M.A. (Education) 3rd Semester

Course Code: EDUCC 111
Course Type – Core

EDUCATIONAL MEASUREMENT AND EVALUATION

Dr. Vishal Sood
Dr. Shashi Kant Sharma
Dr. (Mrs.) Monika Sood

Centre for Distance and Online Education
Himachal Pradesh University
Gyan Path, Summerhill, Shimla - 171005

M.A. (Education) 3rd Semester
Course Code: EDUCC 111
Course Type – Core
Contents

UNIT-1

TAXONOMY OF EDUCATIONAL OBJECTIVES: COGNITIVE, AFFECTIVE AND PSYCHOMOTOR DOMAIN

Structure:

1.1 Introduction:

Teaching and instructions are organized to achieve the educational objectives. The desired behavioral change is brought among the students to realize the educational objectives. The programmed instructional material is designed to specific educational and to some specific instructional objectives. The teaching and instructional objectives are helpful for achieving the educational objectives. Teaching is a purposeful and meaningful process. A teacher has a delimited set of objectives. He should determine the teaching objectives. The educational objectives imply the changes that we try to produce in the child. In the words of B.S.Bloom, “Educational objectives are not only the goals towards which the curriculum is shaped and towards which instruction is guided, but they are also the goals that provide the detailed specification for the construction and use of evaluative techniques”.

1.2 Learning Objectives:

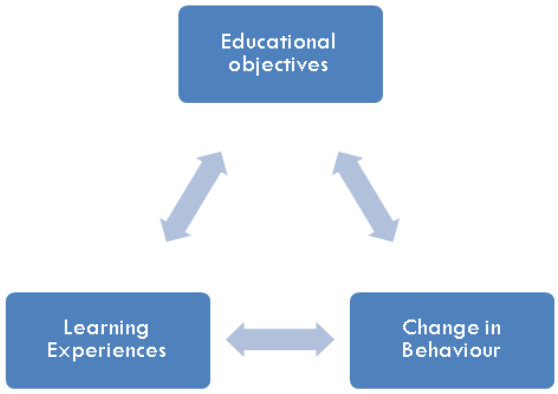
After going through this Unit, learners will be able to:

- Develop an understanding of taxonomy of objectives in the Cognitive domain.
- Develop an understanding of taxonomy of objectives in the Affective domain.
- Develop an understanding of taxonomy of objectives in the psychomotor domain.
- Explain bloom's taxonomy of educational objectives.

1.3 Taxonomy of Educational Objectives:

Taxonomy means a system of classification and in this sense taxonomy like Bloom's Taxonomy presents a system of classification of the objectives in the similar way as Dewey's decimal system tends to classify a number of books in a library.

The taxonomy, of educational objectives has been worked out on the assumption that the teaching-learning process may be conceived as an attempt to change the behavior of the pupils with respect to some subject matter or learning experiences. Behavior is divided into three domains- Cognitive (knowing), affective (feeling) and psychomotor (doing). The taxonomy of educational objectives has also been considered to be belonging to these three domains. Education is a tripolar process and the three poles are Educational Objectives, Learning Experiences and Change in Behaviour.



The learning experiences are provided by teaching activities to bring change in behavior which in turn is evaluated in terms of educational objectives. Thus educational objectives form the basis for teaching activities and evaluation techniques.

Bloom's Taxonomy of Educational Objectives:

One of the most widely used ways of organizing levels of expertise is according to Bloom's Taxonomy of Educational Objectives. Bloom's Taxonomy uses a multi-tiered scale to express the level of expertise required to achieve each measurable student outcome. Organizing measurable student outcomes in this way will allow us to select appropriate classroom assessment techniques for the course.

There are three taxonomies. Which of the three to use for a given measurable student outcome depends upon the original goal to which the measurable student outcome is connected. There are *knowledge-based goals, skills-based goals, and affective goals* (affective: values, attitudes, and interests); accordingly, there is a taxonomy for each. Within each taxonomy, levels of expertise are listed in order of increasing complexity. Measurable student outcomes that require the higher levels of expertise will require more sophisticated classroom assessment techniques.

To determine the level of expertise required for each measurable student outcome, first decide which of these three broad categories (knowledge-based, skills-based, and affective) the corresponding course goal belongs to. Then, using the appropriate Bloom's Taxonomy, look over the descriptions of the various levels of expertise. Determine which description most closely matches that measurable student outcome. Bloom's Taxonomy is a convenient way to describe the degree to which we want our students to understand and use concepts, to demonstrate particular skills, and to have their values, attitudes, and interests affected. It is critical that we determine the levels of student expertise that we are expecting our students to achieve because this will determine which classroom assessment techniques are most appropriate for the course.

Multiple-choice tests also rarely provide information about achievement of skills-based goals. Similarly, traditional course evaluations, a technique commonly used for affective assessment, do not generally provide useful information about changes in student values, attitudes, and interests.

Thus, commonly used assessment techniques, while perhaps providing a means for assigning grades, often do not provide us (or our students) with useful feedback for determining whether students are attaining our course goals. Usually, this is due to a combination of not having formalized goals to begin with, not having translated those goals into outcomes that are measurable, and not using assessment techniques capable of measuring expected student outcomes given the levels of expertise required to achieve them. Using the CIA model of course development, we can ensure that our curriculum, instructional methods, and classroom assessment techniques are properly aligned with course goals.

Note that Bloom's Taxonomy need not be applied exclusively after course goals have been defined. Indeed, Bloom's Taxonomy and the words associated with its different categories can help in the goals-defining process itself. Thus, Bloom's Taxonomy can be used in an iterative fashion to first state and then refine course goals. Bloom's Taxonomy can finally be used to identify which classroom assessment techniques are most appropriate for measuring these goals.

Self- Check Exercise-1

Q.1 Which level of Bloom's Taxonomy involves recalling facts, information, or concepts?

a) Remembering

b) Understanding

c) Applying

d) Analyzing

Q.2 When learners demonstrate comprehension of the material by explaining ideas or concepts in their own words, they are operating at which level of Bloom's Taxonomy?

a) Remembering
b) Understanding
c) Applying
d) Evaluating

Q.3 Applying learned information in new situations or solving problems using acquired knowledge corresponds to which level of Bloom's Taxonomy?

a) Remembering
b) Understanding
c) Applying
d) Analyzing

1.4 Taxonomy of Objectives in the Cognitive Domain:

Bloom and his associates in 1956 developed the cognitive domain on the basis of complexity of mental activities arranged from the lowest to the highest level of functioning as described below:

Bloom's Taxonomy of Educational Objectives for Knowledge-Based Goals		
Level of Expertise	**Description of Level**	**Example of Measurable Student Outcome**
1. Knowledge	Recall, or recognition of terms, ideas, procedure, theories, etc.	When is the first day of Spring?
2. Comprehension	Translate, interpret, extrapolate, but not see full implications or transfer to other situations, closer to literal translation.	What does the summer solstice represent?
3. Application	Apply abstractions, general principles, or methods to specific concrete situations.	What would Earth's seasons be like if its orbit was perfectly circular?
4. Analysis	Separation of a complex idea into its constituent parts and an understanding of organization and relationship between the parts. Includes realizing the distinction between hypothesis and fact as well as between relevant and extraneous variables.	Why are seasons reversed in the southern hemisphere?
5. Synthesis	Creative, mental construction of ideas and concepts from multiple sources to form	If the longest day of the year is in June, why is the northern

	complex ideas into a new, integrated, and meaningful pattern subject to given constraints.	hemisphere hottest in August?
6. Evaluation	To make a judgment of ideas or methods using external evidence or self-selected criteria substantiated by observations or informed rationalizations.	What would be the important variables for predicting seasons on a newly discovered planet?

Let us try to elaborate the above taxonomy of objectives of cognitive domain given by Bloom for its clarity and understanding

1. **Knowledge-**It represents the lowest level ofobjectives belonging to cognitive domain and primarily aims for the acquisition of the knowledge concerning.

i. Specific facts, terminology, methods and processes and

ii. Generalized principles, theories and structures.

2. **Comprehension-** comprehension is based upon the knowledge. If there is no knowledge, there will be no comprehension. On the ladder of the acquisition of cognitive abilities its level is little higher than the knowledge. Specifically, it means the basic understanding of the facts, ideas, methods, process, principles or theories, etc. as a result, what is communicated to learner, he may-

i. Translate or summarize the communicated knowledge in hisown words.

ii. Interpret i.e. cite examples, discriminate, classify, verify or generalize and

iii. Extrapolate i.e. understand by use of knowledge and extend it to other subjects and fields.

3. **Application**- The knowledge is useful only when it is possible to make it employed. The application of an idea, principle or theory may be made possible only when it is grasped and understood properly. Therefore, the category of application automatically involves both the earlier categories i.e. knowledge and comprehension. Under this objective the learner is required to acquire the ability to make use of abstract or generalized ideas, principles in the particular and concrete situations.

4. **Analysis**- analysis refers to an understanding at higher level. It is a complex cognitive process that involves knowledge, comprehension as well as application of an idea, fact, principles or theory. Though the realization of these objectives the learner is expected to acquire the necessary skills in drawing inferences, discriminating, making choices and selection and separating apart the different components or elements of a concept, object or principle.

5. **Synthesis**- The objective belonging to this category aim to help the learner to acquire necessary ability to combine the different elements or

components of an idea, object, concept, or principle as to produce an integrated picture i.e. a figure of wholeness. As a result he may be expected to propagate or present a theory or principle by combining different approaches, ideas or viewpoints.

6. **Evaluation**- Evaluation is the highest level of cognitive domain. It is defined as to take decision in regard to some aim, thought, solution, method, material, etc. The structure of human being is such that he cannot step himself to evaluate everything and take decision. Thus, the thought, decisions and things, which he thinks are beneficial, he evaluates them at higher level and those which are beneficiary are evaluated at lower level. The external and internal evidences are taken into consideration while taking the decision.

Example: Let's take a concept like photosynthesis and create examples of learning objectives aligned with each level of Bloom's Taxonomy in the cognitive domain:

1. **Remembering**:

Define photosynthesis.
List the raw materials required for photosynthesis.
Recall the chemical equation for photosynthesis.

2. **Understanding**:

Explain the process of photosynthesis in your own words.
Summarize the role of chlorophyll in photosynthesis.
Describe how light intensity affects the rate of photosynthesis.

3. **Applying**:

Predict how a decrease in carbon dioxide concentration would affect the rate of photosynthesis.
Design an experiment to investigate the effect of different colored lights on photosynthesis.
Calculate the amount of glucose produced during photosynthesis given a certain amount of carbon dioxide and water.

4. **Analyzing**:

Compare and contrast photosynthesis and cellular respiration.
Analyze the factors that limit the rate of photosynthesis in plants.
Examine the adaptations of plants in different environments related to photosynthesis.

5. **Evaluating**:

Evaluate the importance of photosynthesis in the ecosystem.
Assess the effectiveness of different methods for increasing crop yield through manipulation of photosynthesis.
Critique a scientific study investigating the effects of pollutants on photosynthesis.

6. **Creating**:

Design a model illustrating the process of photosynthesis, including all key components and their interactions.

Develop a multimedia presentation explaining the importance of photosynthesis to a non-scientific audience.
Propose a solution for improving photosynthetic efficiency in crops to address food security issues.

Self- Check Exercise-2

Q.1 Which level of Bloom's Taxonomy involves breaking down information into its component parts and understanding the relationships between them?
a) Understanding
b) Applying
c) Analyzing
d) Evaluating

Q2. Making judgments based on criteria and standards, such as assessing the value or effectiveness of ideas or solutions, is associated with which level of Bloom's Taxonomy?
a) Analyzing
b) Evaluating
c) Creating
d) Remembering

Q3. Synthesizing information to generate new ideas, products, or ways of thinking is characteristic of which level of Bloom's Taxonomy?
a) Analyzing
b) Evaluating
c) Creating
d) Remembering

1.5 Taxonomy of Objectives in the Affective Domain:

Affective domain can be categorized into the following categories. These categories are given in sequence of development and they can be arranged on a continuum from lowest to highest level.

Bloom's Taxonomy of Educational Objectives for Affective Goals		
Level of Expertise	**Description of Level**	**Example of Measurable Student Outcome**
1.Receiving	Demonstrates a willingness to participate in the activity	When I'm in class I am attentive to the instructor, take notes, etc. I do not read the newspaper instead.
2.Responding	Shows interest in the objects, phenomena, or activity by seeking it out or pursuing it for pleasure	I complete my homework and participate in class discussions.

3.Valuing	Internalizes an appreciation for (values) the objectives, phenomena, or activity	I seek out information in popular media related to my class.
4.Organization	Begins to compare different values, and resolves conflicts between them to form an internally consistent system of values	Some of the ideas I've learned in my class differ from my previous beliefs. How do I resolve this?
5.Characterization by a Value or Value Complex	Adopts a long-term value system that is "pervasive, consistent, and predictable"	I've decided to take my family on a vacation to visit some of the places I learned about in my class.

Let us elaborate the above classification for its better understanding:

1. Receiving-It represents the initial category for the objectives belonging to affective domain. For the inculcation of certain interest, attitudes, values or ideas it is essential that learner should be made to receive or attend the desired ideas events or objectives. This category points out towards this necessity and takes into consideration three types of following sequential activities:

i. Firstly, the learner is sensitized or made aware about the existence of certain stimuli.

ii. Then the desired intension or willingness for receiving for receiving or attending the stimuli is created in the learner.

iii. Lastly, the efforts are made for the control of the attention of the learner. He may be trained to pay selective attention and sustain it for a desired period.

2. Responding-Once the learner receives or attends to a particular ideas, event or thing he must be made to respond to it as actively as possible. The responses here do not confine itself in just paying attention or arousal of a simple intention or desire of getting a thing, as in the first category or receiving but manifest themselves in the active behavior like obeying, answering, reading, discussing, recording, writing and reaching to a stimulus, etc.

3. Valuing-When one attends as well as responds to a particular thing, idea or event he is naturally drifted towards taking value judgment about that thing; idea or event. Therefore, this category of valuing depends upon both the former categories i.e. receiving and responding. Here the learner is expected to imbibe a definite value pattern towards different ideas, events and objects. In practice the objectives belonging to this category are usually concerned with the development of typical value patterns attitudes, etc.

4. Organizing-The category of objectives concern with the construction of relatively enduring value structure in the learner by organizing and synthesizing the different value patterns imbibed by him from time to time. Ultimately this category of objective leads the learner to form a set value structure or philosophy of life.

5. Characterizing by a value or value complex-It is the highest level category of objectives of affective domain. At this stage, the learner is destined to imbibe typical characteristics of his individual character i.e. life style of his own. In fact it is end point or ultimate goal of the process of education.

Example: Let's use the topic of environmental conservation to create examples of learning objectives aligned with each level of Bloom's Taxonomy in the affective domain:

1. **Receiving (Receiving the message)**:

Listen attentively during a presentation on the importance of environmental conservation.

Read an article about the impact of deforestation without interruption.

2. **Responding (Responding to the message)**:

Participate actively in a class discussion about different environmental issues.

Express agreement or disagreement with a peer's viewpoint on sustainable living practices.

3. **Valuing (Valuing the message)**:

Express appreciation for the beauty of nature during a field trip to a national park.

Demonstrate a commitment to recycling by consistently separating recyclables from regular waste.

4. **Organizing (Organizing values into a system)**:

Create a personal action plan outlining specific steps to reduce carbon footprint at home and in the community.

Develop a presentation advocating for the implementation of renewable energy sources in the local area.

5. **Characterizing (Internalizing values and integrating them into one's behavior)**:

Act as a leader in organizing a community cleanup event to promote environmental stewardship.

Serve as a mentor to educate others about the importance of biodiversity conservation and sustainable living practices.

Self- Check Exercise-3

Q.1 Which level of the affective domain involves actively engaging with and responding to stimuli or messages?

a) Receiving

b) Responding

c) Valuing

d) Organization

Q.2 When a person demonstrates an appreciation for diversity and cultural differences, they are operating at which level of the affective domain?

a) Receiving
b) Responding
c) Valuing
d) Organization

Q.3 Which level of the affective domain involves adopting a particular belief or attitude as one's own?

a) Receiving
b) Responding
c) Valuing
d) Organization

1.6 TAXONOMY OF OBJECTIVES IN THE PSYCHOMOTOR DOMAIN:

The classification of psychomotor objectives was first produced by Simpson (1966) and later modified by Harrow (1972). These given by harrow are being described under six different categories arranged from lowest to the highest level of functioning:

1. Reflex movements- Reflex movements may be considered as the involuntary motor responses to the various stimuli in the environment. Examples of such reflex movements or actions are: the jerking of hands, the closing of eye lid, stretching of the arms, etc. these movements represents the lowest level of the psychomotor behavior. They are largely controlled by the autonomous nervous system. However, they are very much essenti9al not only for the development of psychomotor abilities but also for the survival of the human beings.

2. Basic Fundamental Movements- These fundamental movements are just a step ahead of the simple reflex movements. They are not as inborn and innate as the reflex movements but a child may be seen to demonstrate such movements in his very early days of life. Their movements in the form of kneeling, creeping, stumbling, walking, jumping, moving hands, neck, head, etc. may be named as basic fundamental movements. They represent the simple basic movements of the body almost requiring no serious attempts or skilled practice for their occurrence.

3. Perceptual Abilities- the development of motor abilities related with the phenomenon of perception belongs to this category of objectives. When some meaning is attached to sensation, it is termed as perception. As a result, the learner is able to derive useful meaning out of the exposure of their senses to various stimuli in the environment. Whatever is perceived by him through his senses becomes an ignition point for the motor behavior. Such type of behavior is a learned behavior. It is always acquired through experience and systematic training.

4. Physical Abilities- there is an urgent need of the development of desirable physical abilities for an effective motor behavior. If one has adequate physical stamina and abilities, he may go ahead in the task of improving his psychomotor behavior. Therefore this category of objectives aims to develop the various physical abilities of the learner like tolerance to bear and stand against rough weather; to do hard labor, to carry the large load, to bend an article, to demonstrate one's physical power in starting, stopping or running an object or machine, etc.

5. Skilled movements- Skilled movements are those complex bodily movements which help in performing the skilled tasks. These movements are to be acquired through an organized and systematic learning process. Their acquisition requires an intelligent understanding and sufficient drill or practice work on the part of the learner. The art of dancing, diving, playing the musical organs, skating, typing, swimming, tailoring, etc. represent such skilled movements. The development of the abilities concerning such skilled movements depends upon the development of the motor abilities described under all the earlier four categories.

6. Non Discussive Communication- this category represents the highest level of the psychomotor behavior. The bodily movements are hereby integrated with the inner feeling and affective behavior of the learner. In this way the non-discussive communication may be defined in terms of the overt behavior activities related with the communication of affective behavior feelings or emotions. This communication may range from a simple behavior express able through posing or facial expression to a complex behavior performed through a highly sophisticated classical dance, sketching, painting or acting.

Bloom's Taxonomy of Educational Objectives for Skills-Based Goals		
Level of Expertise	**Description of Level**	**Example of MeasurableStudent Outcome**
Perception	Uses sensory cues to guide actions	Some of the colored samples you see will need dilution before you take their spectra. Using only observation, how will you decide which solutions might need to be diluted?
Set	Demonstrates a readiness to take action to perform the task or objective	Describe how you would go about taking the absorbance spectra of a sample of pigments?
Guided Response	Knows steps required to complete the task or objective	Determine the density of a group of sample metals with regular and irregular shapes.
Mechanism	Performs task or	Using the procedure described below,

	objective in a somewhat confident, proficient, and habitual manner	determine the quantity of copper in your unknown ore. Report its mean value and standard deviation.
Complex Overt Response	Performs task or objective in a confident, proficient, and habitual manner	Use titration to determine the Ka for an unknown weak acid.
Adaptation	Performs task or objective as above, but can also modify actions to account for new or problematic situations	You are performing titrations on a series of unknown acids and find a variety of problems with the resulting curves, e.g., only 3.0 ml of base is required for one acid while 75.0 ml is required in another. What can you do to get valid data for all the unknown acids?
Organization	Creates new tasks or objectives incorporating learned ones	Recall your plating and etching experiences with an aluminum substrate. Choose a different metal substrate and design a process to plate, mask, and etch so that a pattern of 4 different metals is created.

Example: Let's use the skill of swimming to create examples of learning objectives aligned with each level of the taxonomy of objectives in the psychomotor domain:

1. **Perception (Awareness of the skill)**:
- Identify the different parts of a swimming pool and their functions.
- Recognize the safety rules associated with swimming.

2. **Set (Readiness to perform the skill)**:
- Demonstrate the ability to wear appropriate swimming attire and gear.
- Express willingness to learn and practice swimming strokes.

3. **Guided Response (Imitation of the skill)**:
- Mimic the arm movements of the instructor during a demonstration of the freestyle stroke.
- Follow step-by-step instructions to perform a proper forward dive into the water.

4. **Mechanism (Basic proficiency in the skill)**:
- Perform the breaststroke with coordinated arm and leg movements.
- Execute a shallow dive from the edge of the pool with proper body position and entry technique.

5. **Complex Overt Response (Skillful performance of the skill)**:

- Swim multiple laps using various strokes (e.g., freestyle, backstroke, butterfly) with proficient technique.
- Perform a flip turn at the end of each lap while maintaining speed and efficiency.

6. **Adaptation (Ability to modify the skill as needed)**:

- Adjust swimming technique based on water conditions (e.g., currents, waves) to maintain control and efficiency.
- Modify stroke mechanics to conserve energy during long-distance swimming.

7. **Origination (Creation of new movements or skills)**:

- Develop a synchronized swimming routine incorporating creative movements and formations.
- Invent a new swimming stroke or technique aimed at improving speed or reducing drag.

Self- Check exercise-4

Q.1 Which level of the psychomotor domain involves basic proficiency in performing a skill?
a) Perception
b) Set
c) Mechanism
d) Complex Overt Response

Q.2 When a learner imitates the movements of an instructor during a demonstration, they are operating at which level of the psychomotor domain?
a) Perception
b) Guided Response
c) Mechanism
d) Adaptation

Q.3 Which level of the psychomotor domain requires skillful performance of the task?
a) Set
b) Mechanism
c) Complex Overt Response
d) Adaptation

1.8 Summary:

Bloom's taxonomy serves as the backbone of many teaching philosophies, in particular those that lean more towards skills rather than content. These educators would view content as a vessel for teaching skills. The emphasis on higher-order thinking inherent in such philosophies is based on the top levels of the taxonomy including analysis, evaluation, synthesis and creation. Bloom's taxonomy can be used as a teaching tool to help balance

assessment and evaluative questions in class, assignments and texts to ensure all orders of thinking are exercised in student's learning, including aspects of information searching.

1.9 Glossary:

Educational Objectives: Statements that describe the intended learning outcomes of an educational program, course, or lesson, typically based on Bloom's Taxonomy and specifying what students should know or be able to do.

Learning Outcomes: Statements that describe what students are expected to know, understand, or be able to do as a result of instruction, typically based on educational objectives and aligned with Bloom's Taxonomy.

Cognitive Domain: Refers to the domain of Bloom's Taxonomy that encompasses intellectual skills and abilities related to thinking, understanding, and problem-solving.

Affective Domain: Refers to the domain of Bloom's Taxonomy that encompasses attitudes, beliefs, values, and emotions, influencing students' motivation, engagement, and behavior.

Psychomotor Domain: Refers to the domain of Bloom's Taxonomy that encompasses physical skills and abilities related to movement, coordination, and manual dexterity.

1.10 Answers to Self-Check Exercises

Self Check exercise-1

1. a) Remembering
2. b) Understanding
3. c) Applying

Self Check exercise-2

1 c) Analyzing
2 b) Evaluating
3 c) Creating

Self Check exercise-3

1. b) Responding
2. c) Valuing
3. c) Valuing

Self Check exercise-4

1. c) Mechanism
2. b) Guided Response
3. c) Complex Overt Response

1.11 References/ Suggested Readings:

- Ebel, Robert L.(1966) "Measuring Educational Achievement, Prentice Hall of India Pvt. Ltd. Pp. 481

- Gronlund, N. E. (1976), Measurement and Evaluation in Teaching. McMillan, USA.
- Hopkins, C.D. and Antes, R.L. (1990).Classroom measurement and evaluation. Itasca, Illinois: Peacock.
- Izard, J. (1991). Assessment of learning in the classroom. Geelong, Vic.: Deakin University.
- Izard, J. (1997). Content Analysis and Test Blueprints. Paris: International Institute for Educational Planning.
- Mehrens, W.A. and Lehmann, I.J. (1984).Measurement and evaluation in education and psychology.(3rd Ed.) New York: Holt, Rinehart and Winston.
- Nandra, I.D.S.(2011). Learning Resources and Assessment of Learning.Patiala, 21st Century Publications.
- Taiwo, Adediran A. (2005). Fundamentals of Classroom Testing. New Delhi: Vikas Publishing House Pvt. Ltd.
- Walter W. Cook (1958). Educational Measurement. Washington D.C.: American Council on Education.
- Withers, G. (1997).Item Writing for Tests and Examinations. Paris: International Institute for Educational Planning.

1.12 Terminal Questions:

Q1. Discuss briefly the taxonomy of objectives in the cognitive domain..

Q2. Describe the of taxonomy of objectives in the Affective domain

Q3. Discuss the taxonomy of objectives in the Psychomotor domain.

