage of individuals in given populations that lie below this point.

- **Speed test:** This type of test is applicable on the individuals to know the mental speed.

4.6 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The three basic characteristics of a good test are: (i) validity, (ii) reliability and (iii) objectivity.

2. The four types of norms are: (i) age norm, (ii) grade norm, (iii) percentile norm and (iv) standard score norm.

3. There are three types of standard scores used in testing. These are as follows:

    (i) Z-Score

    (ii) T-Score

    (iii) Stanines

4. Z-Score is calculated as

5. Two types of test according to the ability of the students are: (i) speed test and (ii) power test.

6. Standardized tests measure the common objectives of a wide variety of schools, have standard procedures for administration and scoring, and provide norms for interpreting the scores.

7. The process of standardization includes these elements:

- Standardization of the content

- Standardization of methods of administration

- Standardization of methods of scoring

- Standardization of process of interpretation

8. Standardized tests follow four steps for its construction: (i) planning, (ii) preparation, (iii) try-out and (iv) evaluation.

4.7 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What are the characteristics of a good and standard test?

2. Differentiate between grade norm and percentile norm.

3. What are the characteristics of standardized test?

4. Name the different classification of tests.

Long-Answer Questions

1. Write a detailed note on norms in measurement.

2. Discuss in detail the steps of construction of standardized test.

3. List the similarities and differences between standardized test and teacher- made test.

4. What are the characteristics of a standardized test?

5. Write in detail the disadvantages of age norm and grade norm.

4.8 FURTHER READING

Annastasi, A. 1976. Psychological Testing, 4th Ed. New York: McMillan Publishing Co.

Block, J. H, L. W. Anderson. 1975. Mastery Learning in Classroom Instruction. New York: McMillan Camp.

Bloom, B. S. 1956. Taxonomy of Objectives of Cognitive Domains. New York: Longmans Green &Co.

Bloom, B. S. 1956. Taxonomy of Objectives (Affective Doming,). New York: Longmans Green & Co.

Chronbach, L. J, 1970. Essentials of Psychological Testing, 3 Ed. New York: Harper.

Catel, R. B. Personality: A Systematic Theoretical and Factual Study. New York: McGraw Hill.

Ebel, R.L. 1972. Essentials of Educational Measurement New Jersey: Englewood Cliff

Edwards, A.L. 1957. Techniques of Attitude Scale Construction. New York: Application Century.

Frank S. Freeman. Theory & Practice of Psychological Testing. Mumbai: IBII Publishing Co.

Grunlund, N. E. 1970. Stating Behavioural Objectives for Classroom Instruction. New York: McrMillan.

Grunlund, N. E. 1973. Preparing Criterion Referenced Tests for Classroom Instruction. New York: McMillan.

Grunlund, N. E. 1976. Measurement and Evaluation in Teaching (3rd Ed). New York: McrMillan Publishing Co.

Goslin, D. A. 1967. Teachers and Testing. New York: Sage Foundation.

Grunlford, J. P. 1967. The Nature of Hunian Intelligence. New York: McGraw Hill.

Grunlund, N. E. 1959. Sociometry in the Classroom. New York: Harpt & Row.

Julian, C., Stanley and Hopkins. 1972. Educational and Psychological Measurement and Evaluation. New Delhi: Prentice Hall of India Pvt. Ltd.

Lindermon, Richard, H. Educational Measurement. Mumbai: B. S. Taraporevala Sons & Co.

Lynurn, H. B. 1971. Test Scores and What They Mean. New Jersey: Prentice Hall.

Ten Brink, T. D. 1974. Evaluation: A Practical Guide for Teachers. New York: McGraw Hill.

Bernard, H. W., D.W Fullmer. Principles of Guidance: A Basic Text (Indian Edn.). New Delhi: Allied Pub.

Crow Lester, D., A. Crow. An Introduction to Guidance Principles and Practices.

New York: American Books Co.

Downing, Lester, N. 1964. Guidance and Counselling Services. New York: McGraw Hill.

Fusterm J. N. Psychological Counselling in India. Mumbai: McMillan.

Koehhar, S. K. 1987. Educational Vocational Guidance in Secondary Schools. New Delhi: Sterling Publisher Pvt. Ltd.

Hoose, William, N. Van. (ed). 1979. Counselling and Guidance in 20th Century.

Boston: Noughton Muffin Co.

Khorshed, A. W. Guidance Movement in India, Guidance Services. New Delhi: NCERT.
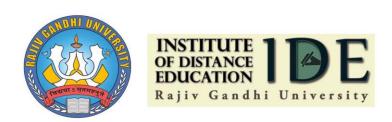
Pandey, K. P. Educational and Vocational Guidance in India. Varanasi: Vishwavidyalaya Prakashan.

Pasricha, Pren. 1976. Guidance and Counselling in India Education. New Delhi: NCERT.

Swamy, R. V. (ed). 1971. Guidance Services in Colleges and Universities. Bangalore:

University and Directorate of Employment and Training.

Traxier, R. E., R. D. Norla. Techniques of Guidance. New York: Harper and Road.

Institute of Distance Education

Rajiv Gandhi University

A Central University

Rono Hills, Arunachal Pradesh