UNIT-2

EDUCATIONAL MEASUREMENT: CONCEPT, NEED AND SCOPE

Structure:

2.1 Introduction:

The term measurement and evaluation are often used in Psychology and education. Measurement and evaluation are the very old processes which are not only used in behavioural science but it is an origin of physical sciences and arithmetic. The development of measurement goes side by side the human development. Measurement may be understood as the comparison of a quantity with an appropriate scale for the purpose of determining the numerical value on the scale that corresponds to the quantity to be measured. Measurement is the process of systematically assigning numbers to objects and their properties, to facilitate the use of mathematics in studying and describing objects and their relationships. Some types of measurement are fairly concrete: for instance, measuring a person's weight in pounds or kilograms, or their height in feet and inches or in meters. Note that the particular system of measurement used is not as important as a consistent set of rules: we can easily convert measurement in kilograms to pounds, for instance. Although any system of units may seem arbitrary (try defending feet and inches to someone who grew up with the metric system!), as long as the system has a consistent relationship with the property being measured, we can use the results in calculations.

2.2 Learning Objectives:

After going through this Unit, learners will be able to:

- Develop an understanding of educational measurement.
- Develop an understanding of need, purpose and scope of educational measurement.
- Explain the functions of educational measurement.

2.3 Concept of Educational Measurement:

Educational measurement refers to the use of educational assessments and the analysis of data such as scores obtained from educational assessments to infer the abilities and proficiencies of students. Educational measurement is the assigning of numerals to traits such as achievement, interest, attitudes, aptitudes, intelligence and performance.

The aim of theory and practice in educational measurement is typically to measure abilities and levels of attainment by students in areas such as reading, writing, mathematics, science and so forth. Traditionally, attention focuses on whether assessments are reliable and valid. In practice, educational measurement is largely concerned with the analysis of data from educational assessments or tests. Typically, this means using total scores on assessments, whether they are multiple choice or open-ended and marked using marking rubrics or guides.

Let us make the meaning of term measurement more clear with the help of definitions given below:

Carter V Good. "Measurement may be understood as the comparison of a quantity with an appropriate scale for the purpose of determining the numerical value on the scale that corresponds to the quantity to be measured."

Remmers, Gaze and Rummel. "Measurement refers to observations that can be expressed quantitatively and answers the question, how much."

Mahesh Bhargava."Measurement is the process of assigning symbols or numerical to observations, objects or events in some meaningful or consistent manner according to rule."

The analysis of above definition may clearly reveal that measurement is nothing but a process of quantification i.e. assigning units of measurements or numerical values to the types of characteristics observed in the behavior or nature of an individual or object during some observation or testing.

Measurement in education refers to the process of assessing students' knowledge, skills, abilities, and other characteristics relevant to learning and academic achievement. It involves the systematic collection of data to evaluate students' progress, diagnose areas of strength and weakness, inform instructional decision-making, and determine the effectiveness of educational programs. Here are some key concepts related to measurement in education:

1. **Assessment**: Assessment is the process of gathering information about students' performance, typically through various methods such as tests, quizzes, projects, observations, and portfolios. Assessment can be formative (ongoing and used to provide feedback for improvement) or summative (evaluative, used to make judgments about student achievement).
2. **Validity**: Validity refers to the extent to which an assessment measures what it is intended to measure. A valid assessment accurately reflects the knowledge, skills, or attributes it is designed to assess.

3. **Reliability**: Reliability refers to the consistency and stability of assessment results. A reliable assessment produces consistent scores when administered repeatedly under similar conditions.
4. **Norm-Referenced vs. Criterion-Referenced**: Norm-referenced assessment compares a student's performance to the performance of a larger group (norm group), whereas criterion-referenced assessment measures a student's performance against a predetermined set of criteria or standards.
5. **Standardized Testing**: Standardized tests are assessments administered and scored under uniform conditions, with established procedures for administration and scoring. They are often used for large-scale assessments to compare students' performance across schools, districts, or regions.
6. **Formative Assessment**: Formative assessment occurs throughout the learning process to provide feedback for improvement. It helps teachers identify students' strengths and weaknesses and adjust instruction accordingly.
7. **Summative Assessment**: Summative assessment occurs at the end of a learning period to evaluate students' overall achievement. It is typically used for grading, promotion, or certification purposes.
8. **Rubrics**: Rubrics are scoring guides that outline criteria for evaluating students' performance on tasks or assignments. They provide clear expectations and criteria for assessment, facilitating consistent and fair evaluation.
9. **Authentic Assessment**: Authentic assessment tasks mirror real-world situations and require students to apply knowledge and skills in meaningful contexts. Examples include projects, performances, and portfolios.
10. **Data-Informed Decision Making**: Measurement data are used to inform instructional decisions, curriculum development, and educational policy. By analyzing assessment results, educators can identify areas for improvement, monitor progress, and make evidence-based decisions to enhance student learning outcomes.

Measurement in education plays a crucial role in promoting accountability, ensuring educational equity, and supporting continuous improvement in teaching and learning practices. Effective measurement practices help educators tailor instruction to meet students' diverse needs, foster academic growth, and prepare students for success in school and beyond.

Self- Check Exercise-1

Q.1 _____ refers to the process of assessing students' knowledge, skills, and abilities.

Q.2 _____ is the extent to which an assessment measures what it is intended to measure.

Q.3 Standardized tests are administered and scored under _____ conditions.

Q.4 _____ assessment occurs throughout the learning process to provide feedback for improvement.

Q.5 Rubrics provide _____ for evaluating students' performance on tasks or assignments.

2.4 Need, Purpose and Scope of Educational Measurement:

To explain the need of measurement, three assumptions must be made.

First, the schools exist in order to accomplish certain aims and these aims can be expressed in terms of desired changes in pupil behavior.

Second, instructional programs in schools are formulated in order to accomplish these objectives.

Third, objectives or aims are not likely to be accomplished successfully unless provision is made for continuing evaluation of the instructional programs.

Measurement, therefore, looks to be essential if evaluative process is to be accurately and effectively carried out. Measurement can be useful not only in evaluating a total program of instruction but also in providing information concerning the progress and development of the individual pupil. More specifically as indicated by Lindeman and Merenda (1979), it can answer the questions such as given below:

1. What are the characteristics of pupils at the time they enter the system (to know the status).
2. Considering the general ability and aptitudes of the pupils in a given school system, how does their achievement in various subject-matter areas compare with that of students of similar ability and aptitude in other school system (to compare the statuses).
3. To what extent are the instructional objectives of the school and the individual classroom teacher being achieved through the instructional processes and methods employed (to assess the efficacy of teaching methods in the light of instructional objectives).
4. Which children entering the school system require a specialized instruction in order to take the fullest advantage of their exceptional ability or to deal effectively with special learning problems? Which special instructional processes and methods and what special programmes must be developed for achieving maximum individualization of instruction (to match instruction with student ability and to provide remedial teaching).
5. What advice should be given to individual students as they develop educational and vocational plans for the future (to provide educational and vocational guidance).
6. How can students are helped to develop realistic self-images so that they will be able to formulate goals that are consistent with their aptitude and abilities.
7. How can new students be properly [placed so that their instruction will be consistent prior learning and with their aptitude and ability?

8. How can information concerning the characteristics of individual pupils be made available in suitable form to outside agencies such s colleges, universities and prospective employers?
9. How can information concerning school programmes policies and objectives be best gathered and conveyed to parents, community and the school management or the government?
10. When more than one method of instruction is available which one tends to be most effective in maximizing pupil achievement?

Questions such as above may be adequately answered by measurement. Prof. A.K. Singh (1986) has enlisted the following functions of measurement:

- Selection of students and school personnel.
- Classification of students and teachers into various categories.
- Comparison of students, classes, programmers and methods of teaching etc.
- Guidance and counseling, both to pupils and teachers.
- Research.
- Improving class-room instruction.

Purpose of Educational Measurement:

1. **Assessing Student Learning**: The primary purpose of educational measurement is to evaluate students' knowledge, skills, abilities, and other attributes relevant to learning and academic achievement. This assessment helps educators understand students' strengths and weaknesses, identify areas for improvement, and tailor instruction to meet individual needs.
2. **Monitoring Progress**: Educational measurement provides a means for monitoring students' progress over time. By assessing students at regular intervals, educators can track their growth, identify learning trends, and intervene when necessary to support struggling students or challenge high achievers.
3. **Informing Instructional Decision-Making**: Measurement data inform instructional decision-making by providing valuable insights into students' learning needs and preferences. Educators use assessment results to design and adjust instructional strategies, differentiate instruction, and provide targeted support to address students' diverse learning styles and abilities.
4. **Evaluating Educational Programs**: Educational measurement helps evaluate the effectiveness of educational programs, curriculum, and instructional interventions. By assessing students' learning outcomes, educators and policymakers can determine whether educational initiatives are achieving their intended goals and make evidence-based decisions to improve program quality and effectiveness.
5. **Promoting Accountability**: Educational measurement promotes accountability by holding educators, schools, and educational systems accountable for student learning outcomes. Assessment data are used to assess the performance of educational institutions, inform stakeholders about

progress and challenges, and guide resource allocation and policy development to improve educational outcomes for all students.

Scope of Educational Measurement:

1. **Formative Assessment**: Formative assessment occurs during the learning process to provide ongoing feedback for improvement. It focuses on identifying students' strengths and weaknesses, monitoring progress, and guiding instructional adjustments to enhance learning outcomes.
2. **Summative Assessment**: Summative assessment occurs at the end of a learning period to evaluate students' overall achievement. It typically involves assessing students against predetermined standards or criteria and making judgments about their proficiency or readiness for advancement.
3. **Standardized Testing**: Standardized tests are used for large-scale assessments to compare students' performance across schools, districts, or regions. These tests are administered and scored under uniform conditions, allowing for consistent evaluation and benchmarking of student achievement.
4. **Authentic Assessment**: Authentic assessment tasks mirror real-world situations and require students to apply knowledge and skills in meaningful contexts. Examples include projects, performances, portfolios, and simulations, which assess students' ability to transfer learning to authentic situations and solve real-world problems.
5. **Individualized Assessment**: Individualized assessment involves tailoring assessment strategies to meet the unique needs and characteristics of individual students. It may include alternative assessments, accommodations, and modifications to ensure equitable access to assessment opportunities for all learners.

Overall, the purpose and scope of educational measurement encompass a wide range of assessment practices aimed at promoting student learning, informing instructional decision-making, evaluating program effectiveness, and fostering accountability in educational settings.

Self- Check Exercise-2

Q.1 Which of the following best describes the need for educational measurement?

a) To rank students based on their intelligence.
b) To evaluate students' learning and academic achievement.
c) To enforce strict discipline in schools.
d) To create competition among students.

Q.2 Which aspect falls within the scope of educational measurement?

a) Evaluating teachers' performance.
b) Assessing students' progress over time.
c) Deciding school holidays.
d) Promoting extracurricular activities.

Q.3 What is the primary purpose of educational measurement?

a) To create standardized tests for all students.

b) To evaluate the effectiveness of educational programs.

c) To increase workload for teachers.

d) To discourage students from learning.

2.5 Functions of Educational Measurement:

Findly (1963) has classified the purposes served by tests and measurement under three inter-related categories:

a. Instructional functions,
b. Administrative functions, and
c. Guidance functions.

a. Instructional functions:

One great function of measurement and testing is the improvement of instruction in the class-room. The programme of measurement serves this function in the following manner:

1. Measurement stimulus teachers to clarify and refine meaningful course objectives:

Participation of teaching staff in selecting as well as constructing measuring tools has resulted in improved instruments on one hand and on the other hand, it has resulted in clarifying objectives of instruction and in making them real and meaningful. Dr. Benjamin S. Bloom observed that when teachers have actively participated in defining objectives and construction of evaluating tools, they return to the learning problems with great vigor and remarkable creativity, their teaching is greatly improved.

2. Measurement provides a means of feedback to the teacher:

Feedback from measurement helps the teacher provide more appropriate instructional guidance for individual students as well as for the class as a whole. Well-designed tests may also be of value for pupil self-diagnosis since they help students' identity areas of specific weaknesses.

3. Measurement motivates Learning and teaching:

As a general rule, students pursue their studies more diligently if they expect to be evaluated. In the intense competition for a student's time course without examinations are often squeezed out of priority and usually ignored by students and teachers alike. When queried students have consistently reported greater study and learning. The anticipation of a forthcoming test affects pupil's intention to remember instructional content.

4. Measurement influences Retention Positively:

Kruger's classical study shows that examination not only stimulate review (learning and over-learning), but also positively influence retention performance.

b. Administrative Functions:

Apart from instructional advantages of measurement and examination, there are quite a few administrative benefits accruing from it. Some of such functions are summarized below:

1. Measurement Provides a Mechanism for Quality Control for a school or school system:

Measurement provides local, state or national norms which form dependable basis for assessing certain curricular strengths and weaknesses. In the absence of such norms, instructional inadequacies maybe go unnoticed and school system, though deficient, may feel contented with what is going therein.

2. Measurement is useful for research in education and psychology:

Measurement is the back-bone of all educational research. Tests provide useful data for deciding which innovative programmes are better or poorer than the conventional ones in facilitating the attainment of specific goals. Research in the process of learning and teaching employing different models and strategies very largely depends upon objectives and comprehensive measurement and testing.

3. Measurement enables better decisions on classification and placement:

Grouping children by their ability levels is an example of classification for which tests can be of immense value. Educational and vocational placements are also facilitated by measurement and testing.

4. Measurement increases the quality of selection decisions:

Scholastic aptitude and achievement test scores have repeatedly demonstrated their value in identifying who are or are not likely to succeed in various classes, or schools or colleges or programmes. Certain jobs require special skills that are best assessed by well-designed tests. Tests are the primary criteria for identifying the gifted or retarded children, or the recipients of various awards, prizes and distinctions.

5. Measurement can be useful means of accreditation, mastery or certification:

Tests on which standards of performance has been established allow the demonstration of competence or knowledge that may have been acquired in an unconventional way. The examinee may thereby receive some deserved credit, or certificate or authorization. Such certification serves a useful purpose for further admissions, selections or jobs.

c. Guidance Functions:

The third major function of measurement or testing is in the field of educational and vocational guidance. Some important advantages are as follows:

1. Tests can be of value in diagnosing an individual's special aptitudes and abilities:

Obtaining measures of scholastic aptitude, achievement, interest and personality is often an important aspect of counseling process. The use of information from standardized test and inventories can be helpful for guiding the selection of a school or college, the choosing of an appropriate course of study, discovering unrecognized abilities, and so on. Fickle and Millman (1957) have remarked that:

“when a student because of proper use of test results, is well adjusted and challenged in his school classes, happy with his curriculum and aware of his abilities and interests with respect to his educational and vocational future, then not only he himself benefits, but in the long run teachers, counselors, school administrators, college personnel employers and other benefit.”

2. Tests and examination provide measurements upon which school decisions are based.

Instructionally tests provide feedback, motivation and retention. Administratively tests facilitate quality control, programme evaluation, research, classification, comparison, placement, selection, accreditation, mastery and certification. In guidance, measurement serves to diagnose special aptitudes or abilities and thus facilitate course choices remedial teaching and career preparation.

To Summarize, the functions of educational measurement are diverse and integral to the educational process. Here's an overview of its key functions:

1. **Assessing Learning**: One of the primary functions of educational measurement is to assess students' learning and academic achievement. Through various assessment methods such as tests, quizzes, projects, and presentations, educators gauge students' understanding of concepts, mastery of skills, and application of knowledge.
2. **Monitoring Progress**: Educational measurement enables educators to monitor students' progress over time. By assessing students at regular intervals, educators can track their growth, identify areas for improvement, and provide timely interventions or support to ensure continued progress.
3. **Informing Instructional Decision-Making**: Measurement data provide valuable insights into students' learning needs, preferences, and challenges, which inform instructional decision-making. Educators use assessment results to design and adjust instructional strategies, differentiate instruction, and provide targeted support to meet students' diverse learning styles and abilities.
4. **Evaluating Educational Programs**: Educational measurement helps evaluate the effectiveness of educational programs, curriculum, and instructional interventions. By assessing students' learning outcomes, educators and policymakers can determine whether educational initiatives are achieving their intended goals and make evidence-based decisions to improve program quality and effectiveness.
5. **Promoting Accountability**: Measurement data are used to hold educators, schools, and educational systems accountable for student learning outcomes. Assessment results are used to assess the performance of educational institutions, inform stakeholders about progress and challenges, and guide resource allocation and policy development to improve educational outcomes for all students.
6. **Facilitating Differentiation**: Educational measurement supports differentiation by providing information about students' individual learning needs, strengths, and weaknesses. Educators can use assessment data to

tailor instruction, provide additional support or enrichment opportunities, and personalize learning experiences to meet the diverse needs of students.

7. **Supporting Educational Planning and Policy Development**: Measurement data inform educational planning and policy development at various levels, from classroom instruction to district-wide initiatives and national education policies. By analyzing assessment results, policymakers can identify areas for improvement, allocate resources effectively, and develop evidence-based policies to enhance educational quality and equity.

8. **Promoting Continuous Improvement**: Educational measurement fosters a culture of continuous improvement by providing feedback on students' progress, instructional effectiveness, and program outcomes. Educators use assessment data to identify areas for growth, set goals for improvement, and implement targeted interventions or initiatives to enhance teaching and learning practices.

Overall, the functions of educational measurement are multifaceted, encompassing assessment for learning, accountability, program evaluation, instructional improvement, and educational policy development. By providing valuable information about student learning and performance, measurement plays a crucial role in promoting educational excellence, equity, and innovation.

Self- Check exercise -3

Q.1 One of the primary functions of measurement in education is to assess students' learning and academic _____.

Q.2 Measurement data provide valuable insights into students' learning needs, preferences, and challenges, which inform _____ decision-making.

Q.3 Educational measurement helps evaluate the effectiveness of educational programs, curriculum, and instructional interventions by assessing students' learning _____.

2.6 Summary:

Measurement in education is a vital process aimed at assessing students' knowledge, skills, abilities, and other attributes relevant to learning and academic achievement. Its primary purpose is to evaluate student learning, monitor progress, inform instructional decision-making, evaluate program effectiveness, and promote accountability. Educational measurement encompasses various assessment practices, including formative assessment, which provides ongoing feedback for improvement, and summative assessment, which evaluates overall achievement. Standardized testing is often used for large-scale assessments, while authentic assessment tasks mirror real-world situations to assess students' ability to apply knowledge and skills authentically. Measurement data help educators understand students' learning needs, identify areas for improvement, tailor instruction, and support

struggling students. Overall, measurement in education plays a crucial role in promoting student learning, fostering continuous improvement, and ensuring accountability in educational settings.

2.7 Glossary:

- **Assessment**: The process of gathering information about students' knowledge, skills, abilities, and other attributes relevant to learning and academic achievement.
- **Validity**: The extent to which an assessment measures what it is intended to measure.
- **Reliability**: The consistency and stability of assessment results, indicating the degree to which the assessment produces consistent scores under similar conditions.
- **Formative Assessment**: Assessment conducted during the learning process to provide ongoing feedback for improvement and guide instructional decision-making.
- **Summative Assessment**: Assessment conducted at the end of a learning period to evaluate students' overall achievement and make judgments about proficiency or readiness for advancement.

2.8 Answers to Self- Check Exercises:

Self- check Exercise-1

Answer1. Assessment
Answer2. Validity
Answer3. Uniform
Answer4. Formative
Answer5. Criteria

Self- check Exercise-2

Answer1: b) To evaluate students' learning and academic achievement.
Answer2: b) Assessing students' progress over time.
Answer3: b) To evaluate the effectiveness of educational programs.

Self- check Exercise-3

Answer 1: Achievement
Answer 2: Instructional
Answer 3: Outcomes

2.9 References/ suggested Readings:

- Ebel, Robert L.(1966) “Measuring Educational Achievement, Prentice Hall of India Pvt. Ltd. Pp. 481
- Gronlund, N. E. (1976), Measurement and Evaluation in Teaching. McMillan, USA.
- Hopkins, C.D. and Antes, R.L. (1990).Classroom measurement and evaluation. Itasca, Illinois: Peacock.

- Izard, J. (1991). Assessment of learning in the classroom. Geelong, Vic.: Deakin University.
- Izard, J. (1997). Content Analysis and Test Blueprints. Paris: International Institute for Educational Planning.
- Mehrens, W.A. and Lehmann, I.J. (1984).Measurement and evaluation in education and psychology.(3rd Ed.) New York: Holt, Rinehart and Winston.
- Nandra, I.D.S.(2011). Learning Resources and Assessment of Learning.Patiala, 21st Century Publications.
- Taiwo, Adediran A. (2005). Fundamentals of Classroom Testing. New Delhi: Vikas Publishing House Pvt. Ltd.
- Walter W. Cook (1958). Educational Measurement. Washington D.C.: American Council on Education.
- Withers, G. (1997).Item Writing for Tests and Examinations. Paris: International Institute for Educational Planning.

2.10 Terminal Questions:

Q.1 What is the primary function of measurement in education?

Q.2 How does measurement in education support instructional decision-making?

Q.3 Explain the role of measurement in evaluating educational programs.

Q.4 How does measurement contribute to promoting accountability in education?

UNIT-3

CRITERION AND NORM REFERENCED MEASUREMENT

Structure:

3.1 Introduction:

Measurement/Assessment in education refers to the process of gathering and analyzing information about students' knowledge, skills, abilities, and other characteristics relevant to learning and academic achievement. It plays a crucial role in evaluating students' progress, diagnosing areas of strength and weakness, informing instructional decision-making, and measuring the effectiveness of educational programs. Assessment in education is a multifaceted process that encompasses various methods, purposes, and considerations. When used effectively, assessment promotes student learning, informs instructional decision-making, and contributes to educational improvement and accountability. Measurement in education is essential for promoting student learning, guiding instructional practices, and improving educational outcomes. By employing valid, reliable, and fair assessment methods aligned with educational objectives, educators can effectively measure student progress and make informed decisions to support student success.

3.2 Learning Objectives:

After going through this Unit, learners will be able to:

- Understand criterion referenced tests.
- Develop an understanding of norm referenced test.

3.3 Criterion Referenced Tests:

A **criterion-referenced test** is a style of test which uses test scores to generate a statement about the behavior that can be expected of a person with that score. Most tests and quizzes that are written by school teachers can be considered criterion-referenced tests. In this case, the objective is simply

to see whether the student has learned the material. Criterion-referenced assessment can be contrasted with norm-referenced assessment
Criterion-referenced testing was a major focus of psychometric research in the 1970s.

Definitions: Criterion-referenced tests measure how well a test taker has mastered a specific set of learning objectives or criteria. Instead of comparing students' performance to that of others, the focus is on whether they meet predefined standards of performance.

- American **Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (2014)**: Criterion-referenced tests are assessments that measure the extent to which a test taker has mastered particular learning objectives or content, without reference to the performance of others. These tests are designed to determine whether a test taker has achieved a particular level of proficiency or mastery of specific skills, knowledge, or competencies, as defined by predetermined criteria or standards.
- Educational **Testing Service (ETS)**: Criterion-referenced tests are assessments that provide information about an individual's performance in relation to clearly defined criteria or standards. The focus of these tests is on whether the test taker has attained specific learning objectives or competencies, rather than comparing their performance to that of others. CRTs are used to evaluate mastery of content, skills, or competencies and to inform decisions about instructional planning, curriculum development, and student progress.

A common misunderstanding regarding the term is the meaning of *criterion*. Many, if not most, criterion-referenced tests involve a cut score, where the examinee passes if their score exceeds the cut score and fails if it does not (often called a mastery test). The *criterion* is not the cut score; the criterion is the domain of subject matter that the test is designed to assess. For example, the criterion may be "Students should be able to correctly add two single-digit numbers," and the cut score may be that students should correctly answer a minimum of 80% of the questions to pass.
The criterion-referenced interpretation of a test score identifies the relationship to the subject matter. In the case of a mastery test, this does mean identifying whether the examinee has "mastered" a specified level of the subject matter by comparing their score to the cuts core. However, not all criterion-referenced tests have a cut score, and the score can simply refer to a person's standing on the subject domain Because of this common misunderstanding, criterion-referenced tests have also been called standards-based assessments by some education agencies, as students are assessed with regards to standards that define what they "should" know, as defined by the state.

Criterion Referenced Test

A criterion-referenced test is a test that provides a basis for determining a candidate's level of knowledge and skills in relation to a well-defined domain of content. Often one or more performance standards are set on the test score scale to aid in test score interpretation. Criterion-referenced tests, a type of test introduced by Glaser (1962) and Popham and Husek (1969), are also known as domain-referenced tests, competency tests, basic skills tests, mastery tests, performance tests or assessments, authentic assessments, objective-referenced tests, standards-based tests, credentialing exams, and more. What all of these tests have in common is that they attempt to determine a candidate's level of performance in relation to a well-defined domain of content. This can be contrasted with norm-referenced tests, which determine a candidate's level of the construct measured by a test in relation to a well-defined reference group of candidates, referred to as the norm group. So it might be said that criterion-referenced tests permit a candidate's score to be interpreted in relation to a domain of content, and norm-referenced tests permit a candidate's score to be interpreted in relation to a group of examinees. The first interpretation is content-centered, and the second interpretation is examinee-centered.

Criterion-Referenced Tests (CRTs) are assessments designed to measure whether students have achieved specific learning objectives or criteria. Unlike norm-referenced tests, which compare students' performance to that of a norm group, CRTs focus on evaluating students' mastery of predefined standards or criteria. Here's an overview of Criterion-Referenced Tests:

Purpose:

1. **Assessing Mastery**: CRTs are used to determine whether students have mastered specific learning objectives, competencies, or standards.
2. **Evaluating Curriculum Alignment**: CRTs help educators assess the alignment between instructional objectives, curriculum, and assessment practices.
3. **Informing Instruction**: Results from CRTs provide valuable feedback to educators, informing instructional decision-making and guiding targeted interventions or support.

Key Features:

1. **Criteria-Based**: CRTs are designed based on specific criteria or standards that students are expected to achieve.
2. **Objective-Referenced**: Performance on CRTs is compared against predetermined performance standards rather than other students' performance.
3. **Absolute Interpretation**: Scores on CRTs are interpreted in absolute terms, indicating whether students have met, exceeded, or not met the established criteria.

4. **Directly Aligned with Curriculum**: CRTs are closely aligned with instructional objectives, ensuring that assessment measures what is taught.

Examples:

1. **State Standards Tests**: Assessments aligned with state academic standards, measuring students' proficiency in specific subject areas.
2. **End-of-Chapter Assessments**: Assessments administered at the end of a unit or chapter to evaluate students' mastery of learning objectives.
3. **Performance Tasks**: Tasks or projects designed to assess students' ability to apply knowledge and skills in authentic contexts, with scoring rubrics tied to specific criteria.
4. **Skills Assessments**: Assessments designed to measure students' proficiency in specific skills or competencies, such as writing, problem-solving, or scientific inquiry.

Benefits:

1. **Clear Feedback**: CRTs provide clear feedback to students and educators about students' strengths and areas for improvement.
2. **Curriculum Alignment**: CRTs help ensure that instruction and assessment practices are aligned with curriculum objectives and standards.
3. **Individualized Instruction**: Results from CRTs inform individualized instruction, allowing educators to tailor interventions to meet students' specific learning needs.

Considerations:

1. **Validity and Reliability**: CRTs must be valid and reliable, accurately measuring what they intend to measure and producing consistent results.
2. **Fairness**: CRTs should be fair and equitable for all students, free from bias or discrimination.
3. **Interpretation of Scores**: Educators must understand how to interpret CRT scores in relation to established criteria or standards.

Criterion-Referenced Tests play a crucial role in assessing student learning and informing instructional decision-making by providing clear, objective measures of students' mastery of specific learning objectives or standards.

Self-Check Exercise-1

Q.1 Which of the following best describes the function of Criterion-Referenced Tests (CRTs)?

a) Comparing students' performance to that of a norm group.
b) b) Measuring mastery of specific learning objectives or criteria.
c) c) Ranking students based on their performance relative to peers.
d) d) Providing feedback on students' progress over time.

Q.2 Criterion-Referenced Tests focus on evaluating individual students' performance against predetermined standards or criteria. True or False?

Q.3 Match the following terms:

a) Criterion-Referenced Tests	1. Measures whether students have achieved particular standards or criteria.
b) Mastery	2. The level of proficiency or expertise demonstrated by students.
c) Specific learning objectives	3. Clearly defined goals or targets for student learning.
d) Individual performance	4. Focuses on how well students perform relative to predetermined criteria, rather than comparing them to others.

3.4 Norm Referenced Test

Norm-referencing is based on the assumption that a roughly similar range of human performance can be expected for any student group. There is a strong culture of norm-referencing in higher education. It is evident in many commonplace practices, such as the expectation that the mean of a cohort's results should be a fixed percentage year-in year-out (often this occurs when comparability across subjects is needed for the award of prizes, for instance), or the policy of awarding first class honours sparingly to a set number of students, and so on. In contrast, criterion-referencing, as the name implies, involves determining a student's grade by comparing his or her achievements with clearly stated criteria for learning outcomes and clearly stated standards for particular levels of performance. Unlike norm-referencing, there is no pre-determined grade distribution to be generated and a student's grades is in no way influenced by the performance of others. Theoretically, all students within a particular cohort could receive very high (or very low) grades depending solely on the levels of individuals' performances against the established criteria and standards.

It is not always possible to be entirely objective and to comprehensively articulate criteria for learning outcomes: some subjectivity in setting and interpreting levels of achievement is inevitable in higher education. This being the case, sometimes the best we can hope for is to compare individuals' achievements relative to their peers. Norm-referencing, on its own — and if strictly and narrowly implemented — is undoubtedly unfair. With norm-referencing, a student's grade depends – to some extent at least – not only on his or her level of achievement, but also on the achievement of other students. This might lead to obvious inequities if applied without thought to any other considerations. For example, a student who fails in one year may well have passed in other years! The potential for unfairness of this kind is most likely in smaller student cohorts, where norm-referencing may force a spread of grades and exaggerate differences in achievement. Alternatively, norm-referencing might artificially compress the range of difference that actually

exists. The essential characteristic of norm-referencing is that students are awarded their grades on the basis of their ranking within a particular cohort. Norm-referencing involves fitting a ranked list of students' 'raw scores' to a pre-determined distribution for awarding grades. Usually, grades are spread to fit a 'bell curve' (a 'normal distribution' in statistical terminology), either by qualitative, informal rough-reckoning or by statistical techniques of varying complexity. For large student cohorts (such as in senior secondary education), statistical moderation processes are used to adjust or standardize student scores to fit a normal distribution. This adjustment is necessary when comparability of scores across different subjects is required (such as when subject scores are added to create an aggregate ENTER score for making university selection decisions). Norm-referenced score interpretations compare test-takers to a sample of peers. The goal is to rank students as being better or worse than other students. Norm-referenced test score interpretations are associated with traditional education. Students who perform better than others pass the test, and students who perform worse than others fail the test.

Definition: Norm-Referenced Tests (NRTs) are assessments that compare an individual's performance to that of a norm group, allowing for the interpretation of scores in relation to the performance of a larger population. The purpose of NRTs is to rank individuals based on their relative standing or performance compared to others in the norm group, rather than evaluating their mastery of specific learning objectives or criteria. NRTs are commonly used for comparative purposes, such as college admissions, selection for gifted programs, or identifying students in need of additional support or intervention.

Norm-Referenced Tests (NRTs) are assessments designed to compare an individual's performance to that of a norm group, typically a representative sample of the population. These tests provide information about how an individual's performance ranks relative to others in the norm group. Here's a brief overview of Norm-Referenced Tests:

Purpose:

1. **Comparative Assessment**: NRTs are used to compare an individual's performance to that of a norm group, allowing for the interpretation of scores in relation to the performance of a larger population.
2. **Ranking and Selection**: NRTs rank individuals based on their relative standing or performance compared to others in the norm group, making them useful for selection purposes, such as college admissions, employment, or program placement.
3. **Identifying Relative Strengths and Weaknesses**: NRTs provide information about individuals' relative strengths and weaknesses compared to their peers, helping identify areas where additional support or intervention may be needed.

Key Features:

1. **Norm Group**: NRTs use a norm group, which is a representative sample of the population, to establish norms or reference points for interpreting scores.
2. **Percentile Ranks**: Scores on NRTs are often reported as percentile ranks, indicating the percentage of individuals in the norm group who scored at or below a given score.
3. **Comparative Interpretation**: Interpretation of scores on NRTs involves comparing an individual's performance to that of the norm group, rather than evaluating mastery of specific learning objectives or criteria.
4. **Standardization**: NRTs are typically standardized to ensure consistency in administration, scoring, and interpretation across different test administrations and populations.

Examples:

1. **Standardized Achievement Tests**: Tests administered to students to assess their academic achievement in specific subject areas, such as reading, mathematics, or science.
2. **College Entrance Exams**: Exams like the SAT or ACT are norm-referenced assessments used for college admissions, with scores compared to those of a national sample of test-takers.
3. **Personality and Aptitude Tests**: Some personality and aptitude tests, such as the Myers-Briggs Type Indicator (MBTI) or the Wechsler Adult Intelligence Scale (WAIS), use norm-referenced scoring to compare individuals' results to those of a norm group.

Considerations:

1. **Population Characteristics**: The composition of the norm group should be representative of the population of interest to ensure the validity and fairness of score interpretation.
2. **Test Fairness**: NRTs should be fair and free from bias to ensure equitable assessment opportunities for all individuals.
3. **Interpretation**: It's important to interpret scores on NRTs in context, considering factors such as the characteristics of the norm group and the purpose of the assessment.

In summary, Norm-Referenced Tests provide valuable information about individuals' relative standing or performance compared to a norm group, making them useful for comparative assessment, ranking, and selection purposes. However, it's essential to interpret scores in context and consider the characteristics of the norm group when using NRTs for decision-making.

Self- Check Exercise-2

Q.1 Which of the following best describes the purpose of Norm-Referenced Tests (NRTs)?

a) Measuring mastery of specific learning objectives.

b) Comparing an individual's performance to that of a norm group.

c) Providing feedback on students' progress over time.

d) Assessing students' relative strengths and weaknesses.

Q.2 Norm-Referenced Tests provide information about how an individual's performance ranks relative to others in the norm group. True or False?

Q.3 Match the following terms:

a) Norm-Referenced Tests	1. Assessments designed to compare an individual's performance to that of a norm group.
b) Percentile Ranks	2. Scores on NRTs often reported as percentile ranks, indicating relative standing compared to the norm group.
c) Comparative Assessment	3. Purpose of NRTs, such as college admissions or program placement.
d) College Admissions	4. Examples of NRTs, such as the SAT or ACT.

3.5 Comparison of Criterion-Referenced and Norm-Referenced Measurement:

In psychology and education the information obtained from the tests are generally evaluated by norm-referenced measurement. One of the criteria of a good psychological test is norms. No test is a good test unless and until the norms are developed for it. On the basis of the test scores of an individual, one knows whether he is below average, average or above average in the group.

Criterion-referenced and norm-referenced measurements are differentiated on the basis of their aim. When the achievement of students has to be done in reference to some specific group then the measurement is norm-referenced. In criterion-referenced measurement the evaluation of ability is done on some criteria, e.g., some cut point for admission to some school or college.

Criterion-referenced and norm-referenced measurement can easily be understood by their comparison. These two measurements can easily be understood by following points:

1. Criterion and norm-referenced measurements are differentiated on the basis of information received by two measurements. The total attained knowledge of a student is known through norm-referenced measurement, while criterion-referenced measurement tells about specific educational aims. Through criterion-referenced measurement one knows how much of the specific objects of education are attained by the student and what is not attained. Norm-referenced measurement tells the number of questions which a student has solved. For example if a student had solved 7 questions out of

ten, then this information is received through norm-referenced measurement but if the three questions which he had not solved belongs to a particular area then we know that the student has not learned a particular phenomenon. This information is obtained only through criterion-referenced measurement.

2. The absolute amount of knowledge obtained by a student is known through criterion-referenced measurement, while norm-referenced measurement tells how much a student has attained in comparison to other students in the group.

3. In criterion-referenced measurement, through the educational methods one tries to know the amount of attained knowledge of a student, while in norm-referenced measurement the success of a student is evaluated in relative terms.

4. In criterion-referenced measurement certain items are included keeping in mind the specific objectives and they are confined to these objectives only, while in norm-referenced measurement, these items are generally spread over a wider area.

Dimension	**Criterion-Referenced Tests**	**Norm-Referenced Tests**
Definition	Measures student performance against predetermined criteria or standards.	Compares individual student performance to a norm group.
Purpose	- To determine whether each student has achieved specific skills or concepts. - To find out how much students know before instruction begins and after it has finished. - Determines mastery of specific skills or content.	- To rank each student with respect to the achievement of others in broad areas of knowledge. - To discriminate between high and low achievers. - Ranks students relative to one another.
Standards	Predetermined criteria or standards are explicit and specific.	No predetermined standards; comparison to norm group.
Content	- Measures specific skills which make up a designated curriculum. These skills are identified by teachers and curriculum experts. - Each skill is expressed as an instructional objective.	- Measures broad skill areas ample from a variety of textbooks, syllabus, and the judgments of curriculum experts.
Item	- Each skill is tested by at least	- Each skill is usually tested by

Characteristics	four items in order to obtain an adequate sample of student performance and to minimize the effect of guessing. - The items which test any given skill are parallel in difficulty.	less than four items. - Items vary in difficulty. - Items are selected that discriminate between high and low achievers.
Score Interpretation	- Each individual is compared with a preset standard for acceptable achievement. The performance of other examinee is irrelevant. - A student's score is usually expressed as a percentage. - Student achievement is reported for individual skills. - Based on whether students meet, exceed, or fall short of criteria.	- Each individual is compared with other examinee and assigned a score--usually expressed as a percentile, a grade equivalent score, or a stanine. - Student achievement is reported for broad skill areas, although some norm-referenced tests do report student achievement for individual skills. - Interpretation based on comparison to norm group
Feedback	Detailed and specific, highlighting areas of strength and areas needing improvement relative to criteria.	May focus on how a student's performance compares to peers.
Application	Common in competency-based education systems.	Often used in competitive selection processes.

Self- Check Exercise- 3

Q.1 What is the primary focus of criterion-referenced evaluation?

Q.2 How does norm-referenced evaluation differ from criterion-referenced evaluation in terms of standards?

Q.3 Explain the purpose of norm-referenced evaluation in educational assessment.

3.6 Summary:

Criterion-referenced evaluation and norm-referenced evaluation are two fundamental methods used in educational assessment, each offering unique perspectives on student performance. Criterion-referenced evaluation focuses on measuring individual achievement against predetermined criteria or standards, providing insight into whether students have mastered specific skills or content. In contrast, norm-referenced evaluation compares students' performance to that of a norm group, ranking them relative to their peers.

Criterion-referenced evaluation emphasizes mastery of specific objectives and provides detailed, criterion-based feedback to guide improvement. On the other hand, norm-referenced evaluation is often used for comparative purposes, such as competitive selection processes, where ranking and comparison are essential. While both approaches serve important roles in education, they differ in their emphasis, interpretation, and application, catering to diverse educational contexts and objectives.

3.7 Glossary:

Criterion-Referenced Evaluation: An assessment approach that measures individual student performance against predetermined criteria or standards.

Norm-Referenced Evaluation: An assessment approach that compares individual student performance to that of a norm group, ranking students relative to one another.

Standards: Predetermined criteria or benchmarks that define what students are expected to know or be able to do in criterion-referenced evaluation.

Criteria: Specific components or elements used to evaluate student performance in criterion-referenced evaluation.

Mastery: Achievement of a specified level of proficiency or competence in criterion-referenced evaluation, indicating that students have met the predetermined standards or criteria.

3.8 Answers to Self Check Exercises:

Self-Check Exercise-1

Answer1: b) Measuring mastery of specific learning objectives or criteria.

Answer2: True.

Answer3: a) 4 b) 2 c) 3 d) 1

Self- Check Exercise-2

Answer1: b) Comparing an individual's performance to that of a norm group.

Answer2: True.

Answer3: a) 1 b) 2 c) 3 d) 4

Self- Check Exercise-3

Answer1: The primary focus of criterion-referenced evaluation is to measure individual student performance against predetermined criteria or standards, assessing whether students have mastered specific skills or content.

Answer2: In norm-referenced evaluation, there are no predetermined standards. Instead, student performance is compared to that of a norm group. In criterion-referenced evaluation, standards are explicit and specific, outlining what students are expected to know or be able to do.

Answer3: The purpose of norm-referenced evaluation is to rank students relative to one another. It provides comparative information about students'

performance within a norm group, helping to identify the highest and lowest performers.

3.9 References/ Suggested Readings:

- Ebel, Robert L.(1966) “Measuring Educational Achievement, Prentice Hall of India Pvt. Ltd. Pp. 481
- Gronlund, N. E. (1976), Measurement and Evaluation in Teaching. McMillan, USA.
- Hopkins, C.D. and Antes, R.L. (1990).Classroom measurement and evaluation. Itasca, Illinois: Peacock.
- Izard, J. (1991). Assessment of learning in the classroom. Geelong, Vic.: Deakin University.
- Izard, J. (1997). Content Analysis and Test Blueprints. Paris: International Institute for Educational Planning.
- Mehrens, W.A. and Lehmann, I.J. (1984).Measurement and evaluation in education and psychology.(3rd Ed.) New York: Holt, Rinehart and Winston.
- Nandra, I.D.S.(2011). Learning Resources and Assessment of Learning.Patiala, 21st Century Publications.
- Taiwo, Adediran A. (2005). Fundamentals of Classroom Testing. New Delhi: Vikas Publishing House Pvt. Ltd.
- Walter W. Cook (1958). Educational Measurement. Washington D.C.: American Council on Education.
- Withers, G. (1997).Item Writing for Tests and Examinations. Paris: International Institute for Educational Planning.

3.10 Terminal Questions:

Q.1 Provide an example of a Criterion-Referenced Test and explain how it measures mastery of specific learning objectives.

Q.2 Discuss the benefits and limitations of Criterion-Referenced Tests in educational assessment.

Q.3 Explain how Norm-Referenced Tests are used in college admissions.

Q.4 Discuss the strengths and limitations of Norm-Referenced Tests in educational assessment.

UNIT 4: MEASUREMENT OF ACHIEVEMENT

Structure

4.1 Introduction:

Achievement generally refers to the successful accomplishment or attainment of goals, objectives, or standards. In an educational context, achievement specifically relates to the knowledge, skills, or competencies that individuals acquire or demonstrate as a result of their learning experiences. Despite the complexity, intangibility, and delayed fruition of many educational achievements and despite the relative imprecision of many of the techniques of educational measurement, there are logical grounds for believing that all important educational achievements can be measured. To be important, an educational achievement must lead to a difference in behavior. The person who has achieved more must in some circumstances behave differently from the person who has achieved less. If such a difference cannot be observed and verified no grounds exist for believing that the achievement is important.

4.2 Learning Objectives:

After completing this Unit, the learner will be able to;

- Understand the Concept of Achievement
- Understand the various dimensions such as academic, personal, and social achievement.
- Understand the meaning and significance of achievement tests.

4.3 Meaning of Achievement:

The purpose of achievement testing is to measure some aspect of the intellectual competence of human beings: what a person has learned to know or to do. Teachers use achievement tests to measure the attainments of their students. Employers use achievement tests to measure the competence of prospective employees. Professional associations use achievement tests to

exclude unqualified applicants from the practice of the profession. In any circumstances where it is necessary or useful to distinguish persons of higher from those of lower competence or attainments, achievement testing is likely to occur.

The varieties of intellectual competence that may be developed by formal education, self-study, or other types of experience are numerous and diverse. There is a corresponding number and diversity of types of tests used to measure achievement.

Measurement, in its most fundamental form, requires nothing more than the verifiable observation of such a difference. If person A exhibits to any qualified observer more of a particular trait than person B, then that trait is measurable. By definition, then, any important achievement is potentially measurable.

Many important educational achievements can be measured quite satisfactorily by means of paper and pencil tests. But in some cases the achievement is so complex, variable, and conditional that the measurements obtained are only rough approximations. In other cases the difficulty lies in the attempt to measure something that has been alleged to exist but that has never been defined specifically. Thus, to say that all important achievements are potentially measurable is not to say that all those achievements have been clearly identified or that satisfactory techniques for measuring all of them have been developed.

Here are some key aspects of the meaning of achievement:

Attainment of Goals: Achievement involves reaching specific goals or objectives set by educators, institutions, or individuals themselves. These goals may include mastering academic content, developing critical thinking skills, or demonstrating proficiency in specific subject areas.

Demonstration of Competence: Achievement often involves demonstrating competence or proficiency in a particular domain or skill. This may be assessed through various means, such as tests, projects, presentations, or performance evaluations.

Recognition of Effort and Progress: Achievement recognizes the effort and progress made by individuals in their learning journey. It acknowledges both the process of learning and the outcomes achieved.

Personal Growth and Development: Achievement is not solely about academic success but also encompasses personal growth and development. It may involve overcoming challenges, developing resilience, and acquiring transferable skills that contribute to lifelong learning and success.

Measurement and Evaluation: Achievement is often measured and evaluated through assessments and evaluations. These assessments may include standardized tests, classroom assignments, performance tasks, and teacher observations, among others.

Contextual and Relative: Achievement is contextual and relative, meaning it can vary depending on individual circumstances, cultural backgrounds, and

educational contexts. What constitutes achievement for one person or group may differ from another.

Self- Check Exercise-1

Q.1 What does achievement generally refer to?

a) Successful accomplishment of goals
b) Failure to meet expectations
c) Mediocrity in performance
d) Lack of effort in learning

Q. 2 In an educational context, achievement specifically relates to:

a) Accumulation of wealth
b) Acquisition of material possessions
c) Attainment of knowledge and skills
d) Social status

Q. 3 Which of the following is NOT a characteristic of achievement?

a) Recognition of effort and progress
b) Personal growth and development
c) Measurement through subjective criteria
d) Attainment of specific goals or standards

4.4 Measurement Of Achievement:

"Measurement of achievement" refers to the process of assessing and quantifying a person's performance or proficiency in a particular domain or skill. It involves the use of various assessment tools and techniques to evaluate how well individuals have mastered specific learning objectives or standards. The measurement of achievement plays a crucial role in education and other fields, providing valuable information about students' progress, strengths, and areas needing improvement.

In education, the measurement of achievement encompasses a wide range of assessment methods, including tests, quizzes, projects, presentations, and performance tasks. These assessments may be designed to measure different types of achievement, such as cognitive skills (e.g., problem-solving, critical thinking), academic knowledge (e.g., subject-specific content), or psychomotor skills (e.g., manual dexterity, physical fitness).

The process of measuring achievement typically involves several steps:

Setting Objectives or Standards: Establishing clear learning objectives or standards that define what students are expected to know or be able to do.

Selecting Assessment Methods: Choosing appropriate assessment methods and tools to measure achievement in alignment with the established objectives or standards.

Administering Assessments: Conducting assessments to collect data on students' performance and achievement.

Scoring and Evaluation: Evaluating students' responses or performances based on predetermined criteria or scoring rubrics.

Interpreting Results: Analyzing assessment results to understand students' strengths, weaknesses, and overall achievement levels.
Providing Feedback: Providing feedback to students based on their performance to support their learning and growth.
Using Results for Decision-Making: Using assessment data to inform instructional planning, curriculum development, and interventions to enhance student learning and achievement.
Effective measurement of achievement requires careful consideration of assessment validity, reliability, fairness, and authenticity. Valid assessments accurately measure what they are intended to measure, while reliable assessments produce consistent results over time and across different contexts. Fair assessments ensure that all students have an equal opportunity to demonstrate their achievement, regardless of factors such as background or disability. Authentic assessments provide meaningful tasks and contexts that reflect real-world applications of knowledge and skills.
Overall, the measurement of achievement is essential for monitoring progress, guiding instruction, and promoting continuous improvement in education and beyond.

Self-Check Exercise-2:

Q. 1 Learning outcomes of achievement represent the specific knowledge, skills, or competencies that individuals are expected to acquire or demonstrate as a result of their educational _____.

Q. 2 The objectives of teaching the topic of achievement encompass several key goals aimed at promoting student learning, growth, and _____.

Q.3 Effective measurement of achievement requires careful consideration of assessment validity, reliability, _____, and authenticity.

4.5 Achievement Tests:

An **achievement test** is a test of developed skill or knowledge. The most common type of achievement test is a standardized test developed to measure skills and knowledge learned in a given grade level, usually through planned instruction, such as training or classroom instruction. Achievement tests are often contrasted with tests that measure aptitude, a more general and stable cognitive trait. Achievement test scores are often used in an educational system to determine what level of instruction for which a student is prepared. High achievement scores usually indicate a mastery of grade-level material, and the readiness for advanced instruction. Low achievement scores can indicate the need for remediation or repeating a course grade.
Under No Child Left Behind, achievement tests have taken on an additional role of assessing proficiency of students. Proficiency is defined as the amount of grade-appropriate knowledge and skills a student has acquired up to the point of testing. Better teaching practices are expected to increase the amount

learned in a school year, and therefore to increase achievement scores, and yield more "proficient" students than before.

When writing achievement test items, writers usually begin with a list of content standards (either written by content specialists or based on state-created content standards) which specify exactly what students are expected to learn in a given school year. The goal of item writers is to create test items that measure the most important skills and knowledge attained in a given grade-level. The number and type of test items written is determined by the grade-level content standards. Content validity is determined by the representatives of the items included on the final test. Achievement tests are assessments designed to measure the knowledge, skills, or abilities that individuals have acquired in a specific subject area or domain. There are several types of achievement tests, each tailored to assess different aspects of learning and achievement.

Types of Achievement Tests:

Standardized Achievement Tests: These tests are designed to measure students' performance in a particular subject area using standardized procedures and scoring. They often assess a broad range of content within a specific grade level or educational domain and are administered to large groups of students. Examples include state standardized tests, national assessments (e.g., SAT, ACT), and international assessments (e.g., PISA).

Subject-Specific Achievement Tests: These tests focus on assessing proficiency in a specific subject area, such as mathematics, reading, science, or social studies. Subject-specific achievement tests may cover a range of topics within the subject area and are often used to evaluate students' mastery of curriculum standards or learning objectives in that subject.

Diagnostic Achievement Tests: Diagnostic tests are designed to identify students' strengths and weaknesses in a particular subject area or skill domain. They provide detailed information about students' current levels of understanding and can help educators identify areas where additional instruction or support may be needed. Diagnostic tests may be administered before or during instruction to inform teaching practices and curriculum planning.

Formative Assessment: While not a traditional achievement test, formative assessment techniques are used to monitor students' progress and understanding throughout the learning process. Formative assessments provide feedback to both students and teachers and can help guide instructional decisions in real-time. Examples of formative assessment techniques include quizzes, exit tickets, classroom discussions, and peer/self-assessment.

Summative Assessment: Summative assessments are administered at the end of a unit, course, or school year to evaluate students' overall achievement and mastery of learning objectives. These assessments often take the form of

comprehensive exams, final projects, or standardized tests and are used to assign grades or make decisions about student progression or graduation.

Criterion-Referenced Tests: Criterion-referenced tests measure students' performance against specific learning criteria or standards. These tests assess whether students have mastered predefined objectives or competencies, rather than comparing their performance to that of other students. Criterion-referenced tests are commonly used in competency-based education systems and to assess mastery of specific skills or knowledge.

Norm-Referenced Tests: Norm-referenced tests compare students' performance to that of a norm group, typically a representative sample of students from the same grade level or age group. These tests provide information about how individual students' performance ranks relative to their peers. Norm-referenced tests are often used for comparative purposes, such as identifying high and low achievers or making admissions decisions.

These are some of the most common types of achievement tests used in education. Each type serves different purposes and provides valuable information about students' learning and achievement in various subject areas and skill domains.

Importance and Limitations of Achievement Tests:

Achievement tests play important roles in education, in government, in business and industry, and in the professions. If they were constructed more carefully and more expertly, and used more consistently and more wisely, they could do even more to improve the effectiveness of these enterprises.

But achievement tests also have limitations beyond those attributable to hasty, inexpert construction or improper use. In the first place, they are limited to measuring a person’s command of the knowledge that can be expressed in verbal or symbolic terms. This is a very large area of knowledge, and command of it constitutes a very important human achievement; but it does not include all knowledge, and it does not represent the whole of human achievement. There is, for example, the unverbalized knowledge obtained by direct perceptions of objects, events, feelings, relationships, etc. There are also physical skills and behavioral skills, such as leadership and friendship, that are not highly dependent on command of verbal knowledge. A paper and pencil test of achievement can measure what a person knows about these achievements but not necessarily how effectively he uses them in practice.

In the second place, while command of knowledge may be a necessary condition for success in modern human activities, it is by no means a sufficient condition. Energy, persistence, and plain good fortune, among other things, combine to determine how successfully he uses the knowledge he possesses. A person with high achievement scores is a better bet to succeed than one with low achievement scores, but high scores cannot guarantee success.

Self-Check Exercise-3:

Q 1. Achievement tests are designed to measure:
a) Personality traits
b) Physical fitness
c) Knowledge, skills, or abilities
d) Social behavior
Q 2. Which of the following is NOT a characteristic of achievement tests?
a) Administered at the end of the school year
b) Measure specific learning objectives
c) Provide feedback on student performance
d) Assess mastery of content or skills
Q 3. Which of the following is an example of an achievement test?
a) Personality inventory
b) Physical fitness test
c) End-of-course exam
d) Career aptitude test
Q 4. The primary purpose of achievement tests is to:
a) Assess personality traits
b) Evaluate physical abilities
c) Measure academic performance
d) Determine career aptitude

4.6 Summary:

Achievement and achievement tests play integral roles in educational assessment, providing valuable insights into students' learning, growth, and mastery of knowledge and skills. Achievement encompasses the successful accomplishment or attainment of goals, objectives, or standards, reflecting both academic and personal success. Achievement tests are specifically designed assessments that measure individuals' knowledge, skills, or abilities in a particular subject area or domain. These tests can be norm-referenced, comparing students' performance to that of a norm group, or criterion-referenced, measuring performance against specific criteria or standards. They serve various purposes, including diagnosing learning needs, monitoring progress, evaluating mastery of learning objectives, and informing instructional decisions. Achievements tests are administered through standardized procedures, with careful consideration given to validity, reliability, fairness, and authenticity. Feedback from achievement tests guides instruction, supports student learning and growth, and informs educational decision-making. Overall, achievement and achievement tests are essential components of the educational assessment landscape, providing valuable information to educators, students, parents, and policymakers about learning outcomes and academic achievement.

4.7 Glossary:

· **Achievement**: The successful accomplishment or attainment of goals, objectives, or standards, often involving the acquisition of knowledge, skills, or competencies.
· **Assessment**: The process of collecting, analyzing, and interpreting information about student learning and performance.
· **Achievement Tests**: Assessments designed to measure the knowledge, skills, or abilities that individuals have acquired in a specific subject area or domain.
· **Scoring**: The process of assigning numerical or descriptive ratings to student responses or performances based on predetermined criteria or rubrics.
· **Feedback**: Information provided to students about their performance on assessments, highlighting strengths and areas needing improvement to support learning and growth.
· **Fairness**: The extent to which an assessment provides all students with an equal opportunity to demonstrate their knowledge, skills, or abilities, regardless of factors such as background or disability.
· **Bias**: Systematic errors or inaccuracies in assessment items or procedures that unfairly advantage or disadvantage certain groups of students.

4.8 Answers to Self-Check Exercises:

Self- Check Exercise-1

Answer1: a) Successful accomplishment of goals
Answer2: c) Attainment of knowledge and skills
Answer3: c) Measurement through subjective criteria

Self- Check Exercise-2

Answer1: experiences
Answer2: success
Answer3: fairness

Self- Check Exercise-3

Answer 1: c) Knowledge, skills, or abilities
Answer 2: a) Administered at the end of the school year
Answer3: c) End-of-course exam
Answer4: c) Measure academic performance

4.9 References/Suggested Readings:

- Ebel, Robert L.(1966) “Measuring Educational Achievement, Prentice Hall of India Pvt. Ltd. Pp. 481
- Gronlund, N. E. (1976), Measurement and Evaluation in Teaching. McMillan, USA.
- Hopkins, C.D. and Antes, R.L. (1990).Classroom measurement and evaluation. Itasca, Illinois: Peacock.

- Izard, J. (1991). Assessment of learning in the classroom. Geelong, Vic.: Deakin University.
- Izard, J. (1997). Content Analysis and Test Blueprints. Paris: International Institute for Educational Planning.
- Mehrens, W.A. and Lehmann, I.J. (1984).Measurement and evaluation in education and psychology.(3rd Ed.) New York: Holt, Rinehart and Winston.
- Nandra, I.D.S.(2011). Learning Resources and Assessment of Learning.Patiala, 21st Century Publications.
- Taiwo, Adediran A. (2005). Fundamentals of Classroom Testing. New Delhi: Vikas Publishing House Pvt. Ltd.
- Walter W. Cook (1958). Educational Measurement. Washington D.C.: American Council on Education.
- Withers, G. (1997).Item Writing for Tests and Examinations. Paris: International Institute for Educational Planning.

4.10 Terminal Questions:

Q.1 What is the definition of achievement in an educational context?

Q.2 What are achievement tests, and what do they measure?

Q.3 What is the purpose of achievement tests in education?

Q.4 How are achievement tests administered, and what factors are considered in their development?

Q.5 What role does feedback from achievement tests play in education?

UNIT 5- MEASUREMENT OF ATTITUDE AND SKILLS

Structure

5.1 Introduction:

The measurement of attitude and skills plays a crucial role in educational assessment, providing valuable insights into individuals' beliefs, behaviors, competencies, and performance. Attitudes reflect individuals' feelings, beliefs, and behavioral tendencies toward specific objects, people, groups, or situations, while skills represent their abilities or competencies developed through learning, practice, and experience.

Measurement of attitude and skills involves systematically collecting, analyzing, and interpreting data to assess individuals' attitudes, beliefs, behaviors, or competencies using standardized procedures and instruments. This process aims to provide accurate and reliable information about individuals' attitudes and skills, informing decision-making in education, training, and professional development.

In educational settings, measuring attitudes helps educators understand students' motivation, engagement, and perceptions of learning experiences, providing insights into their social and emotional development. Assessing skills, on the other hand, enables educators to evaluate students' mastery of specific competencies or learning objectives, guiding instructional planning and support.

Various assessment methods can be used to measure attitudes and skills, including surveys, questionnaires, interviews, observations, performance assessments, and self-assessments. These methods may involve quantitative or qualitative data collection techniques, depending on the nature of the attitudes or skills being assessed and the desired outcomes of the assessment.

Validity and reliability are critical considerations in the measurement of attitude and skills, ensuring that assessment results are accurate, meaningful, and consistent over time and across different contexts. Validity refers to the extent to which an assessment measures what it is intended to measure, while reliability refers to the consistency and stability of assessment results.

Feedback plays a vital role in the measurement of attitude and skills, providing individuals with information about their performance, strengths, and areas needing improvement. Constructive feedback helps individuals reflect on their attitudes and skills, set goals for improvement, and monitor their progress over time.

Overall, the measurement of attitude and skills is essential for promoting learning, growth, and development in educational and professional settings. By systematically assessing individuals' attitudes and skills, educators can identify areas of strength and areas needing improvement, tailor instruction to meet learners' needs effectively, and support their ongoing development and success.

5.2 Learning Objectives:

After completion of this Unit, the learner will be able to;

- Understand the concept and importance of measuring attitude.
- Analyse the importance of skill measurement in educational process.

5.3 Measurement of Attitude:

Attitude is such a complex affair that it cannot be completely described. Thurston has used the concept of attitude to denote 'the sum total of a man's inclinations and feelings prejudice or bias, pre-conceived notions, ideas, threats and convictions about any specific topic." Thus a man's attitude about pacifism means all that he feels and thinks about war and peace. It is obviously a subjective and personal affair.

The concept opinion is used to denote a verbal expression of attitude. If a man says that we made a mistake in entering the U.N.O., it would be called his opinion. But it also shows that his attitude is anti- U.N.O. thus, opinion is a verbal expression of attitude.

"An attitude is essentially a form of anticipatory response, a beginning of action not necessarily completed." –**K. Young**.

"An attitude is a mental and neutral state of readiness, exerting directive or dynamic influence upon the individual's response to all objects and situations with which it is related."- **BRITT**

Characteristics of Attitudes

The following characteristics may be noted:

- Unlimited range of attitudes; our likes, dislikes, food we take, everything is an aspect of attitude.
- It is a position towards outer objects, either for or against.
- There are individual differences in attitudes.
- Attitudes are the basis of behavior as they lead to strike, war, voting, etc.
- They may be overt or covert.
- They are integrated into an organized system.

- They always imply a subject-object relationship.

Measurement of Attitudes

The following dimensions or properties of attitudes are important in the measurement of attitudes:

i. **Direction**: i.e., for or against any issue.

ii. **Degree**: i.e., amount of favorableness or unfavourableness on a continuum.

iii. **Strength or Intensity**: i.e., how strong an as attitude is.

iv. **Salience**: i.e., freedom or spontaneity with which it is manifested.

v. **Contrivance or Consistency**: i.e., hoe does an individual maintains his attitudes under different conditions.

There are various techniques for the measurement of attitudes. The projective techniques include Rorschach, ThematicAppreciation Test (T.A.T), World association Test and sentence Completion Test. All these can be utilized for measuring attitudes. Questionnaires, inventories, situational test and interviews are also helpful. Perhaps the most important technique of measuring attitudes is the 'scaling' technique. **Thurston and Chave** mastered this technique by constructing an unique attitude scale.

Attitude Measurement

Perhaps the most straightforward way of finding out about someone's attitudes would be to ask them. However, attitudes are related to self-image and social acceptance (i.e. attitude functions).

In order to preserve a positive self-image, people's responses may be affected by social desirability. They may not well tell about their true attitudes, but answer in a way that they feel socially acceptable.

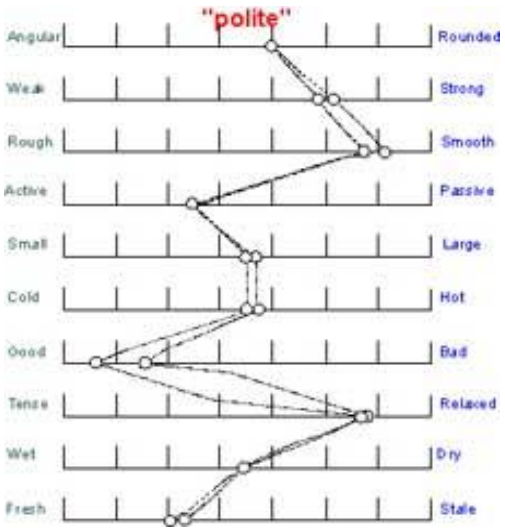
Given this problem, various methods of measuring attitudes have been developed. However, all of them have limitations. In particular the different measures focus on different components of attitudes – cognitive, affective and behavioral – and as we know, these components do not necessarily coincide.

Attitude measurement can be divided into two basic categories

- **Direct Measurement** (**likert scale** and semantic differential)
- **Indirect Measurement** (projective techniques)

Semantic Differential

The semantic differential technique of Osgood et al. (1957) asks a person to rate an issue or topic on a standard set of **bipolar adjectives** (i.e. with opposite meanings), each representing a **seven point scale**.



To prepare a semantic differential scale, you must first think of a number of words with opposite meanings that are applicable to describing the subject of the test.

For example, participants are given a word, for example 'car', and presented with a variety of adjectives to describe it. Respondents tick to indicate how they feel about what is being measured.

In the picture (above), you can find Osgood's map of people's ratings for the word 'polite'. The image shows ten of the scales used by Osgood. The image maps the average responses of two groups of 20 people to the word 'polite'.

The semantic differential technique reveals information on three basic dimensions of attitudes: evaluation, potency (i.e. strength) and activity.

• **Evaluation** is concerned with whether a person thinks positively or negatively about the attitude topic (e.g. dirty – clean, and ugly - beautiful).

• **Potency** is concerned with how powerful the topic is for the person (e.g. cruel – kind, and strong - weak).

• **Activity** is concerned with whether the topic is seen as active or passive (e.g. active – passive).

Using this information we can see if a person's feeling (evaluation) towards an object is consistent with their behavior. For example, a place might like the taste of chocolate (evaluative) but not eat it often (activity). The evaluation dimension has been most used by social psychologists as a measure of a person's attitude, because this dimension reflects the affective aspect of an attitude.

Evaluation of Direct Methods

An attitude scale is designed to provide a valid, or accurate, measure of an individual's social attitude. However, as anyone who has every "faked" an attitude scales knows there are shortcomings in these self report scales of attitudes. There are various problems that affect the validity of attitude scales. However, the most common problem is that of social desirability.

Socially desirability refers to the tendency for people to give "socially desirable" to the questionnaire items. People are often motivated to give

replies that make them appear “well adjusted”, unprejudiced, open minded and democratic. Self report scales that measure attitudes towards race, religion, sex etc. are heavily affected by socially desirability bias.

Respondents who harbor a negative attitude towards a particular group may not wish be admit to the experimenter (or to themselves) that they have these feelings. Consequently, responses on attitude scales are not always 100% valid.

Projective Techniques

To avoid the problem of social desirability, various indirect measures of attitudes have been used. Either people are unaware of what is being measured (which has ethical problems) or they are unable consciously to affect what is being measured.

Indirect methods typically involve the use of a projective test. A **projective test** is involves presenting a person with an ambiguous (i.e. unclear) or incomplete stimulus (e.g. picture or words). The stimulus requires interpretation from the person. Therefore, the person’s attitude is inferred from their interpretation of the ambiguous or incomplete stimulus.

The assumption about these measures of attitudes it that the person will “project” his or her views, opinions or attitudes into the ambiguous situation, thus revealing the attitudes the person holds. However, indirect methods only provide general information and do not offer a precise measurement of attitude strength since it is qualitative rather than quantitative. This method of attitude measurement is not objective or scientific which is a big criticism.

Examples of projective techniques include:

- Rorschach Inkblot Test
- Thematic Apperception Test (or TAT)
- Draw a Person Task

Thematic Apperception Test

Here a person is presented with an ambiguous picture which they have to interpret. The Thematic Apperception Test (TAT) taps into a person's unconscious mind to reveal the repressed aspects of their personality.

Although the picture, illustration, drawing or cartoon that is used must be interesting enough to encourage discussion, it should be vague enough not to immediately give away what the project is about.

TAT can be used in a variety of ways, from eliciting qualities associated with different products to perceptions about the kind of people that might use certain products or services.

The person must look at the picture(s) and tell a story. For example:

- What has led up to the event shown.
- What is happening at the moment.
- What the characters are thinking and feeling, and
- What the outcome of the story was.

Draw a Person Test

Figure drawings are **projective diagnostic techniques** in which an individual is instructed to draw a person, an object, or a situation so that cognitive, interpersonal, or psychological functioning can be assessed. The test can be used to **evaluate children** and adolescents for a variety of purposes (e.g. self-image, family relationships, cognitive ability and personality).

A projective test is one in which a test taker responds to or provides ambiguous, abstract, or unstructured stimuli, often in the form of pictures or drawings.

While other projective tests, such as the **Rorschach Technique** and **Thematic Apperception Test**, ask the test taker to interpret existing pictures, figure drawing tests require the test taker to create the pictures themselves. In most cases, figure drawing tests are given to children. This is because it is a simple, manageable task that children can relate to and enjoy.

Some figure drawing tests are primarily measures of cognitive abilities or cognitive development. In these tests, there is a consideration of how well a child draws and the content of a child's drawing. In some tests, the child's self-image is considered through the use of the drawings.

In other figure drawing tests, interpersonal relationships are assessed by having the child draw a family or some other situation in which more than one person is present. Some tests are used for the evaluation of child abuse. Other tests involve personality interpretation through drawings of objects, such as a tree or a house, as well as people.

Finally, some figure drawing tests are used as part of the diagnostic procedure for specific types of psychological or neuropsychological impairment, such as central nervous system dysfunction or mental retardation.

Despite the flexibility in administration and interpretation of figure drawings, these tests require skilled and trained administrators familiar with both the theory behind the tests and the structure of the tests

themselves. Interpretations should be made with caution and the limitations of projective tests should be considered.

It is generally a good idea to use projective tests as part of an overall test battery. There is little professional support for the use of figure drawing, so the examples that follow should be interpreted with caution.

The **House-Tree-Person (HTP)** test, created by Buck in 1948, provides a measure of a self-perception and attitudes by requiring the test taker to draw a house, a tree, and a person.

- The picture of the house is supposed to conjure the child's feelings toward his or her family.
- The picture of the tree is supposed to elicit feelings of strength or weakness. The picture of the person, as with other figure drawing tests, elicits information regarding the child's self-concept.

The HTP, though mostly given to children and adolescents, is appropriate for anyone over the age of three.

Evaluation of Indirect Methods

The major criticism of indirect methods is their lack of objectivity. Such methods are unscientific and do not objectively measure attitudes in the same way as a Likert scale.

There is also the ethical problem of deception as often the person does not know that their attitude is actually being studied when using indirect methods.

The advantages of such indirect techniques of attitude measurement are that they are less likely to produce socially desirable responses, the person is unlikely to guess what is being measured and behavior should be natural and reliable.

Self-Check Exercise-1

Q.1 What is the primary purpose of measuring attitudes in educational settings?
a) To assess physical fitness
b) To evaluate academic achievement
c) To understand students' beliefs and perceptions
d) To monitor attendance records
Q.2 Which of the following is a common method for measuring attitudes?
a) Blood pressure monitoring
b) IQ testing
c) Surveys and questionnaires
d) Physical fitness assessments

5.4 Measurement of Cognitive And Non-Cognitive Skills:

Cognitive Skills

Measures of cognition have been developed and refined over the past century. Cognitive ability has multiple facets. Psychologists distinguish between fluid intelligence (the rate at which people learn) and crystallized intelligence (acquired knowledge).Achievement tests are designed to capture crystallized intelligence whereas IQ tests like Raven's progressive matrices (1962) are designed to capture fluid intelligence.

This new understanding of cognition is not widely appreciated. Many use IQ tests, standardized achievement tests, and even grades as interchangeable measures of "cognitive ability" or intelligence. Scores on IQ tests and standardized achievement tests are strongly correlated with each other and with grades. However, these general indicators of "cognition" measure different skills and capture different facets of cognitive ability.

Measuring Non-cognitive Skills

We use the term non-cognitive skills to describe the personal attributes not thought to be measured by IQ tests or achievement tests. These attributes go by many names in the literature, including soft skills, personality traits, non-cognitive abilities, character skills, and socio emotional skills. These different names connote different properties. "Traits" suggests a sense of permanence and possibly also of heritability. "Skills" suggests that these attributes can be learned. In reality, the extent to which these personal attributes can change lies on a spectrum. Both cognitive and non cognitive skills can change and be changed over the life cycle, but through different mechanisms and with different ease at different ages. We use the term skill because all attributes can be shaped.

Although non-cognitive skills are overlooked in most contemporary policy discussions and in economic models of choice behaviour, personality psychologists have studied these skills for the past century. Psychologists primarily measure non-cognitive skills by using self-reported surveys or observer reports. They have arrived at a relatively well-accepted taxonomy of non-cognitive skills called the Big Five, with the acronym **OCEAN**, which

stands for: *Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism*.

A Task-Based Framework for Identifying and Measuring Skills

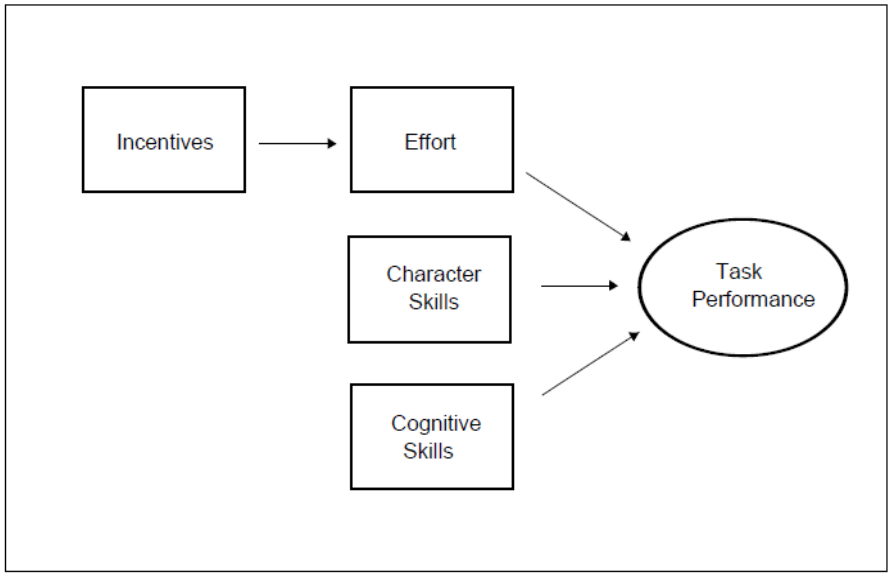
A leading personality psychologist defines personality (non-cognitive) traits (skills) as follows:

Personality traits are the relatively enduring patterns of thoughts, feelings, and behaviours that reflect the tendency to respond in certain ways under certain circumstances.

Roberts' definition of personality ("non-cognitive" skills) and the one favoured by Almlund et al. suggests that all psychological measurements are calibrated on measured behaviour or "tasks" broadly defined. A task could be taking an IQ test, answering a personality questionnaire, performing a job, attending school, completing secondary school, participating in crime, or performing in an experiment run by a social scientist.

Figure below depicts how performance on a task can depend on incentives, effort, and cognitive and non-cognitive skills. Performance on different tasks depends on these components to different degrees. People can compensate for their shortfalls in one dimension by having strengths in other dimensions.

Figure: Determinants of Task Performance



Many believe that personality skills can only be assessed by self-reported questionnaires that elicit skills like the Big Five. However, performance on any task or any observed behaviour can be used to measure personality and other skills. For example, completing high school requires many other skills besides those measured by

Achievement tests, including showing up in school, paying attention, and behaving in class.

Inferring skills from performance on tasks requires standardizing all of the other contributing factors that produce the observed behaviors. The inability to parse and localize behaviors that depend on a single skill or ability gives rise

to a fundamental problem of assessing the contribution of any particular skill to the successful performance on any task (or measure). This problem is commonly ignored in empirical research that studies how cognitive and non-cognitive skills affect outcomes. There are two distinct issues that need to be addressed in designing measures of skills based on performance of any task.

First, behaviour depends on incentives created by situations. Different incentives

elicit different amounts of effort on the tasks used to measure skills. Accurately measuring non-cognitive skills requires standardizing for the effort applied in any task. **Second**, performance on most tasks depends on multiple skills. Not standardizing for incentives and other relevant skills that determine performance on a particular task used to measure a particular skill can produce misleading estimates of that particular skill.

Measuring Skills Using Behaviours

Ralph Tyler, one of the two scholars suggested using measures of behaviour such as performance, participation in student activities, and other observations by teachers and school administrators to complement achievement tests while evaluating students and schools. They develop and apply methods to use high school grades to measure both cognitive and non cognitive skills. They show that non-cognitive skills promote educational attainment, beneficial labour market outcomes, and health.

Some criticize this approach and argue that it is tautological to use measures of behaviour to predict other behaviours even though the measures are taken early in life to predict later life behaviours. As suggested by Figure, all tasks or behaviours can be used to infer a skill as long as the measurement accounts for other skills and aspects of the situation. In addition, many of the recent studies in economics use early measures of behaviours to predict behaviours in adulthood.

Self-reported scales should not be assumed to be more reliable than behaviours, although personality psychologists often assume so. The question is which measurements are most predictive and which can be implemented in practice. The literature suggests that there are objective measurements of non-cognitive skills that are not plagued by reference bias.

Self-Check Exercise-2:

Q.1 What is the primary purpose of measuring skills in educational settings?

a) To assess physical health

b) To evaluate academic knowledge

c) To understand students' abilities and competencies

d) To monitor attendance records

Q.2 Which of the following is a common method for measuring skills?

a) IQ testing

b) Surveys and questionnaires

c) Performance assessments
d) Blood pressure monitoring

5.5 Summary:

Education is an extensive, diverse and complex enterprise, not only in learns of the achievements it seeks to develop, but also in terms of the means by which it seeks to develop. Our understanding of the nature and process of education is far from perfect. Hens, it is easy to agree that we do not know how to measure all important education outcomes. But, in principle, all important outcomes of education are measurable. They may not even be measurable in principle using only paper and pencil tests. But if they are known to be important, they must be measurable. To be important, an outcome of education must make an observable difference. That is, at some time, under some circumstances a person who has more of it must behave differently from a person who has less of it. If different degree or amounts of an education achievement never make any observation difference, what evidence can be found to show that it is in fact important? But if such difference can be observed, then the achievement is measurable for all that measurement requires is verifiable observation of a more –less relationship.

5.6 Glossary:

· **Attitude**: A psychological tendency that reflects an individual's beliefs, feelings, and behavioral tendencies toward a particular object, person, group, or situation.

· **Measurement**: The process of assigning numerical or descriptive values to attributes or characteristics of individuals, objects, or phenomena using standardized procedures and instruments.

· **Skills**: Abilities or competencies that individuals develop through learning, practice, and experience, enabling them to perform tasks or activities effectively and efficiently.

· **Attitude Measurement**: The process of assessing individuals' attitudes toward specific topics, issues, or objects using standardized instruments such as surveys, questionnaires, or scales.

· **Skill Assessment**: The process of evaluating individuals' abilities or competencies in specific domains or areas of performance through observation, demonstration, testing, or other assessment methods.

5.7 Answers To Self Check Exercises:

Self-Check Exercise 1:

Answer1: c) To understand students' beliefs and perceptions
Answer2: c) Surveys and questionnaires

Self- check Exercise-2

Answer1: c) To understand students' abilities and competencies

Answer2: c) Performance assessments

5.8 References/ Suggested Readings:

- Ebel, Robert L.(1966) “Measuring Educational Achievement, Prentice Hall of India Pvt. Ltd. Pp. 481
- Gronlund, N. E. (1976), Measurement and Evaluation in Teaching. McMillan, USA.
- Hopkins, C.D. and Antes, R.L. (1990).Classroom measurement and evaluation. Itasca, Illinois: Peacock.
- Izard, J. (1991). Assessment of learning in the classroom. Geelong, Vic.: Deakin University.
- Izard, J. (1997). Content Analysis and Test Blueprints. Paris: International Institute for Educational Planning.
- Mehrens, W.A. and Lehmann, I.J. (1984).Measurement and evaluation in education and psychology.(3rd Ed.) New York: Holt, Rinehart and Winston.
- Nandra, I.D.S.(2011). Learning Resources and Assessment of Learning.Patiala, 21st Century Publications.
- Taiwo, Adediran A. (2005). Fundamentals of Classroom Testing. New Delhi: Vikas Publishing House Pvt. Ltd.
- Walter W. Cook (1958). Educational Measurement. Washington D.C.: American Council on Education.
- Withers, G. (1997).Item Writing for Tests and Examinations. Paris: International Institute for Educational Planning.

5.9 Terminal Questions:

Q1. What is the purpose of measuring attitudes in educational settings?

Q2. How can attitudes be assessed effectively? Provide examples of assessment methods.

Q3. What role does feedback play in skill development and assessment?

UNIT-6 EVALUATION IN EDUCATION

Structure:

6.1 Introduction:

Evaluation is a methodological area that is closely related to, but distinguishable from more traditional social research. Evaluation utilizes many of the same methodologies used in traditional social research, but because evaluation takes place within a political and organizational context, it requires group skills, management ability, political dexterity, sensitivity to multiple stakeholders and other skills that social research in general does not rely on as much. Here we introduce the idea of evaluation and some of the major terms and issues in the field.

Educational evaluation is a systematic process of collecting, analyzing, and interpreting data to assess the effectiveness, quality, and impact of educational programs, policies, practices, and interventions. It aims to provide information that can be used to make informed decisions, improve teaching and learning, and enhance educational outcomes for students, educators, and stakeholders. Educational evaluation encompasses a wide range of activities and approaches, including assessment of student learning, evaluation of instructional practices, analysis of curriculum and instructional materials, and assessment of educational policies and initiatives.

6.2 Learning Objectives:

After going through this Unit, learners will be able to:

- Develop an understanding of educational evaluation
- Develop an understanding of functions of evaluation.

6.3 Meaning and Concept Of Educational Evaluation:

A glimpse of the following definitions may reveal the meaning and nature of the term of evaluation**.**

Carter V. Good. "Evaluation is the process of ascertaining or judging the value or amount or something by use of a standard of appraisal."

Stufflebeam and others. "Evaluation is the process of delineating, obtaining and providing useful information for judging decision alternatives"

Remmers, Gage and Rummel. "Evaluation is not just a testing programme. Tests are but one of many different techniques such as observation, checklist, questionnaires, interviews etc. that may contribute to the total evaluation programme."

Torgerson and Adams."To evaluate is ascertain the value of a process or a thing. Thus educational evaluation is the passing of judgment on the degree of worthwhileness of a teaching process or learning experience."

Indian Education Commission (1966)."It is now agreed that evaluation is a continuous process, forms an integral part of the total system of education and is ultimately related to educational objectives. It exercises a great influence on the pupil's study habits and the teacher's method of instruction and thus helps not only to measure educational achievements but also improves it."

An analysis of the above definitions may clearly reveal the following facts about the nature and characteristics of the term evaluation.

- Evaluation is more comprehensive term than measurement or testing.
- It represents continuous process and overall efforts for knowing about the progress of the learner.
- It improves quantitative as well as qualitative description of the outcomes of a teaching-learning process.
- It helps in knowing about the changes in the behavior related to the domains of the learner's behavior as a result of the process of teaching-learning.
- It provides greater scope and flexibility for the use of variety of means and techniques rather than limiting itself to certain tests or conventional examinations.
- It represents a comprehensive plan of better testing and measurement for inquiring into the quality of the output in the light of the set objectives.
- It provides sufficient value judgment about the progress of the learner, teacher's effort and effectiveness of the instructional programmes.

Key components of educational evaluation include;

Purpose and Goals: Clarifying the purpose and goals of the evaluation, including identifying the specific questions or issues to be addressed and the intended uses of evaluation findings

Stakeholder Involvement: Engaging stakeholders, including students, teachers, administrators, parents, policymakers, and community members, in the evaluation process to ensure that their perspectives and priorities are considered.

Data Collection: Collecting relevant data using various methods and instruments, such as surveys, interviews, observations, tests, and document analysis, to gather information on educational processes, outcomes, and impacts.

Data Analysis: Analyzing and interpreting data to identify patterns, trends, strengths, weaknesses, and areas for improvement in educational programs, practices, and outcomes.

Reporting and Dissemination: Communicating evaluation findings to stakeholders in clear, concise, and accessible formats, such as reports, presentations, and dashboards, to inform decision-making and promote accountability and transparency.

Use of Findings: Using evaluation findings to inform decision-making, improve educational practices, allocate resources effectively, and enhance the quality and effectiveness of educational programs and services.

Educational evaluation can be conducted at various levels, including classroom, school, district, state, and national levels, and across different domains, such as academic achievement, instructional quality, school climate, student engagement, and equity and access. It can also encompass formative evaluation, conducted during program implementation to provide ongoing feedback and support, as well as summative evaluation, conducted at the end of a program or initiative to assess its overall effectiveness and impact.

Overall, educational evaluation serves as a valuable tool for promoting continuous improvement, accountability, and evidence-based decision-making in education, ultimately contributing to the enhancement of teaching and learning and the achievement of educational goals and outcomes.

Self-Check Exercise-1

Q.1 What is the primary purpose of educational evaluation?

a) To assign grades to students

b) To improve teaching and learning

c) To rank schools based on performance

d) To increase enrollment rates

Q.2 Who are the key stakeholders involved in educational evaluation?

a) Only teachers

b) Only students

c) Teachers, students, administrators, parents, policymakers

d) Only administrators

Q.3 Which of the following is NOT a component of educational evaluation?

a) Data analysis

b) Test preparation

c) Stakeholder involvement

d) Reporting and dissemination

6.4 Functions Of Educational Evaluation:

There are several advantages of evaluation in the field of education. How much a child knows/ how much success a teacher achieved in his class? How to adjust the teaching for the maximum benefit of students? The answer for all these questions can be answered only through evaluation. Teacher knows about the progress of students only through evaluation. He guides them on this basis. One can be familiar about the problems of students through evaluation and provide them guidance. Similarly, the activities of school can be improved on the basis of evaluation.

Following points may be summarized about the function of evaluation:

- Evaluation helps to classify the students on the basis of their achievement.
- Evaluation helps the teacher to identify the difficulties, problems and weaknesses of the students and provide them their solution.
- Evaluation stimulates the students and gives them the inspiration to study.
- Evaluation gives guidance to the teacher.
- Evaluation helps to make the body and mind disciplined.
- Evaluation clarifies the objectives of teaching.

Educational evaluation serves several key functions in the field of education. Here are some of its primary functions:

Assessment of Learning Outcomes: One of the main functions of educational evaluation is to assess the effectiveness of educational programs, interventions, and practices in achieving desired learning outcomes. This includes evaluating students' academic achievement, skills development, and mastery of content standards or learning objectives.

Improvement of Teaching and Learning: Educational evaluation provides feedback and data to educators, administrators, and policymakers to identify strengths and weaknesses in teaching and learning processes. By analyzing evaluation findings, stakeholders can make informed decisions to improve curriculum, instruction, assessment practices, and educational policies.

Accountability and Quality Assurance: Evaluation helps to ensure accountability and quality assurance in education by assessing the performance and effectiveness of educational institutions, programs, and personnel. It provides evidence to stakeholders, such as parents, taxpayers, and policymakers, about the use of resources, the attainment of goals, and the overall quality of education.

Decision-Making and Policy Development: Evaluation findings inform decision-making and policy development at various levels of the education system, including classroom, school, district, state, and national levels. Policymakers use evaluation data to allocate resources, set priorities, implement reforms, and develop evidence-based policies to improve educational outcomes.

Identification of Needs and Priorities: Educational evaluation helps identify the needs, priorities, and challenges facing students, educators, schools, and communities. By collecting and analyzing data on student performance, demographic trends, school climate, and other factors, stakeholders can target resources and interventions to address specific needs and disparities.

Promotion of Equity and Social Justice: Evaluation plays a critical role in promoting equity and social justice in education by assessing access, participation, and outcomes for all students, particularly those from marginalized or underserved populations. Evaluation findings can highlight disparities in educational opportunities and outcomes and inform efforts to address inequities and ensure equal access to high-quality education for all students.

Professional Development and Capacity Building: Evaluation provides valuable feedback to educators and school leaders for professional development and capacity building. By identifying areas for improvement and best practices, evaluation helps educators enhance their teaching strategies, instructional practices, and classroom management techniques to meet the diverse needs of students effectively.

Monitoring and Evaluation of Educational Policies and Reforms: Evaluation monitors the implementation and impact of educational policies, initiatives, and reforms over time. It assesses the effectiveness, efficiency, and sustainability of reforms, identifies barriers to implementation, and recommends adjustments or modifications to improve outcomes and achieve desired goals.

Overall, educational evaluation serves as a critical tool for promoting continuous improvement, accountability, and evidence-based decision-making in education, ultimately contributing to the enhancement of teaching and learning and the achievement of educational goals and outcomes.

Self-Check Exercise-2

Q.1 How does educational evaluation contribute to accountability in education?

a) By assigning grades to students
b) By promoting continuous improvement
c) By assessing the effectiveness of educational programs
d) By determining teacher salaries

Q.2 Which of the following is NOT a function of educational evaluation?

a) Identifying areas for improvement
b) Promoting equity and social justice
c) Determining teacher salaries
d) Enhancing accountability and transparency

6.5 Principles Of Evaluation:

Principle of evaluation can be described in the following forms:

1. The selection of apparatuses for any work depends on the aim of the work. Similarly, the tools for evaluation cannot be decided until and unless its aim is clear. The selection of apparatus should be according to aims of evaluation.
2. While using the tools for evaluation, it should be kept in mind that it fulfills our aim. There may be several tools available for evaluation, but they differ in aims. Thus, all of them may not be helpful in fulfilling the same aim. Thus, the use of tools for evaluation should be according to aims of evaluation.
3. For complete evaluation of a person or group, different means of evaluation should be adopted, because by one means you cannot obtain the total evaluation of a person or group.
4. The evaluation must have complete knowledge of the utility of tools. What characteristics should be there in a tool for it being good or bad, should properly be examined.
5. One should not take evaluation as an end in itself, but it should be used for other higher aims.
6. While evaluating, the evaluator must keep in mind the moral aspect of evaluation.

Educational evaluation is guided by several key principles that ensure its effectiveness and relevance in informing decision-making, promoting accountability, and improving educational outcomes. Here are some principles of educational evaluation along with their importance:

Validity: Evaluation should measure what it intends to measure accurately. Validity ensures that evaluation methods and instruments assess the intended learning outcomes, instructional practices, or program objectives. Importance: Valid evaluation results provide reliable information for decision-making, program improvement, and accountability.

Reliability: Evaluation should produce consistent and dependable results over time and across different contexts. Reliability ensures that evaluation findings are stable and reproducible, reflecting true changes in educational practices or outcomes rather than random error. Importance: Reliable evaluation data are essential for making informed decisions, monitoring progress, and assessing the effectiveness of interventions.

Fairness and Equity: Evaluation should be fair and equitable, treating all individuals and groups with respect and ensuring equal opportunities for participation and success. Fairness and equity involve addressing biases, cultural differences, and diverse needs in evaluation processes and outcomes. Importance: Fair and equitable evaluation promotes trust, engagement, and inclusivity, enhancing the validity and credibility of evaluation findings.

Transparency: Evaluation should be transparent, with clear objectives, methods, criteria, and procedures communicated to stakeholders. Transparency involves openness, honesty, and accountability in data collection, analysis, interpretation, and reporting. Importance: Transparent

evaluation processes build trust, credibility, and buy-in from stakeholders, fostering collaboration, understanding, and support for evaluation efforts.

Utility: Evaluation should be useful and relevant to stakeholders' needs, providing timely and actionable information to inform decision-making, improve practices, and achieve desired outcomes. Utility involves aligning evaluation goals, methods, and findings with stakeholders' interests, priorities, and goals. Importance: Utilizable evaluation findings drive evidence-based decision-making, resource allocation, and program improvement efforts, maximizing the impact and effectiveness of educational initiatives.

Feasibility: Evaluation should be feasible, considering practical constraints such as time, cost, and available resources. Feasibility involves balancing the rigor and complexity of evaluation methods with the practicalities of implementation and sustainability. Importance: Feasible evaluation approaches ensure that evaluation efforts are realistic, efficient, and sustainable, maximizing the likelihood of successful implementation and uptake.

Ethical Considerations: Evaluation should adhere to ethical principles and standards, protecting the rights, privacy, and confidentiality of participants and stakeholders. Ethical considerations involve obtaining informed consent, maintaining confidentiality, avoiding harm, and ensuring the responsible use of evaluation data. Importance: Ethical evaluation practices safeguard the well-being and dignity of individuals and communities, maintaining trust, integrity, and credibility in evaluation processes and outcomes.

Continuous Improvement: Evaluation should support continuous learning, reflection, and improvement in educational practices and outcomes. Continuous improvement involves using evaluation findings to identify strengths, areas for growth, and opportunities for innovation and adaptation. Importance: Continuous improvement processes foster a culture of learning, innovation, and excellence in education, driving ongoing progress and achievement of desired outcomes.

Overall, adherence to these principles ensures that educational evaluation is rigorous, relevant, ethical, and impactful, contributing to the enhancement of teaching and learning, organizational effectiveness, and student success in educational settings.

Self-Check Exercise-3

Q.1 Which principle of evaluation ensures that evaluation methods and instruments accurately measure what they intend to measure?

a) Validity
b) Reliability
c) Transparency
d) Utility

Q.2 Which principle of evaluation refers to the consistency and stability of evaluation results over time and across different contexts?

a) Validity
b) Reliability
c) Fairness and Equity
d) Transparency

6.6 Summary:

Educational evaluation is a systematic process that plays a crucial role in assessing the effectiveness, quality, and impact of educational programs, policies, practices, and interventions. Its primary purpose is to improve teaching and learning by providing feedback and data to educators, administrators, and policymakers. Through assessment of learning outcomes, evaluation helps ensure accountability, promote equity and social justice, and inform evidence-based decision-making and policy development. Principles such as validity, reliability, fairness, transparency, utility, feasibility, ethical considerations, and continuous improvement guide the evaluation process. Overall, educational evaluation serves as a critical tool for promoting continuous improvement, accountability, and evidence-based decision-making in education, ultimately contributing to the enhancement of teaching and learning and the achievement of educational goals and outcomes.

The generic goal of most evaluations is to provide "useful feedback" to a variety of audiences including sponsors, donors, client-groups, administrators, staff, and other relevant constituencies. Most often, feedback is perceived as "useful" if it aids in decision-making. But the relationship between an evaluation and its impact is not a simple one -- studies that seem critical sometimes fail to influence short-term decisions, and studies that initially seem to have no influence can have a delayed impact when more congenial conditions arise. Despite this, there is broad consensus that the major goal of evaluation should be to influence decision-making or policy formulation through the provision of empirically-driven feedback.

6.7 Glossary:

Educational Evaluation: The systematic process of collecting, analyzing, and interpreting data to assess the effectiveness, quality, and impact of educational programs, policies, practices, and interventions.

Feedback: Information provided to stakeholders based on evaluation findings to guide reflection, learning, and improvement.

Accountability: The responsibility for achieving desired outcomes and using resources effectively, demonstrated through evidence-based evaluation.

6.8 Answers To Self Check Exercises:

Self-Check Exercise-1

Answer1: b) To improve teaching and learning
Answer2: c) Teachers, students, administrators, parents, policymakers
Answer3: b) Test preparation

Self-Check Exercise-2

Answer1: c) By assessing the effectiveness of educational programs

Answer2: c) Determining teacher salaries

Self-Check Exercise-3

Answer1: a) Validity

Answer2: b) Reliability

6.9 References/ Suggested Readings:

- Ebel, Robert L.(1966) “Measuring Educational Achievement, Prentice Hall of India Pvt. Ltd. Pp. 481
- Gronlund, N. E. (1976), Measurement and Evaluation in Teaching. McMillan, USA.
- Hopkins, C.D. and Antes, R.L. (1990).Classroom measurement and evaluation. Itasca, Illinois: Peacock.
- Izard, J. (1991). Assessment of learning in the classroom. Geelong, Vic.: Deakin University.
- Izard, J. (1997). Content Analysis and Test Blueprints. Paris: International Institute for Educational Planning.
- Mehrens, W.A. and Lehmann, I.J. (1984).Measurement and evaluation in education and psychology.(3rd Ed.) New York: Holt, Rinehart and Winston.
- Nandra, I.D.S.(2011). Learning Resources and Assessment of Learning.Patiala, 21st Century Publications.
- Taiwo, Adediran A. (2005). Fundamentals of Classroom Testing. New Delhi: Vikas Publishing House Pvt. Ltd.
- Walter W. Cook (1958). Educational Measurement. Washington D.C.: American Council on Education.
- Withers, G. (1997).Item Writing for Tests and Examinations. Paris: International Institute for Educational Planning.

6.10 Terminal Questions:

Q.1 What is educational evaluation?

Q.2 What are the purposes of educational evaluation?

Q.3 How do educators use data from educational evaluation?

Q.4 What are some challenges in educational evaluation?

UNIT-7 Difference in Measurement and Evaluation

Structure

7.1 Introduction:

Measurement and evaluation are foundational aspects of education, serving distinct yet interconnected purposes. Measurement involves the systematic process of assessing and quantifying various aspects of student learning and educational outcomes. This includes using tools like tests, quizzes, and performance assessments to gather data on students' knowledge, skills, and abilities. The primary goal of measurement is to provide educators with objective information about students' academic progress and achievement, enabling them to track learning trajectories and identify areas needing improvement.

Evaluation, on the other hand, extends beyond measurement by encompassing the broader assessment of educational programs, policies, and practices. It involves analyzing collected data to make informed judgments about the effectiveness, quality, and impact of these educational efforts. Evaluation considers multiple factors such as the alignment of curriculum with learning objectives, the efficiency of instructional strategies, and the overall outcomes achieved by students. By examining both quantitative measures and qualitative insights from stakeholders, evaluation helps educators and policymakers understand what works well and where adjustments are necessary to enhance educational outcomes.

Together, measurement and evaluation play essential roles in shaping educational practices and policies. They provide the foundation for data-driven decision-making, continuous improvement in teaching and learning, and accountability in educational systems. By employing rigorous measurement techniques and comprehensive evaluation frameworks, educators can ensure that their efforts are effective, equitable, and aligned with the goals of providing high-quality education for all students.

7.2 Learning Objectives:

- Explain the principles of evaluation.
- Difference between measurement and evaluation
- Understand current trends in evaluation.

7.3 Meaning of Measurement in Education:

Measurement in education plays a fundamental role in understanding and evaluating students' knowledge, skills, and overall academic progress. In its essence, measurement refers to the systematic process of assessing various aspects of learning and educational outcomes through quantifiable means. This process encompasses a wide array of assessments, evaluations, and data-driven analyses that educators use to gauge students' comprehension, mastery of subjects, and developmental growth over time. Beyond individual student assessments, measurement in education extends to evaluating the effectiveness of educational programs, guiding instructional decisions, and informing policies aimed at enhancing educational outcomes. By providing objective insights into students' learning journeys, measurement not only supports educators in tailoring their teaching strategies but also contributes to fostering an environment conducive to continuous improvement and educational equity. This introduction sets the stage by highlighting the importance of measurement in education, emphasizing its role in both assessing student learning and guiding educational practices for optimal outcomes.

In education, the term "measurement" refers to the process of assessing or evaluating various aspects of learning and educational outcomes. It involves quantifying and describing students' knowledge, skills, abilities, or other characteristics. Here are some key aspects and meanings of measurement in education:

Assessment: Measurement often involves assessing students' knowledge, skills, and understanding. This can include tests, quizzes, projects, portfolios, and other forms of evaluation.

Evaluation: Measurement also includes the evaluation of educational programs, policies, and practices to determine their effectiveness and impact on student learning.

Quantification: It focuses on quantifying aspects of learning, such as scores, grades, ratings, or rankings, to provide objective information about student performance.

Data-driven decision making: Measurement provides data that educators and administrators use to make informed decisions about curriculum, instructional methods, interventions, and resource allocation.

Accountability: Measurement plays a crucial role in accountability systems in education, where student performance data may be used to assess the effectiveness of schools, teachers, and educational policies.

Monitoring progress: Educators use measurement to monitor students' progress over time, identify areas where additional support may be needed, and track the effectiveness of interventions.

Formative and summative purposes: Measurements can be used formatively (to inform instruction and learning) or summatively (to summarize achievement at the end of a unit, course, or program).

Overall, measurement in education is essential for understanding, improving, and ensuring the quality and effectiveness of educational experiences and outcomes for students. It encompasses a range of methods and purposes aimed at fostering learning and growth within educational settings.

Self- Check Exercise-1

Q 1: Which of the following best describes the purpose of measurement in education?

A) To assign grades to students

B) To assess learning and educational outcomes

C) To rank students based on their performance

D) To evaluate teachers' performance

Q 2: What is the primary focus of measurement in education?

A) Monitoring student behavior

B) Assessing physical fitness

C) Evaluating academic achievement

D) Ranking students based on popularity

Q 3: Which term refers to the systematic process of evaluating students' knowledge, skills, and abilities?

A) Assessment

B) Examination

C) Categorization

D) Classification

Q 4: How is measurement used in education to support instructional decisions?

A) By assigning homework

B) By providing feedback to students

C) By informing curriculum adjustments

D) By conducting classroom observations

Q 5: What role does measurement play in educational accountability?

A) Assessing students' hobbies

B) Evaluating teacher salaries

C) Monitoring school performance

D) Assigning extracurricular activities

7.4 Meaning of Evaluation in Education

Evaluation in education is a vital process that serves multiple purposes aimed at improving educational outcomes and enhancing the overall effectiveness of educational programs and practices. At its core, evaluation involves

systematically gathering and analyzing data to assess various aspects of educational endeavors, including student learning outcomes, program effectiveness, and the impact of policies and interventions. Through rigorous assessment methods such as tests, surveys, observations, and qualitative analyses, evaluation provides valuable insights into what students have learned, how well educational programs are meeting their objectives, and where improvements are needed. This feedback loop not only informs instructional decisions and curriculum development but also supports accountability by offering transparent measures of educational quality to stakeholders. Moreover, evaluation in education promotes continuous improvement by identifying strengths to build upon and areas for enhancement, thereby fostering a dynamic learning environment that strives for excellence and equity in education. By integrating evaluation into the fabric of educational practice, institutions can adapt and innovate to meet the evolving needs of students and society effectively.

Evaluation in education refers to the process of systematically assessing various aspects of educational programs, policies, or practices to determine their effectiveness, impact, and overall quality. It involves gathering and analyzing data to make informed decisions aimed at improving educational outcomes and enhancing the learning experiences of students. Here are key aspects and purposes of evaluation in education:

Assessment of Learning Outcomes: Evaluation involves assessing what students have learned and achieved as a result of their educational experiences. This can include measuring academic knowledge, skills development, and understanding of key concepts.

Program Effectiveness: Evaluating educational programs helps determine whether they are achieving their intended goals and objectives. It involves analyzing program components, instructional methods, curriculum design, and resource allocation to assess their effectiveness.

Feedback and Improvement: Evaluation provides feedback to educators, administrators, and policymakers about what is working well and areas that need improvement. This feedback informs decisions about curriculum revisions, instructional strategies, and professional development opportunities for teachers.

Accountability: Evaluation serves as a means of accountability in education, providing stakeholders such as parents, policymakers, and the public with information about the performance of schools, teachers, and educational systems. It ensures transparency and encourages continuous improvement.

Decision-Making: Educational evaluation supports data-driven decision-making processes. By analyzing evaluation results, educators and administrators can make informed decisions about resource allocation, programmatic changes, and interventions to enhance student learning outcomes.

Policy Development: Evaluation findings contribute to the development and refinement of educational policies at local, regional, and national levels. It helps policymakers understand the impact of policies on students, teachers, and schools, guiding future initiatives and reforms.

Continuous Improvement: Evaluation is integral to fostering a culture of continuous improvement in education. By regularly assessing and reflecting on practices and outcomes, educators can adapt and refine their approaches to better meet the needs of diverse learners.

Overall, evaluation in education plays a crucial role in ensuring educational quality, promoting accountability, and driving continuous improvement across educational settings. It involves a systematic approach to gathering, analyzing, and interpreting data to support informed decision-making and enhance student learning outcomes.

Self-Check Exercise-2

Q 1: What is the primary purpose of evaluation in education?
A) Assigning grades to students
B) Assessing learning outcomes
C) Ranking students based on performance
D) Monitoring attendance rates

Q 2: Evaluation in education helps to:
A) Measure student behavior
B) Improve educational outcomes
C) Increase teacher salaries
D) Assign extracurricular activities

Q 3: Which of the following is a key aspect of evaluation in education?
A) Providing homework assignments
B) Analyzing program effectiveness
C) Conducting classroom observations
D) Assigning group projects

Q 4: How does evaluation support decision-making in education?
A) By assigning classroom seating
B) By informing curriculum revisions
C) By scheduling school events
D) By organizing field trips

Q 5: What role does evaluation play in fostering continuous improvement in education?
A) Assigning grades to students
B) Analyzing student hobbies
C) Identifying areas for enhancement
D) Monitoring school facilities

7.5 Difference between Measurement and Evaluation:

Measurement and evaluation are two distinct yet interconnected processes within the realm of education.

Measurement in education refers to the systematic process of assessing or quantifying various aspects of learning and educational outcomes. It involves gathering data through methods such as tests, quizzes, observations, and performance assessments to obtain objective information about students' knowledge, skills, abilities, or other characteristics. The primary purpose of measurement is to provide quantifiable data that educators can use to gauge the extent of student learning and academic achievement. For example, measuring students' scores on standardized tests helps to determine their proficiency in specific subjects or skills. Measurement is crucial for monitoring progress, identifying areas where additional support may be needed, and informing instructional decisions in the classroom.

On the other hand, evaluation in education goes beyond mere measurement. Evaluation is a broader and more comprehensive process that involves the systematic assessment and interpretation of data to make judgments about the quality, effectiveness, and value of educational programs, policies, practices, or outcomes. While measurement focuses on obtaining data and assessing specific aspects of learning, evaluation seeks to understand the overall impact and effectiveness of educational efforts. It considers multiple factors such as student learning outcomes, program goals and objectives, instructional strategies, curriculum design, and the alignment of educational practices with intended outcomes. Evaluation also incorporates qualitative aspects, such as stakeholders' perceptions and feedback, to provide a holistic understanding of educational effectiveness.

In summary, measurement provides the quantitative data necessary to assess specific aspects of student learning and achievement, while evaluation involves interpreting these data within a broader context to make informed judgments about educational quality, effectiveness, and improvement. Both measurement and evaluation are essential components of educational assessment and accountability, contributing to ongoing efforts to enhance learning outcomes and educational experiences for students.

S.N.	**Measurement**	**Evaluation**
1	It refers to the status (quantity and magnitude) of achievement of the student. It only answers the question “how much”. It is thus quantitative.	It refers to the value judgment (the quality or worth) we attach to the magnitude or quantity of the achievement of the student. It answers the question “what value that how much is”. It is thus qualitative.
2	At a time, measurement involves a single aspect of	Evaluation is more comprehensive concept. It includes achievements

	achievement, a single variable i.e. mathematical ability or scientific skill etc. it is limited in scope.	but also the attitudes, interests, ideals, ways of thinking, work habits, personal and social adaptability. So its scope is very wide.
3	It is objective.	It is subjective, by and large.
4	It is static, at the most periodical.	It is a continuous, never ending process.
5	It is absolute.	It is a relative.
6	It has true zero, as starting reference point. A true zero point means complete absence of the trait being measured.	There is only arbitrary zero point in evaluation. By arbitrary zero point is meant appoint which does not represent the absence of trait being measured.
7	It does not look before and after.	It takes into account the past achievements and also the future goals.
8	The unit of measurements is fixed and constant throughout the measurement.	The unit of measurement is not fixed and varies during the process of evaluation.
9	It is more accurate and predictable because of its true zero point.	It is less accurate and predictable because of its arbitrary zero point.
10	It is direct.	It is indirect.

Self-check Exercise-3

Q1. Which of the following best describes measurement?

A) Making judgments based on criteria

B) Assigning numbers or symbols to attributes

C) Evaluating quality and effectiveness

D) Interpreting data within a broader context

Q2. What is the primary purpose of measurement?

A) To make informed decisions

B) To describe and quantify attributes

C) To assess quality and value

D) To interpret data within a context

Q3. Which term involves making judgments based on criteria or standards?

A) Measurement

B) Evaluation

C) Quantification

D) Objectivity

Q4. In which process is objectivity and replicability most important?

A) Measurement

B) Evaluation

C) Quantification
D) Interpretation
Q5. What is a characteristic of evaluation that distinguishes it from measurement?
A) Focus on quantification
B) Use of standardized methods
C) Subjective judgment
D) Collecting raw data

7.6 Summary

Measurement and evaluation are integral processes in various disciplines, each serving distinct purposes and employing different methodologies. Measurement involves the systematic assignment of numbers or symbols to attributes or characteristics according to standardized rules. It focuses on quantifying variables such as test scores in education, physical properties like temperature in science, or performance metrics in business. The primary goal of measurement is to provide objective, replicable data that describes or categorizes these attributes.

In contrast, evaluation goes beyond measurement by making subjective judgments or assessments based on established criteria or standards. It involves interpreting the measured data within a broader context to assess quality, effectiveness, or value. Evaluations are often used to make informed decisions, recommendations, or comparisons, such as evaluating the success of educational programs, assessing employee performance, or reviewing healthcare outcomes. Unlike measurement, which deals with collecting and quantifying data, evaluation focuses on the qualitative interpretation and significance of that data to guide actions and improvements.

Together, measurement and evaluation form a critical cycle in fields such as education, healthcare, business, and social sciences, providing essential tools for understanding, decision-making, and continuous improvement. While measurement lays the groundwork by quantifying variables objectively, evaluation applies subjective judgment and context to draw meaningful conclusions and drive informed actions.

7.7 Glossary

- **Measurement** provides raw data or descriptions using standardized methods, focusing on objectivity and quantifiability.
- **Evaluation** interprets data within a broader context, involves subjective judgment, and aims to assess quality or effectiveness based on criteria or standards.

7.8 Answers to Self-Check Exercises

Exercise-1

Answer 1. B) To assess learning and educational outcomes
Answer 2. C) Evaluating academic achievement
Answer 3. A) Assessment
Answer 4. C) By informing curriculum adjustments
Answer 5. C) Monitoring school performanceExercise-2

Exercise-2

Answer 1. B) Assessing learning outcomes
Answer 2. B) Improve educational outcomes
Answer 3. B) Analyzing program effectiveness
Answer 4. B) By informing curriculum revisions
Answer 5. C) Identifying areas for enhancement

Exercise -3

Answer 1. B) Assigning numbers or symbols to attributes
Answer 2. B) To describe and quantify attributes
Answer 3. B) Evaluation
Answer 4. A) Measurement
Answer 5. C) Subjective judgment

7.9 References/ Suggested Readings

- Ebel, Robert L.(1966) “Measuring Educational Achievement, Prentice Hall of India Pvt. Ltd. Pp. 481
- Gronlund, N. E. (1976), Measurement and Evaluation in Teaching. McMillan, USA.
- Hopkins, C.D. and Antes, R.L. (1990).Classroom measurement and evaluation. Itasca, Illinois: Peacock.
- Izard, J. (1991). Assessment of learning in the classroom. Geelong, Vic.: Deakin University.
- Izard, J. (1997). Content Analysis and Test Blueprints. Paris: International Institute for Educational Planning.
- Mehrens, W.A. and Lehmann, I.J. (1984).Measurement and evaluation in education and psychology.(3rd Ed.) New York: Holt, Rinehart and Winston.
- Nandra, I.D.S.(2011). Learning Resources and Assessment of Learning.Patiala, 21st Century Publications.
- Taiwo, Adediran A. (2005). Fundamentals of Classroom Testing. New Delhi: Vikas Publishing House Pvt. Ltd.
- Walter W. Cook (1958). Educational Measurement. Washington D.C.: American Council on Education.
- Withers, G. (1997).Item Writing for Tests and Examinations. Paris: International Institute for Educational Planning.

7.10 Terminal Questions

1. Explain the concept of measurement. How does measurement differ from simple observation or description? Provide examples to illustrate your explanation.

2. Describe the process of evaluation. How does evaluation differ from measurement? Provide examples of evaluation in different fields to illustrate your explanation.
3. Discuss the importance of measurement and evaluation in decision-making processes. Provide examples to illustrate how these processes contribute to improving outcomes in various fields.

UNIT-8 Current Trends in Evaluation-I

Structure

8.1 Introduction

In recent years, evaluation practices have evolved significantly, driven by emerging trends that emphasize effectiveness, relevance, and adaptability across various fields. One prominent trend is Utilization-Focused Evaluation (UFE), which prioritizes stakeholder engagement throughout the evaluation process to ensure that findings are useful and actionable for decision-making and program improvement. There's also a notable shift towards Outcome and Impact Evaluation, which goes beyond traditional output measures to assess broader outcomes and societal impacts using mixed-method approaches for a more comprehensive understanding.

8.2 Learning Objectives

After completing this unit, the learners will be able to;

- Know about the various trends in evaluation
- Understand the Grading System of Evaluation
- Understand the Open Book Examination System

8.3 Grading System of Evaluation

Grading in education is the process of applying standardized measurements of varying levels of achievement in a course. Another way the grade point average (GPA) can be determined is through extra-curricular activities. Grades can be assigned as letters (generally A through F), as a range (for example 1 to 6), as a percentage of a total number of questions answered correctly, or as a number out of a possible total (for example out of 20 or 100). In some countries, all grades from all current classes are averaged to create a GPA for the marking period. The GPA is calculated by taking the number of grade points a student earned in a given period of time of middle school through high school. GPAs are also calculated for undergraduate and graduate students in most universities. The GPA can be used by potential employers or educational institutions to assess and

compare applicants. A cumulative grade point average is a calculation of the average of all of a student's total earned points divided by the possible amount of points. This grading system calculates for all of his or her complete education career.

Grading System in Evaluation

- When students' level of performance are classified into a few classificatory units using letter grades, the system of assessment is called grading system.
- In grading, classification is made on 5 point, 7 point, 9 point scale etc.
- Fundamentally a 'grade' is a score.
- Grading considered to be more scientific way of evaluation as it solves the problems of borderline cases by including all the pupils falling within a wider range with in the same grade.
- Identifies the students' performance level within a wide range.
- Grading involves the use of set of specialized symbols whose meaning ought to be objectively defined in terms of specific criteria.
- **In 5 point scale grading system the categories are**

Score %	**Letter grade**	**interpretation**
75 and above	A	Outstanding
60 – 74	B	Very good
45 – 59	C	Good
30 – 44	D	Average
Below 30	E	Below average

- **In 9 point scale grading system the categories are**

Score %	**Letter grade**	**interpretation**
90 – 100	A+	Outstanding
80 – 89	A	Excellent
70 – 79	B+	Very good
60 – 69	B	Good
50 – 59	C+	Above average
40 – 49	C	Average
30 – 39	D+	Marginal
20 – 29	D	Need improvement
Below 20	E	Need improvement

Types of Grading

Direct Grading

- Judgment of any given phenomenon (achievement, skills, personality traits etc.) by the evaluator in terms of most appropriate letter grade without assigning the scores.
- In examinations, the evaluator will award a particular grade to the answer for each individual question on the basis of its quality.
- Then the Grade Point Average (GPA) calculated for obtaining the overall grade of the student in a particular subject.

Letter grade	A	B	C	D	E	F	G	H	I
Numerical Value of the Grade	9	8	7	6	5	4	3	2	1

Calculation of GPA

Q.No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Obtained Grade	F	B	A	G	D	H	E	E	B	F	C	B	E	D	B
Grade Point	4	8	9	3	6	2	5	5	8	4	7	8	5	6	8

$$GPA = \frac{4 + 8 + 9 + 3 + 6 + 2 + 5 + 5 + 8 + 4 + 7 + 8 + 5 + 6 + 8}{15}$$

= 5.86 = 6*Hence Overall Grade will be 'D'*

Direct Grading

- Massive and time consuming.
- Mistakes can't be ruled out.
- Suitable for small groups.
- Not feasible for large-scale examinations.

Indirect Grading

- Method of awarding grades through marks.
- Marks are awarded to the individual questions on the basis of the prescribed marking scheme.
- Find out total mark and converted to grade.

➢ This can be done in two ways: Absolute and Relative

Absolute Grading

- Some fixed ranges of scores are determined in advance for each grade.
- The score obtained by a student in a subject is converted into corresponding grade.eg.

Score %	Letter grade	interpretation
90 – 100	A+	Outstanding
80 – 89	A	Excellent
70 – 79	B+	Very good

60 – 69	B	Good
50 – 59	C+	Above average
40 – 49	C	Average
30 – 39	D+	Marginal
20 – 29	D	Need improvement
Below 20	E	Need improvement

Relative Grading

- The grade range is not fixed in advance.
- The fundamental premise of relative grading is that if an evaluation result is plotted on a graph, the graph will be assumes the form of a Normal Probability Curve

Procedure to find out Relative Grading

- Group of students are divided into different groups.
- Find out number of students in each grade using statistical methods.
- Using this number of students in each grade, find out the range of each grade.
- Corresponding grades are given to each students who fall in different ranges.

Example: Suppose there are 50 students.

- Marks in 50

1. 45 45 44 43 42 42 41 39 38
2. 35 35 33 32 31 30 30 30 29
3. 28 27 27 27 26 25 25 24 24
4. 23 23 22 22 21 21 21 20 20
5. 18 18 18 18 11 10 9 8 7

Let the group is divided into 5 Grades (A, B, C, D, E)

Find out number of students in each grade using statistical methods.

- A Grade 5 students
- B Grade 10 students
- C Grade 20 students
- D Grade 10 students
- E Grade 5 students

Using this number of students in each grade, find out the range of each grade.

Then Range of Each Grade are

- 50 – 43 A Grade
- 42 – 32 B Grade
- 31 – 22 C Grade
- 21 – 18 D Grade
- Below 18 E Grade

Grading system in education in India has some integrated features that help students, teachers, and parents build uniformity and advanced educational ground. They are:

1. It reduces examination-related pressure.
2. It assesses a student's performance based on their results and the given assignments, projects, and depth of knowledge. This process keeps students motivated and affects their studies.
3. Students are categorized into different grades, so they can quickly rectify where they are lacking and where they are strong.
4. Teachers can identify the average and below-average students separately. Also, they allow improving and working on the students' rectified weaknesses.
5. Parents get opportunities to understand their child's weaknesses and correct them with care and proper guidance. Also, this grading system finds the aspects that parents need to focus on.
6. It enhances uniformity and helps children to upgrade themselves and perform according to the grading assessment.

Self-Check Exercise-1

Q.1 In a point grading system, what do points typically represent?
A) Achievement levels
B) Relative performance
C) Specific criteria
D) Cumulative scores

Q2. How are grades determined in a point grading system?
A) Based on a predetermined distribution
B) By comparing students' scores
C) According to the total points earned
D) Using rubrics and standards

Q3. In a course using a point grading system, a student earns 85 points out of 100. What grade would this typically correspond to?
A) A
B) B
C) C
D) D

Q4. What advantage does a point grading system offer over other grading methods?
A) Allows for easy comparison of student performance
B) Provides subjective assessment of student work
C) Focuses on improvement over time
D) Incorporates holistic evaluation

Q5. How does a point grading system facilitate transparency in grading?
A) By using standardized criteria
B) By assigning letter grades directly
C) By involving peer evaluation

D) By including formative assessments

8.4 Open Book Examination System

Advanced countries are restructuring their educational systems and preparing to make changes in the human to intellectual capital for meeting the economic and manpower demands of the 21st century. Educationalist and academicians opined it is necessary to build a young community who are able to think independently and creatively and able to process the information analytically. Scholars opined that in order to develop creative and independent thinkers, more open-ended tasks that reflect real-life situations – questions involving problem-solving should be incorporated in examinations. It is generally accepted that open book examinations create an enriched environment, offering the student an opportunity to better understand and respond to a particular question. The Central Board of Secondary Education's (CBSE) proposal for an open-book exam from 2013 within its class X and XII board assessments is appreciated. It makes a better option than the present one which reflects a fresh, borne in mind innovative spirit, reshape education in India.

An "open book examination" is one in which examinees are allowed to consult their class notes, textbooks, and other approved material while answering questions. The conventional memory testing examinations must give way to examinations that test the intellectual skills of the student. This is where open book examinations come in. An open book question provides the candidates with the theory the question is examining and then asks them to demonstrate their ability to apply the theory to a scenario.Teaching is transmitting Information in conventional system and tests how much information the students have been able to store in their minds. On the contrary teaching should equip students with the ability to acquire knowledge, to modify existing knowledge on the basis of new experience, to build new knowledge, and to apply available knowledge to solve problems and make intelligent decisions. The main focus of teaching will be on the skills of acquiring, modifying and creating knowledge, that is, on processing information, rather than on the informationcontent itself.Suppose the examination consists of information based questions like, "Who invented the theory of relativity?", "Explain the term 'Standard Deviation' etc., students can then easily find the answers in the textbooks or notes, and copy them in their answer books. In an open book examination, it is meaningless to ask questions like above.The essential difference between closed book examinations and open book examinations is that the former can still be used to evaluate how much the students have memorized, while the latter cannot. On the other hand, if the examination tests the skills of problem solving and critical thinking, then there is no harm in students consulting their text books and class notes. Students cannot copy anything from the textbook. The open-book examination measuring high-level skills such as conceptualizing, problem solving and reasoning corresponds

almost completely to a total, real-world situation, less stressful, less memorization, more room for logical thinking and more room for creative thinking.A paradigm shift necessitate in teaching and learning process from the prescriptive to the provocative, from providing knowledge to empowering students to acquire knowledge, from furnishing students with a finite body of knowledge to equipping them with higher order cognitive skills and from dependency to taking ownership of the learning experience.The teachers should no longer focus on providing 'the right answers' and being the final arbiter of truth, but rather serve as facilitator, trainer, consultant and assessor. The focus will shift from teaching a body of information to teaching how to process and apply it. Teaching and learning must be more interactive throughout the training which facilitates the students to think themselves and ask questions in turn and to acclimatize them to diversity in responses.

Types of Open Book Examination:

One may think of two kinds of open book examinations, say the *restricted* type and the *unrestricted* type. In the restricted type of open book examinations, students are permitted to bring into the examination room one or more specific documents approved by the course instructor. In the unrestricted type of open book examinations, students are free to bring whatever they like.

In the restricted open book examination, students may be permitted to consult printed documents such as the logarithmic tables, dictionaries, or complete works of Shakespeare, but no handwritten material or printed documents which have not had prior approval. One may also need to make sure that the printed documents that students bring do not contain any scribbles on the margin. In this type of examination, the approved documents function more or less as *appendices to the question paper itself.* These examinations are not radically different from closed book examinations. They do not present any special problems, irrespective of the nature of the course.

As mentioned earlier, there are no restrictions on what the students can bring in an unrestricted open book examination. They may bring any books (with or without scribbles on the margin), lecture handouts of the course instructor, or their own handwritten notes. The use of such examinations presupposes certain teaching strategies and types of questions. In particular, it demands that the course focuses on a set of *intellectual skills,* rather than on the *information content,* and that no content based questions be asked in the examination. If the course instructor has concentrated on handing down currently available knowledge, and the question paper contains traditional content based questions like "Write an essay on the difference between British and American English", the use of the unrestricted open book examinations would be disastrous.

When used properly, it will be pointless for students taking the unrestricted open book examinations to consult any material they have brought, because the questions will be designed in such a way the answers will not be found in the textbooks, handouts or class notes. An intelligent student who has had the

experience of such examinations once will not bother to bring anything for the next examination, since (s) he will know that no prepared material will be of any use. The use of these examinations then acts as symbolic gesture that makes the students realize the nature of the course and the examinations, and shocks them into a mode of studying that does not involve cramming.

Impact on Learning Strategies

A more important reason for using open book examinations is that they have a tremendous impact on promoting the right mental sets in both learning and teaching. The most immediate result on students will be that they will stop "mugging" or rote learning. Most students used to conventional examinations think of "studying" as the mechanical memorization of information in textbooks and class notes in order to reproduce it in examinations. Open book examinations will effect a fundamental change in this attitude. If textbooks can be consulted in the examination rooms, why bother to memorize them?

Does this mean that students don't need to "study" for examinations? No. It implies that studying should not be equated with memorizing; instead, it should be understanding concepts, and using these concepts (along with available information) to practice the skills of modifying and building knowledge, thinking critically, and solving problems. In acquiring the right strategies of studying, nothing is as effective as the shocking realization that mugging is of no use in the examinations. Given open book examinations, there will be no more mugging. Once the burden of mugging is taken away, education can be a pleasurable activity, not a painful drudgery. What is learnt with pleasure is learnt more effectively, and retained better.

Impact on Teaching Strategies

The effects of open book exams on teaching strategies will be equally profound. First, the nature of the examination questions will change. They cannot be of the form: "Write an essay on X", "Explain the term Y with examples", "Define the term Z", but will have to be designed carefully and intelligently to test the students' understanding, and the skills of applying that understanding.

If the nature of the examination questions changes, strategies for preparing students to take those examinations will also have to change. It will no longer be enough to paraphrase or simplify the content of the text books in the classroom. Teachers will have to design tasks that will provide exercises for the appropriate mental skills required in each subject. Instead of the teacher talking all the time and students taking down notes, classes will have discussions, questions, and other active processes. In other words, teaching will no longer be the transfer of information from the teacher to student: it will be the training of the mind in certain intellectual skills.

Thus, open book examinations can restore the true meaning of the word education for both teachers and students. It is true that *it will take some time and effort on the part of students and teachers to adapt themselves to the demands of open book examinations.* But the changes will be inevitable.

When combined with the mode of teaching that focuses on thinking skills, they will make education an exciting and enjoyable intellectual adventure, the beginning of a lifelong quest for knowledge.

Self- Check Exercise-2

Q.1 What is a defining feature of an open book examination?

A) Students are allowed to bring textbooks and notes.

B) Exams are conducted in a large hall with multiple students.

C) Students must rely solely on memory and recall.

D) Exams are conducted without any supervision.

Q2. In an open book examination, what is the primary purpose of allowing students to refer to books and notes?

A) To test memory retention

B) To encourage collaboration

C) To assess application of knowledge

D) To limit resources during the exam

Q3. How does an open book examination differ from traditional closed-book exams?

A) It requires less preparation

B) It focuses on problem-solving skills

C) It allows for resource use during the exam

D) It limits the time students have to complete the exam

Q4. What skill does an open book examination primarily assess in students?

A) Memorization

B) Critical thinking

C) Speed and accuracy

D) Test-taking strategies

Q5. Which of the following is a potential challenge of using open book examinations?

A) Encourages deeper understanding of the subject

B) Requires careful design of questions

C) Minimizes reliance on textbooks

D) Decreases exam duration

8.5 Summary

Grading systems and open book examinations are two distinct approaches in educational assessment, each serving specific purposes and catering to different aspects of learning and evaluation.

Grading systems, such as norm-referenced and criterion-referenced grading, provide frameworks for assigning scores or grades to students based on their performance in various assessments. Norm-referenced grading compares students' performance against each other, often using a predetermined distribution curve to assign grades, while criterion-referenced grading assesses students based on their achievement of specific learning objectives

or standards. These systems help educators gauge student progress, provide feedback, and make decisions about educational outcomes and interventions. They emphasize different aspects of student achievement, from relative performance to mastery of content and skills.

In contrast, open book examinations allow students to refer to textbooks, notes, or other resources during the test. This approach aims to assess students' ability to apply knowledge, analyze information, and solve problems rather than relying solely on memorization. Open book exams encourage critical thinking and the application of learned concepts in practical scenarios, reflecting real-world situations where access to information is readily available. However, designing effective open book exams requires careful consideration of question types and assessment criteria to ensure that students demonstrate deep understanding and effective use of resources without compromising the integrity of the assessment process.

In conclusion, while grading systems provide structured methods for evaluating student performance across various metrics, open book examinations offer opportunities for students to demonstrate higher-order thinking skills and practical application of knowledge in a supportive learning environment. Both approaches play integral roles in assessing and promoting learning outcomes, catering to different educational goals and contexts.

8.6 Glossary

Grading System: A method used to assess and assign scores or grades to students based on their performance in various assessments. Types include norm-referenced grading (compares students' performance against each other) and criterion-referenced grading (assesses based on specific learning standards).

Open Book Examination: An assessment format where students are allowed to refer to textbooks, notes, or other resources during the exam. It tests students' ability to apply knowledge, analyze information, and solve problems with access to reference materials.

8.7 Answers to Self-Check Exercises

Exercise-1

Answer1. C) Specific criteria
Answer2. C) According to the total points earned
Answer3. A) A
Answer4. A) Allows for easy comparison of student performance
Answer5. A) By using standardized criteria

Exercise-2

Answer1. A) Students are allowed to bring textbooks and notes.
Answer2. C) To assess application of knowledge
Answer3. C) It allows for resource use during the exam

Answer4. B) Critical thinking
Answer5. B) Requires careful design of questions

8.8 References/ Suggested Readings

- Ebel, Robert L.(1966) “Measuring Educational Achievement, Prentice Hall of India Pvt. Ltd. Pp. 481
- Gronlund, N. E. (1976), Measurement and Evaluation in Teaching. McMillan, USA.
- Hopkins, C.D. and Antes, R.L. (1990).Classroom measurement and evaluation. Itasca, Illinois: Peacock.
- Izard, J. (1991). Assessment of learning in the classroom. Geelong, Vic.: Deakin University.
- Izard, J. (1997). Content Analysis and Test Blueprints. Paris: International Institute for Educational Planning.
- Mehrens, W.A. and Lehmann, I.J. (1984).Measurement and evaluation in education and psychology.(3rd Ed.) New York: Holt, Rinehart and Winston.
- Nandra, I.D.S.(2011). Learning Resources and Assessment of Learning.Patiala, 21st Century Publications.
- Taiwo, Adediran A. (2005). Fundamentals of Classroom Testing. New Delhi: Vikas Publishing House Pvt. Ltd.
- Walter W. Cook (1958). Educational Measurement. Washington D.C.: American Council on Education.
- Withers, G. (1997).Item Writing for Tests and Examinations. Paris: International Institute for Educational Planning.

8.9 Terminal Questions

Q.1 Compare and contrast norm-referenced and criterion-referenced grading systems. How do these systems differ in their approach to assessing student performance?

Q.2 Discuss the advantages and disadvantages of using an open book examination format in educational assessment. How does this format impact student learning and assessment outcomes?

UNIT-9 Current Trends in Evaluation-II

Structure

9.1 Introduction

Self-evaluation and online evaluation systems are integral components of modern assessment practices, offering individuals and organizations effective tools for monitoring progress, achieving objectives, and promoting continuous improvement in diverse domains. They represent advancements in leveraging technology to enhance evaluation processes, adapt to evolving educational and professional needs, and support lifelong learning journeys.

9.2 Learning Objectives

After completing this unit, the learners will be able to;

- Know about the various trends in evaluation
- Understand the self System of Evaluation
- Understand the Online Examination System

9.3 Self- Evaluation

An individual's assessment of himself, his possibilities, his qualities, and his place among other people. As part of the nucleus of the personality, self-evaluation is a very important regulator of behavior. A person's self-evaluation determines his relationships with others, his critical faculties, the demands he make on himself and his attitude towards success and failure. Self-evaluation is linked with the level of the person's aspirations, or the degree of difficulty involved in attaining his goals. A discrepancy between a person's aspirations and his actual capacities leads to an incorrect self-evaluation, as a result of which his behavior may become inappropriate. There may be emotional outbursts and increased anxiety. Self-evaluation is also objectively expressed in a person's evaluation of the possibilities and consequences of the activity of others.

Soviet psychologists have demonstrated the influence of self-evaluation on cognitive activity in man (perceptions, ideas, and the solution of intellectual problems) In addition; Soviet psychologists have defined techniques of

forming adequate self-evaluations and methods of transforming distorted ones by exerting educational influence on the personality.

A self-evaluation system is a structured approach that individuals use to assess their own performance, skills, behaviors, or achievements against established criteria or goals. It is a reflective process that encourages self-awareness, personal development, and accountability. Here's an exploration of its features, advantages, and implementation aspects:

Features of Self-Evaluation System:

Criteria or Standards: Self-evaluation systems typically define clear criteria or standards against which individuals assess themselves. These criteria may include specific competencies, performance metrics, behavioral expectations, or personal goals.

Reflective Process: It involves introspection and reflection on one's strengths, weaknesses, accomplishments, challenges, and areas for improvement. This introspective process encourages individuals to critically analyze their own performance and behaviors.

Goal Setting: Self-evaluation often includes setting new goals or revisiting existing ones based on the assessment findings. This aspect is crucial for personal and professional growth, as it helps individuals align their efforts with desired outcomes.

Feedback Mechanism: Some self-evaluation systems incorporate feedback from others, such as supervisors, peers, or mentors, to provide additional perspectives and insights into performance.

Action Planning: Based on the assessment results, individuals develop action plans to address identified areas for improvement. These plans outline specific steps, timelines, and resources needed to achieve goals.

Advantages of Self-Evaluation System:

Promotes Self-Awareness: It enhances individuals' understanding of their strengths, weaknesses, and development areas, fostering self-awareness and mindfulness.

Encourages Ownership and Accountability: Self-evaluation empowers individuals to take ownership of their own development and accountability for their actions and outcomes.

Supports Personalized Development: Unlike external evaluations, self-evaluation allows individuals to tailor their development plans and goals according to their unique strengths and needs.

Facilitates Continuous Improvement: By regularly assessing and reflecting on performance, individuals can continuously improve their skills, behaviors, and performance over time.

Flexible and Accessible: Self-evaluation systems can be implemented flexibly, allowing individuals to conduct assessments at their own pace and convenience. They can also be adapted to various contexts and environments.

Implementation of Self-Evaluation System:

Establish Clear Objectives: Define the purpose and objectives of the self-evaluation system, such as professional development, performance improvement, or personal goal setting.

Define Evaluation Criteria: Clearly articulate the criteria, standards, or competencies against which individuals will assess themselves. These criteria should be specific, measurable, achievable, relevant, and time-bound (SMART).

Provide Guidance and Resources: Offer guidelines, templates, or tools to support individuals in conducting self-assessments effectively. This may include self-assessment forms, reflection prompts, or goal-setting frameworks.

Encourage Reflection and Feedback: Promote a culture of reflection and constructive feedback where individuals feel comfortable and empowered to assess themselves honestly and seek input from others when needed.

Monitor Progress and Follow-Up: Regularly review and monitor individuals' progress toward their goals. Adjust action plans as necessary based on evolving needs, achievements, or challenges encountered.

Celebrate Success and Learning: Acknowledge and celebrate successes and milestones achieved through the self-evaluation process. Encourage a learning mindset where failures or setbacks are seen as opportunities for growth and improvement.

Self-Check Exercise-1

Q.1 What is the primary goal of a self-evaluation system?

A) To compare performance with peers
B) To assess others' performance
C) To enhance self-awareness and accountability
D) To delegate evaluation tasks

Q.2 What does a self-evaluation system typically involve?

A) External assessment by supervisors
B) Introspection and reflection
C) Automated grading processes
D) Group-based evaluations

Q3. Which aspect of self-evaluation helps individuals identify areas for improvement?

A) Goal setting
B) Reflective process
C) Feedback from supervisors
D) Performance comparisons

Q4. How does a self-evaluation system promote continuous improvement?

A) By delegating assessment tasks
B) By emphasizing external benchmarks
C) By encouraging goal setting and action planning
D) By limiting feedback opportunities

Q5. What advantage does a self-evaluation system offer over external evaluations?

A) Provides diverse perspectives
B) Enhances objectivity
C) Promotes personalized development
D) Reduces individual accountability

9.4 Online Examination System (OES)

Online Examination System (OES) is a Multiple Choice Questions (MCQ) based examination system that provides an easy to use environment for both Test Conductersand Students appearing for Examination. The main objective of OES is to provide all the features that an Examination System must have, with the "interfaces that doesn't Scare it's Users!"

Taxonomy of OES

Users of OES are classified into three categories: Administrators, Test Conductors and Students. Administrators are responsible for management of system users, subjects, tests, questions, results, system backup and recovery, etc. Test conductors are responsible for preparing subjects, tests and questions. Students are the candidates appearing for the Examination.

Features provided in Latest Release are:

- Supports Management of Users, Subjects, Tests, Questions and Results.
- Fully Automated Evaluation and Results Calculation.
- Provides detailed information to Test Conductors.
- Provides test summary, results summary to both student and test conductor.

Advantages of Online Examination

Online examination systems seek to efficiently evaluate the exam partakers thoroughly through a fully automated system that not only saves time but also give fast results. The Online examination system helps to completely automate the old manual procedure of conducting exams. Usually it is done through a Web Based Online Examination Software or and Intranet variant. It also significantly eliminates the need for monitoring while the exam is being taken. All instructions are displayed to the exam taker before the tests begin.

To effectively deliver an examination, 3 major components have to be catered for efficiently.

They are:

1. Creation of exams – Obviously, an exam will have to be created. Examiners can create exams online. The contents also have to be kept securely until the examination starts.
2. Supervision of examination – Students have to be efficiently identified, and screened to ensure that they so not compromise the standards of the exams.

3. Marking of the exams. Marking is the ultimate stage in any examination as it determines the success or failure of the candidate. It is the stage that dictates the next level of success and achievement in life.

A major highlight of using a web based exam software or an online examination system is that it gives a high level of transparency as opposed to the traditional method or remote method. It is almost impossible to compromise exam questions and evaluations because they cannot also be influenced. Most online exams generate their results instantly and it is often possible for the exam taker to get information on his results immediately. Some of the major advantages of online examination include the following:

- **Security and confidentiality**

As has been stated before, the security and confidentiality of an exam are critical if the exam is to retain its value. Prepared exams need to be securely kept. Any leakage will definitely compromise the standard of the exam and may result to a cancelation or a retake. All these features are well addressed using an online system because not only is the content of the exam safely locked away in a database, access to the database is only possible with an authorized personnel. A lot of possibilities are also opened up on exam day as it allows you to conveniently make your own test in a secure environment. Questions can easily be randomized so that no participants see the same questions in the same order. Questions can easily be mixed as each new question is added to the system's database. The questions can then be randomly drawn from the database. This is why most examination bodies have gradually adopted the online platform.

- **Accessibility and Flexibility**

Exams can be conducted anywhere. All a student needs is a personal computer with internet connection. A student does not need a long commute to exam venue as long as these requirements are met. This also means that thousands of students can take the same exams over a wide spread of locations. Gone are days, students move from one region or locality to another in the name of sitting the examination. A lot of lives have been lost in the process due to accident and other similar mishaps. In the same vein, examiners also benefit from this. Examiners do not have to bother with the laborious task of marking exams as this is well taken care of by the system. The system actually marks each exam and presents the result to the student at the end of the exam. Examiners are also afforded the opportunity to create exam online through an online examination system that can present examinations in multiple languages. Multiple exams on multiple subjects for multiple courses can also be set. Exams can also be configured for 24/7 availability. This allows candidates to take exams at their own convenience.

- **Cost saving**

When an exam is placed online, it results in significant cost savings. The cost of paper, copying, and distribution expenses are all reduced or eliminated.

The elimination of paper costs alone is extraordinary. The copying and distribution of assignments to a large class are often unwieldy and inefficient. Administrators anxious to reduce expenditures are likely to strongly favor the transition from paper assignments examinations to using online assessment software.

- **Time management**

Online examination systems make use of computers that helps in saving time. With the widespread availability of computers and the internet, there is a general acceptability and endorsement of this system. The lengthy formalities and processes involved in creating question papers, registering candidates for exams, answer sheet evaluation and declaration of results are completely eradicated with the online exam system. Each student is timed precisely and all results are generated instantaneously. In some cases, a candidate may even be able to get an assessment on failed questions.

- **Statistical analysis**

On a statistical point of view, compared data can be stored over time. This means that different comparative analysis can be done to analyze the outcomes of exams overtime. Depending on the online exam system used, statistical data can be pulled to analyze different data and create reports.

Self- Check Exercise-2

Q.1 What is the primary advantage of using an online evaluation system?
A) Limited accessibility
B) Automated scoring
C) Manual grading
D) Paper-based assessments

Q2. How do online evaluation systems facilitate assessment administration?
A) By restricting access to assessments
B) By allowing assessments only in-person
C) By conducting assessments remotely
D) By eliminating assessment security

Q3. What feature of online evaluation systems enhances assessment reliability?
A) Remote access
B) Real-time reporting
C) Automated scoring
D) Manual data entry

Q4. Which aspect of online evaluation systems supports adaptive testing?
A) Automated grading
B) Secure test delivery
C) Data analytics
D) Customizable assessment content

Q5. What role does data analytics play in online evaluation systems?
A) Automating assessment scheduling

B) Providing insights into assessment performance
C) Restricting access to assessment results
D) Conducting formative assessments

9.5 Summary

The generic goal of most evaluations is to provide "useful feedback" to a variety of audiences including sponsors, donors, client-groups, administrators, staff, and other relevant constituencies. Most often, feedback is perceived as "useful" if it aids in decision-making. But the relationship between an evaluation and its impact is not a simple one -- studies that seem critical sometimes fail to influence short-term decisions, and studies that initially seem to have no influence can have a delayed impact when more congenial conditions arise. Despite this, there is broad consensus that the major goal of evaluation should be to influence decision-making or policy formulation through the provision of empirically-driven feedback.

A well-implemented self-evaluation system can be a powerful tool for personal and professional development, enhancing self-awareness, accountability, and continuous improvement. By fostering a culture of reflection and empowerment, organizations and individuals can leverage self-evaluation to achieve their goals effectively and adapt to changing environments and challenges.

9.6 Glossary

Self-Evaluation: A process where individuals assess their own performance, skills, behaviors, or achievements against predefined criteria or goals.

Online Evaluation System: A digital platform or tool used to conduct assessments, tests, or evaluations electronically via the internet.

9.7 Answers to Self-Check Exercises

Exercise-1

Answer1. C) To enhance self-awareness and accountability
Answer2. B) Introspection and reflection
Answer3. B) Reflective process
Answer4. C) By encouraging goal setting and action planning
Answer5. C) Promotes personalized development

Exercise-2

Answer1. B) Automated scoring
Answer2. C) By conducting assessments remotely
Answer3. C) Automated scoring
Answer4. D) Customizable assessment content
Answer5. B) Providing insights into assessment performance

9.8 References/ Suggested Readings

- Ebel, Robert L.(1966) “Measuring Educational Achievement, Prentice Hall of India Pvt. Ltd. Pp. 481
- Gronlund, N. E. (1976), Measurement and Evaluation in Teaching. McMillan, USA.
- Hopkins, C.D. and Antes, R.L. (1990).Classroom measurement and evaluation. Itasca, Illinois: Peacock.
- Izard, J. (1991). Assessment of learning in the classroom. Geelong, Vic.: Deakin University.
- Izard, J. (1997). Content Analysis and Test Blueprints. Paris: International Institute for Educational Planning.
- Mehrens, W.A. and Lehmann, I.J. (1984).Measurement and evaluation in education and psychology.(3rd Ed.) New York: Holt, Rinehart and Winston.
- Nandra, I.D.S.(2011). Learning Resources and Assessment of Learning.Patiala, 21st Century Publications.
- Taiwo, Adediran A. (2005). Fundamentals of Classroom Testing. New Delhi: Vikas Publishing House Pvt. Ltd.
- Walter W. Cook (1958). Educational Measurement. Washington D.C.: American Council on Education.
- Withers, G. (1997).Item Writing for Tests and Examinations. Paris: International Institute for Educational Planning.

9.9 Terminal Questions

Q.1 Discuss the significance of self-evaluation in personal and professional development. How does the process of self-evaluation contribute to continuous improvement and accountability?

Q.2 Evaluate the advantages and challenges of implementing online evaluation systems in educational and organizational settings. How do these systems impact assessment practices, accessibility, and reliability?

UNIT-10 Continuous and Comprehensive Evaluation

Structure

10.1 Introduction

The purpose of assessment is necessarily to improve the teaching-learning process and material, and be able to review the objectives that have been identified for different school stages, by gauging the extent to which capabilities if the learners have been developed. Examinations play an important part in one's educational career. The present examination system in India is predominately focusing on the intellectual skills mainly and the present and the society further supporting it, the psycho motor and affective domains of holistic learning have not received their due importance. But the aim of education is developing the 'whole child'. Holistic education demands development of all aspects of individual's personality including cognitive, affective and psycho motor domains. In the present scenario it is very stressful for the parents, teachers and students only to be working on cognitive aspects without learning the processes of learning. Teachers' professional self esteem and promotions are geared to the scholastic marks attained by their learner. Focusing on excellence in academics alone undoubtedly result in lop-sided development of personality. In order to bring about the improvement in the quality of the education and the holistic development of the child who is tomorrow's global citizen, evaluation process should focus adequately on both scholastics and non-scholastic areas of development. Hence the focus needs to shift to continuous and comprehensive evaluation.

10.2 Learning Objectives

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

- Understand the Concept of Continuous and Comprehensive Evaluation.
- Explain the features of CCE.

- Explain the objectives of CCE.
- Understand the importance and Functions of CCE.

10.3 Meaning concept and features of Continuous and Comprehensive Evaluation

Meaning of Continuous and Comprehensive Evaluation:

Continuous and comprehensive Evaluation refers to a system of school based assessment that covers all aspects of student's development. It helps in improving student's performance by identifying his/her learning difficulties at regular time intervals right from the beginning of the academic session and employing suitable remedial measures for enhancing their learning performance. By facilitating all-round development of students, providing all the students the same opportunity to display their individual potential, helping the teacher to realize the effectiveness of teacher-learning process, continuous and comprehensive evaluation technique proves itself as a boost to students.

The National Policy on Education (NPE 1986), states that" Comprehensive and Continuous Evaluation should incorporate both scholastic and non-scholastic aspects of evaluation, spread over the total span of instructional time"

The NCF 2005 also recommends that a school-based Continuous and Comprehensive Evaluation system be established in order to

- Reduce stress on the children,
- Make evaluation comprehensive and regular,
- Provide space for the teacher for creative teaching,
- Provide a tool for diagnosis and for producing learners with greater skills.

Concept of Continuous and Comprehensive Evaluation:

Continuous and Comprehensive Evaluation was formulated by ministry of Human Resource Development, Kapil Sibal to decrease the accumulated stress of board exams on the students and to introduce a more uniform and comprehensive pattern in education for the children all over the nation. It helps. It helps in improving student's performance by identifying his/her learning difficulties at regular time intervals right from the beginning of the academic session and employing suitable remedial measures for enhancing their learning performance. Continuous and Comprehensive Evaluation refers to a system of school based assessment that covers all aspects of student's development. Now according to Right of Education act Continuous Comprehensive Evaluation become mandatory at primary level of education. Continuous Comprehensive Evaluation technique is useful to identify difficulties and weakness in learning of students. It is also useful for all round development of the student. But it is necessary to use variety of evaluation tools and technique. Evaluation is a process by which we can collect evidences for student progress. By analyzing collected data we can record

observations about an individual and them teacher can adopt corrective measures for better learning of student. In short, it involves systematic collection, analysis and interpretation of learner' progress both in scholastic and co-scholastic areas of learning to provide constant feedback about the effectiveness of course content, classroom processes and the growth in individual learner. Continuous Comprehensive Evaluation means a method adopted to evaluate various aspects of development of students personality from various dimensions. It is comprehensive because the evaluation is exhaustive and is done on many levels and since the evaluation is all year round, it is continuous. Continuous Comprehensive Evaluation is divided in to 3 parallel parts:

Part-1: This part deals with performance in academic subjects like Science, Math and English instead of marks, grades are given based on the performance all through the year. The academic year is divided into two sessions, in each session; there are Formative and Summative Assessments. The school has liberty in deciding the number of Formative assessments and their percentage by weight, but at the end of the evaluation, the students get a grade for each subject in stead of marks.

Part-2: This part deals with Co-scholastic areas such as life Skills, Attitude and Value. For each of the co-scholastic skills, teachers are again required to give a Grade and a descriptive indicator. Schools can expand these; for example, life skills can include Thinking skills, Social Skills and Emotional skills. Attitude can be judged towards teachers, peers and environment. This part of the CCE aims to tell the students and parents that it is not just the education that is important for an overall development of a child during the schooling years.

Part-3: This part deals again deals with Co-scholastic activities. The idea behind this part is to give simple grades based on activities performed/participated in during the year. This part is divided in two section a literary/Creative/Scientific/Aesthetic Skills, Performing Art, Clubs etc) Health and Physical Education. Teachers are expected to grades students on their involvement with these activities during the year. Students can sticks with areas like literary skills like debates and Declamations or even showcase their talents in more creative fields like art, craft and drama. This way, this part again promotes development of a child in areas other than academics. In the CCE grading system there will be dual formative assessment and single summative assessment for assessment of scholastic areas. In the formative assessment students will be given regular feedback and motivate them to actively involve themselves in self learning. This will help to increase the students' performance level and confidence level. The formative

Assessment is not constrained only to the pencil paper tests. It also has various quizzes, oral testing, projects, assignments etc. The Summative assessment in the CCE is a way of assessment of student's performance at the end of the teaching. The evaluation is of pen-paper test and is carried out

by the schools themselves. This will be held at the end of each term. There will be an evaluation of Co-Scholastic areas like students achievement, Attitudes, Creative and scientific skills, health and physical education and many more.

Features of Continuous and Comprehensive Evaluation:

The continuous aspect of CCE takes care of continual and periodicity aspect of evaluation. Continual means assessment of students in the beginning of instructions (placement evaluation) and assessment during the instructional process (formative evaluation) done informally using multiple techniques of evaluation. Periodicity means assessment of performance done frequently at the end of unit/term (summative)

The comprehensive component of CCE takes care of assessment of all round development of the child's personality. It includes assessment in Scholastic as well as Co-Scholastic aspects of the pupil's growth.

Objectives of The Scheme:

- To develop cognitive, psychomotor and affective skills.
- To lay emphasis on thought process and de-emphasis memorization.
- To make evaluation an integral part of teaching-learning process.
- To use evaluation for improvement of students' achievement and teaching learning strategies on the basis of regular diagnosis followed by remedial instruction.
- To use evaluation as a quality control devise to maintain desired standard of performance.
- To determine social utility, desirability or effectiveness of a programme and take appropriate decisions about the learner, the process of learning and the learning environment.
- To make the process of teaching and learning a learner-centered activity.

The comprehensive component of CCE takes care of assessment of all round development of the child's personality. It includes assessment in Scholastic as well as Co-Scholastic aspects of the pupil's growth. Scholastic aspects include curricular areas or subject specific areas, whereas co-scholastic aspects include life skills Co-Curricular, attitudes, and values.

Functions of Continuous and Comprehensive Evaluation:

- It helps the teacher to organize effective teaching strategies.
- Continuous and comprehensive evaluation helps in regular assessment to the extent and degree of learner's progress.
- Continuous and comprehensive evaluation serves to diagnose weaknesses and permits the teacher to ascertain in individual learner's strengths and weaknesses and their needs.
- It provides immediate feedback to the teacher, who can then decide whether a particular unit class or whether a few individuals are in need of remedial instruction.

- By continuous evaluation, children can know their strengths and weaknesses. It provides the child a realistic self assessment of how he/she studies.
- It can motivate children to develop good study habits, to correct errors, and to direct their activities towards the achievement of desired goals.
- It helps a learner to determine the areas of instruction in which more emphasis is required.
- Continuous and comprehensive evaluation identifies areas of aptitude and interest. It helps in identifying changes in attitudes, and value systems.
- It helps in making decisions for the future regarding choice of subjects, courses and careers.
- It provides information/reports on the progress of students in scholastic and co-scholastic areas and thus helps in predicting the future success of the learner.
- Continuous evaluation helps in bringing awareness of the achievement to the child, teachers and parents from time to time. They can look into the probable cause of the fall in achievement if any, and may take remedial measures of instruction in which more emphasis is required.

Self-Check Exercise-1

Q.1 What is the primary objective of Continuous and Comprehensive Evaluation (CCE) in education?

A) To conduct annual examinations

B) To replace regular assessments

C) To assess students continuously throughout the academic year

D) To eliminate grading systems

Q2. Which assessment method is integral to Continuous and Comprehensive Evaluation (CCE)?

A) Norm-referenced grading

B) Criterion-referenced testing

C) Single-point evaluations

D) Final examinations only

Q3. How does CCE contribute to holistic student development?

A) By focusing solely on academic achievements

B) By integrating scholastic and co-scholastic assessments

C) By eliminating formative assessments

D) By emphasizing annual evaluations

Q4. What aspect of CCE emphasizes regular feedback and improvement?

A) Conducting standardized tests

B) Summative assessments

C) Continuous evaluation

D) Grading on a curve

Q5. What role does CCE play in promoting personalized learning approaches?
A) Standardizing assessment practices
B) Focusing on memorization-based exams
C) Customizing learning experiences based on individual progress
D) Limiting assessments to final exams

10.4 Need and functions of Continuous and Comprehensive Evaluation

Need of Continuous and Comprehensive Evaluation:

Continuous and Comprehensive Evaluation is a boost to students. It helps in reducing streets of students by:

- Identifying learning, progress of students at regular time intervals on small portions of content.
- Employing a variety of remedial measures of teaching based on learning needs and potential of different students.
- Avoiding from using negative comments on the learner's performance.
- Encouraging learning through employment of a variety of teaching aids and techniques.
- Involving learners actively in the learning process.
- Recognizing and encouraging specific abilities of students, who do not excel in academics but perform well in other co curricular areas.

Advantages of Continuous and Comprehensive Evaluation:

- There is no pressure for students to become highly academic because they aim to encourage individuals to choose subjects based on their interests while retaining the importance of academia. They aim to make the students feel more related so they improve on their academic ability without feeling under pressure.
- The CCE system also focuses o holistic education which aims to develop various aspects of a student's personality which ultimately helps them identify what they are better at and stronger at in terms of academics.
- CCE helps in dropping stress of students in different ways like, evaluating learning advancement of students at expected time gaps on small portions of contents.
- Encourage learning through employing different teaching aids and techniques and involving captive activities in the learning process.
- The grading system will helps to give up the use of negative comments on the learner's performance.

Obstacles in Continuous and Comprehensive Evaluation:

- Lack of necessary knowledge and skill to implement continuous comprehensive evaluation.
- Inadequacy of infrastructure and time.
- Difficult to prepare and maintain records.

- Lack of provisions I curriculum for continuous comprehensive evaluation.
- Student attendance and availability of resources.
- Autonomy for colleges/institutions.
- Difficult to pay individual attention towards individual student.
- Minute observations are necessary as evaluation will be done by gradation method.
- The marking scheme will be a challenge.

Remedial Measures:

- Orientation to teachers about continuous comprehensive evaluation.
- Provisions in curriculum
- Planning of activities
- Time and work planning
- Development of Question Banks
- Development of Multiple Choice Questions
- Development of Diagnostic and criterion referenced tests.

Self- Check Exercise-2

Q.1 What is the primary need for implementing Continuous and Comprehensive Evaluation (CCE) in education?

A) To reduce teacher workload
B) To conduct fewer assessments
C) To provide holistic assessment of students
D) To eliminate grading systems

Q.2 Which function of CCE emphasizes assessing students' academic achievements and learning progress?

A) Providing feedback
B) Conducting annual examinations
C) Integrating co-scholastic activities
D) Focusing on memorization-based tests

Q3. How does CCE support personalized learning experiences for students?

A) By standardizing assessment practices
B) By conducting only summative assessments
C) By adapting teaching methods to individual needs
D) By eliminating formative assessments

Q4. What is a key function of CCE in promoting holistic development of students?

A) Conducting single-point evaluations
B) Integrating scholastic and co-scholastic assessments
C) Emphasizing competitive examinations
D) Eliminating extracurricular activities

Q5. Why does CCE emphasize continuous assessment throughout the academic year?

A) To reduce assessment workload for teachers
B) To allow for regular feedback and improvement
C) To limit student participation in assessments
D) To focus only on final examination results

10.5 Role of Teachers in Continuous and Comprehensive Evaluation

While organizing professional development of teachers like in-service training, following points need to be addressed by educators:

- Trainings are not to be organized in a 'top-down' manner, by telling teachers to implement methods or strategies suggested by them for CCE. Teachers need to be suggested how to do assessment by taking examples so that they would get opportunities to discuss, reflect and share their problems.
- Teachers must get the chance for peer discussion and sharing of school practices related to CCE. This process would facilitate mutual and participatory learning.
- Clear understanding on **purposes** of assessment and evaluation procedures is required otherwise it would damage the learning process.
- Under CCE, many states have developed various formats for recording and reporting progress of children. Clarity on different aspects of CCE is required while generating any kind of assessment data. Without such clarity, experience has shown that prescribed formats are not helping in teaching-learning process, rather wasting teaching-learning time.
- Teachers are working in varied and often difficult situations, such as large-size classrooms, in difficult-terrain schools, multi-grade classrooms etc. A uniform recording and reporting format would not serve the purpose of CCE. Prescriptive formats that do not give flexibility to the teacher go against the very spirit of CCE.

In Continuous Comprehensive Evaluation (CCE), teachers play a crucial role in various aspects of the assessment process. Here are some key roles and responsibilities of teachers in CCE:

1. **Designing Assessment Tools**: Teachers are responsible for designing a variety of assessment tools such as tests, quizzes, projects, presentations, etc., that align with the learning objectives and curriculum requirements.
2. **Administering Assessments**: Teachers conduct assessments periodically as per the CCE guidelines. They ensure that assessments are fair, transparent, and conducted in a conducive environment for learning.
3. **Observation and Documentation**: Teachers observe students' progress on a regular basis. They maintain records and document students' performance across different skills and subjects.
4. **Providing Continuous Feedback**: One of the key features of CCE is continuous feedback to students. Teachers provide constructive feedback on

students' strengths and areas needing improvement, which helps in their overall development.

5. **Identifying Learning Gaps**: Through assessments and observations, teachers identify learning gaps or areas where students may need additional support. This enables them to plan targeted interventions.

6. **Monitoring Progress**: Teachers monitor students' progress over time and track improvements or challenges they face. This helps in understanding the effectiveness of teaching strategies and adjusting them as needed.

7. **Supporting Individualized Learning**: CCE encourages personalized learning approaches. Teachers support students individually based on their strengths, weaknesses, and learning styles, fostering a more inclusive and effective learning environment.

8. **Collaboration with Parents**: Teachers communicate regularly with parents or guardians regarding students' progress, strengths, and areas needing improvement. This collaboration ensures holistic development and support for students.

9. **Professional Development**: Teachers engage in professional development activities to stay updated with CCE practices, assessment techniques, and educational trends. This helps them enhance their skills in implementing CCE effectively.

10. **Ensuring Fairness and Objectivity**: Teachers ensure that assessments are fair, objective, and aligned with the CCE principles. They avoid biases and maintain standards to uphold the credibility of assessment results.

Overall, teachers in CCE not only assess students' academic performance but also contribute significantly to their overall development by providing continuous support, feedback, and guidance throughout their learning journey.

Besides this, the following can be done:

1. Short duration trainings of 4-5 days at one time may be considered, preferably during vacations so that teachers’ and children’s learning time is not consumed for trainings.

2. Administrators need to know that they are also a part of teaching-learning process and their role is very important. Regular interactions with teachers to discuss their problems and find solutions can solve many difficulties.

3. It is necessary to give flexibility of time-table to the teacher for designing and evolving her teaching in CCE. CCE cannot work in rigidly bound time-schedules decided by people other than the teacher.

4. Teachers should be encouraged to use locally available resources and opportunities of learning outside the classrooms, which sometimes are not encouraged by administrators.

5. The essence of training programmes attended by teachers, head teachers and other educational personnel must be shared with all implementers. This process would help everyone update their knowledge and also make them understand the rationale of newly recommended changes.

6. Autonomy needs to be given to teachers to undertake the syllabus in a sequence or manner they would like to take. For example, in most of the schools teachers have to take chapters in a sequence suggested by schools. Without this flexibility, a teacher cannot implement CCE or improve his/her teaching.

Self-Check Exercise-3

Q.1 What is one of the primary responsibilities of teachers in CCE?
A) Conducting annual examinations
B) Designing assessment tools
C) Assigning final grades at the end of the year
D) Reviewing textbooks for curriculum alignment

Q.2 In CCE, teachers provide continuous feedback to students primarily to:
A) Evaluate students' performance in exams
B) Discourage students from making mistakes
C) Support students' overall development
D) Motivate students to compete with each other

Q.3 What do teachers do to identify learning gaps in students as part of CCE?
A) Conducting weekly surprise quizzes
B) Administering standardized tests every month
C) Observing students' progress and performance
D) Providing extra homework assignments

Q.4 Collaboration with parents or guardians in CCE primarily focuses on:
A) Assigning homework tasks
B) Sharing students' progress and areas needing improvement
C) Informing about school events only
D) Recommending additional tuition classes

Q.5 How do teachers contribute to personalized learning in CCE?
A) Conducting uniform assessments for all students
B) Providing individualized feedback based on students' needs
C) Setting rigid grading curves
D) Ignoring students' learning preferences

10.6 Summary

The CCE approach believes that teaching-learning is a continuous process that depends on dynamic interactions between the learner, her peers and the teacher. The teacher is the person who spends the maximum time with children in the classroom. Therefore the teacher is the best person to judge children’s learning needs, levels and progress. If any record is to be maintained in assessment, it should be mainly to inform the teacher and the choice as to what records he/she wants to keep, must be with him/her.

Recording of each and every classroom activity is burdensome, impractical and does not help teaching learning. **The teacher should not be forced to record and report continuously, for all her classes or activities.**
This would require that education officials, superiors and inspectors respect the teacher's autonomy, making her feel responsible and worthy of taking charge of children's learning. CCE can only work in non-threatening situations, for both the teacher and the learners, where the charge of teaching-learning is given to them. Here administrators can encourage teachers to concentrate more on assessing the process and interaction in their classroom, rather than the product.

Teachers are central to the success of Continuous and Comprehensive Evaluation (CCE) by designing, implementing, and refining assessment practices that promote holistic student development, personalized learning experiences, and continuous improvement. Their roles encompass not only academic assessment but also nurturing the overall growth and well-being of students in educational settings.

10.7 Glossary

- **Continuous and Comprehensive Evaluation (CCE):** An evaluation system in education that assesses students continuously throughout the year across scholastic and co-scholastic areas.
- **Assessment:** The process of gathering and interpreting evidence of students' learning or performance.
- **Professional Development:** Activities and programs designed to improve educators' knowledge, skills, and effectiveness in teaching and assessment practices.

10.8 Answers to Self-Check Exercises

Exercise-1

Answer1. rC) To assess students continuously throughout the academic year
Answer2. B) Criterion-referenced testing
Answer3. B) By integrating scholastic and co-scholastic assessments
Answer4. C) Continuous evaluation
Answer5. C) Customizing learning experiences based on individual progress

Exercise-2

Answer1. C) To provide holistic assessment of students
Answer2. A) Providing feedback
Answer3. C) By adapting teaching methods to individual needs
Answer4. B) Integrating scholastic and co-scholastic assessments
Answer5. B) To allow for regular feedback and improvement

Exercise-3

Answer1: B) Designing assessment tools
Answer2: C) Support students' overall development
Answer3: C) Observing students' progress and performance
Answer4: B) Sharing students' progress and areas needing improvement
Answer5: B) Providing individualized feedback based on students' needs

10.9 References/ Suggested Readings

- Ebel, Robert L.(1966) “Measuring Educational Achievement, Prentice Hall of India Pvt. Ltd. Pp. 481
- Gronlund, N. E. (1976), Measurement and Evaluation in Teaching. McMillan, USA.
- Hopkins, C.D. and Antes, R.L. (1990).Classroom measurement and evaluation. Itasca, Illinois: Peacock.
- Izard, J. (1991). Assessment of learning in the classroom. Geelong, Vic.: Deakin University.
- Izard, J. (1997). Content Analysis and Test Blueprints. Paris: International Institute for Educational Planning.
- Mehrens, W.A. and Lehmann, I.J. (1984).Measurement and evaluation in education and psychology.(3rd Ed.) New York: Holt, Rinehart and Winston.
- Nandra, I.D.S.(2011). Learning Resources and Assessment of Learning.Patiala, 21st Century Publications.
- Taiwo, Adediran A. (2005). Fundamentals of Classroom Testing. New Delhi: Vikas Publishing House Pvt. Ltd.
- Walter W. Cook (1958). Educational Measurement. Washington D.C.: American Council on Education.
- Withers, G. (1997).Item Writing for Tests and Examinations. Paris: International Institute for Educational Planning.

10.10 Terminal Questions

Q.1 What do you understand by the term CCE?
Q.2 Do you believe that activity based learning can help Indian student
Q.3 Who developed the CCE pattern? What are its main objectives?

UNIT - 11

CHARACTERISTICS OF MEASUREMENT AND EVALUATION TOOLS-I

Structure

11.1 INTRODUCTION

Dear Learner,

Evaluation is an important aspect of educational process. It is very important to evaluate achievement of the students. In the present lesson, we will learn about the basic characteristics of evaluation tools which includes; reliability, validity, objectivity, usability and practicability.' The lesson will also throw light on norms for interpreting test scores.

11.2 LEARNING OBJECTIVES

After studying this unit, you will be able to:

- List down different characteristics of measurement and evaluation tools.
- Explain reliability and ways of its estimation.
- Discuss different types of validity.
- Define objectivity.
- Explain various procedures of developing and using test score norms.

11.3 BASIC CHARACTERISTICS OF GOOD MEASURING TOOLS

Whenever a test or other measuring device is used as part of the data collection process, the validity and reliability of that test is important. Just as we would not use a math test to assess verbal skills, we would not want to use a measuring device for research that was not truly measuring what we purport it to measure. After all, we are relying on the results to show support or a lack of support for our theory and if the data collection methods are erroneous, the data we analyze will also be erroneous. Here we mention some qualities of a test and the techniques of its estimation.

11.3.1 Reliability

Reliability is the consistency of your measurement, or the degree to which an instrument measures the same way each time it is used under the same condition with the same subjects. In short, it is the repeatability of your measurement. A measure is considered reliable if a person's score on the same test given twice is similar. It is important to remember that reliability is not measured, it is estimated. A good instrument will produce consistent scores. An instrument's reliability is estimated using a correlation coefficient of one type or another. Reliability is synonymous with the consistency of a test, survey, observation, or other measuring device. Imagine stepping on your bathroom scale and weighing 140 pounds only to find that your weight on the same scale changes to 180 pounds an hour later and 100 pounds an hour after that. Based on the inconsistency of this scale, any research relying on it would certainly be unreliable. Consider an important study on a new diet program that relies on your inconsistent or unreliable bathroom scale as the main way to collect information regarding weight change. Would you consider their results accurate? A reliability coefficient is often the Statistic of choice in determining the ' reliability of a test. This coefficient merely represents a correlation, which measures the intensity and direction of a relationship between two or more variables.

Test-Retest Reliability: Test-Retest reliability refers to the test's consistency among different administrations. To determine the coefficient for this type of reliability, the same test is given to a group of subjects on at least two separate occasions. If the test is reliable, the scores that each student receives on the first administration should be similar to the scores on the second. We would expect the relationship between he first and Second administration to be a high positive correlation. One major concern with test-retest reliability is what has been terrified the memory effect. This is especially true when the two administrations are dose together in time. For example, imagine taking a short 10-question test on vocabulary and then terv minutes later being asked to complete the same test. Most of us will remember our responses and when we begin to answer again, we may just answer the way we did on the first test rather than reading through the questions carefully. This can create an artificially high reliability coefficient as subjects respond

from their memory rather than the test itself. When a pre-test and post-test for an experiment is the same, the memory effect can play a role in the results.

Parallel Forms Reliability: One way to assure that memory effects do not occur is to use a different pre- and post-test. In order for these two tests to be used in this manner, however, they must be parallel or equal in what they measure. To determine parallel forms reliability, a reliability coefficient is calculated on the scores of the two measures taken by the same group of subjects. Once again, we would expect a high and positive correlation is we are to say the two forms are parallel.

Inter-Rater Reliability: Whenever observations of behavior are used as data in research, we want to assure that these observations are reliable. One way to determine this is to have two or more observers rate the same subjects and then correlate their observations. If, for example, rater A observed a child act out aggressively eight times, we would want rater B to observe the same amount of aggressive acts. If rater B witnessed 16 aggressive acts, then we know at least one of these two raters is incorrect. If there ratings are positively correlated, however, we can be reasonably sure that they are measuring the same construct of aggression. It does not, however, assure that they are measuring it correctly, only that they are both measuring it the same.

Self-Check Exercise-1

1. What is reliability in the context of measurement?
   a) The accuracy of a measurement.
   b) The consistency of a measurement.
   c) The ability to measure different things.
   d) The validity of a measurement.
2. What does a high positive correlation in test-retest reliability indicate?
   a) The test measures different constructs each time.
   b) The test is not reliable.
   c) The test produces similar scores on different occasions.
   d) The test has low reliability.

11.3.2 Validity

Validity is the extent to which a test measures what it claims to measure. It is vital for a test to be valid in order for the results to be accurately applied and interpreted. Validity isn't determined by a single statistic; but by a body of research that demonstrates the relationship between the test and the behavior it is intended to measure. There are three types of validity: It is the strength of our conclusions, inferences or propositions. Validity refers to the degree in which our test or other measuring device is truly measuring what we intended it to measure. The test question "1 + 1 = _____" is certainly a valid basic addition question because it is truly measuring a student's ability to perform basic addition. It becomes less valid as a measurement of advanced addition because as it addresses some required knowledge for addition, it does not represent all of knowledge required for an advanced understanding of addition. On a test designed to measure knowledge of American History, this question becomes completely invalid. The ability to add two single digits has nothing do with history.

For many constructs, or variables that are artificial or difficult to measure, the concept of validity becomes more Complex. Most of us agree that “1 + 1 = _____" would represent basic addition, but does this question also represent the construct of intelligence? Other constructs include motivation, depression, anger, and practically any human emotion or trait. If we have a difficult time defining the construct, we are going to have an even more difficult time measuring it. Construct validity is the term given to a test that measures a construct accurately and there are different types of construct validity that we should be concerned with. Three of these, concurrent validity, content validity, and predictive validity are discussed below.

Concurrent Validity: Concurrent Validity refers to a measurement device’s ability to vary directly with a measure of the same construct or indirectly with a measure of an opposite construct. It allows you to show that your test is valid by comparing it with an already valid test. A new test of adult intelligence, for example, would have concurrent validity if it had a high positive correlation with the Wechsler Adult Intelligence Scale since the Wechsler is an accepted measure of the construct we call intelligence. An obvious concern relates to the validity of the test against which you are comparing your test. Some assumptions must be made because there are many who argue the Wechsler scales, for example, are not good measures of intelligence.

Content Validity: Content validity is concerned with a test’s ability to include or represent all of the content of a particular construct. The question “1 + 1 =_____” may be a valid basic addition question. Would it represent all of the content that makes up the study of mathematics? It may be included on a scale of intelligence, but does it represent all of intelligence? The answer to these questions is obviously no. To develop a valid test of intelligence, not only must there be questions on math, but also questions on verbal reasoning, analytical ability, and every other aspect of the construct we call intelligence. There is no easy way to determine content validity aside from expert opinion.

Predictive Validity: In order for a test to be a valid screening device for some future behaviour, it must have predictive validity. The entrance tests must possess predictive validity. The main concern with these, and many other predictive measures is predictive validity because without it, they would be worthless.

Self-Check Exercise-2

1. Which type of validity refers to a test’s ability to measure all aspects of a particular construct?
   a) Concurrent Validity
   b) Predictive Validity
   c) Content Validity
   d) Construct Validity

2. Which type of validity is important for tests used as screening devices for future behavior?
   a) Concurrent Validity
   b) Predictive Validity

c) Content Validity
d) Inter-Rater Reliability

11.3.3 Objectivity, Usability and Practicability

It refers to fairness and uniformity in the test scoring procedure. Examiner / rater bias is therefore non-existent in MCQ based tests. That is why they are called objective tests. Further, the analysis, of the test data is undertaken statistically which further assesses various dimensions of the tests to make them more precise and accurate instruments.

MCQs are practically useful and efficient especially in a large scale testing situation unlike descriptive tests which are more resource intensive and demand time and money. The tests should be easy in administration and scoring and interpretation. These qualities fall under practicality. The tests should be feasible & usable. Quality of being usable means the test can be used in context to the objective to be achieved. Usability (practicality) includes ease in administration, scoring, interpretation and application, low cost, proper mechanical make - up. It should measure the objective to be achieved.

Self-Check Exercise-3

1. Why MCQ-based tests are considered objective?
    a) They are easy to administer.
    b) They are less time-consuming.
    c) They eliminate examiner/rater bias.
    d) They require less statistical analysis.

2. Which quality is NOT mentioned as part of the practicality of a test?
    a) Ease in administration
    b) High cost
    c) Scoring and interpretation
    d) Proper mechanical make-up

11.4 NORMS FOR INTERPRETATION OF TEST SCORES

Norms means distribution of scores on a particular test by a well-established group of people. There are different types of norm samples:

National norms: based on particular population.

Local norms: based on a region within a country

Within group norms: usually based on national norms for some well identified sub-group within the population.

It is important to know about the sample used to norm a test because an individual's score on a test has meaning only in the context of the standardization sample. In psychological testing, scores on tests seldom have any concrete meaning, so that interpretation of a score on a test is always relative to how others scored on that test.

11.4.1 Developing & Using Test Norms to Compare Performance

A test norm is a set of scalar data describing the performance of a large number of people on that test. Test norms can be represented by two important statistics: Means and Standard Deviations. The Standard Deviation (SD) tells us how spread out the distribution is around a central point. The standard deviation is, in effect, an average of the departure of people's scores from the group mean. This is a measure of how spread out the group's scores typically are. The standard deviation gives an indication of the degree of dispersion of the scores from the mean, and an estimate of the variability of results in the total sample. It is usefully employed in comparing the variability of different groups. It is used to develop Test Norms because, unlike percentiles, measures based on how much an individual deviates from the mean can be mathematically manipulated and compared. Let us suppose that we have two distributions, Group A and Group B, with the same mean but the performance of people Group A is more spread out. Group A will have a larger standard deviation than Group B, which has smaller individual differences. In Group B, each score is closer to the mean and this means there is a smaller standard deviation.

A. Describing Performance using Z-Scores

The Z score, based on the mean and standard deviation, indicates how many standard deviations above or below the mean a score is. A Z score is merely a raw score, which has been changed to standard deviation units.

The standard Z score is calculated by the formula:

$$Z = \frac{\left(\mathrm{X}\text{-}\overline{\mathrm{X}}\right)}{SD}$$

WhereZ - standard Z score

X = each individual raw score

X (bar) = mean score

SD= standard deviation

Usually when standard scores are used they are interpreted in relation to the normal distribution curve. Standard Z scores in standard deviation units are marked out either side of the mean. Those above the mean are positive, and those below the mean negative in sign. Therefore, by the calculation of the standard Z score an individual's score can be viewed in relation to the rest of the distribution. The standard score is very important when comparing scores from different tests. Before these scores can be properly compared they must be converted to a common scale such as the standard Z score. The Z-score can then be used to express an individual's score on different tests in terms of norms. One important advantage in using the normal distribution curve as a basis for test norms is that the standard deviation has a precise relationship with the area described by the curve. One standard deviation

above and below the mean will include approximately 68% of the sample, as illustrated here. Z scores have a mean of 0 and a standard deviation of 1. Because of this, Z scores can be rather difficult and cumbersome to handle because most of them are decimals and approximately half of them can be expected to be negative. To remedy these drawbacks various transformed standard score systems have been derived. These simply entail multiplying the obtained Z score by a new standard deviation and adding it to a new mean. Both of these steps are devised to eradicate decimals and negative numbers.

B. Describing Performance Using Z-Scores

The t in t score stands for transformed. A T score is a transformed z score. T scores and stens (see below) take the Z score and transform it into a more effective and user- friendly format. However critically, both still describe the level of deviation a particular score has from the mean. The T score (transformed score) is the most common standardized norm system for ability tests. The T score is derived from the Z score, transformed to a new scale, so that the: T score mean is set at 50 and the standard deviation at 10.

Calculate the T score using the following formula:

Z score X10+50

If Z score = 2

T score = 2 x 10 + 50

= 20 + 50-70

Although the T score can easily be calculated, it is most often used by reference to a norm table. In order to identify the relevant T score, you find the raw score in the body of the table and read across to the right hand side. The advantage of T scores over percentiles is that they are a linear scale with equal units of measurement. Thus they can be mathematically manipulated and give results which can be directly compared both within and between individuals. Their disadvantage is that they are not so easily explained or meaningful to those not knowledgeable about testing.

C. Describing Performance Using Stens

Stens or Standard Tenths are the most commonly used scales for comparing individual differences. The normal distribution is divided into ten stens. Unlike percentiles, stens are equal units of measurement, and unlike percentiles they are not influenced by clustering around the midpoint.

A Sten is based on a mean of 5.5 and a standard deviation of 2.

Calculate the sten using the following formula:

Z score x 2 + 5.5

If Z score =2

Sten =2x2+5.5 = 9.5

Normally, the sten is taken to the nearest whole number, with a minimum value of 1 and. a maximum of 10. Stens are commonly used as the norm system for personality questionnaires. However, they are also a useful system for ability tests as they encourage us to think in bands of scores but still provide sufficient discrimination between applicants.

Self-Check Exercise-4

1. What does a standard deviation (SD) indicate in a set of test norms?

a) The average score of the group.
b) The degree of dispersion of scores from the mean.
c) The highest score in the group.
d) The median score of the group.

2. How is a Z score interpreted in relation to the normal distribution curve?

a) It indicates the raw score of an individual.
b) It indicates how many standard deviations a score is from the mean.
c) It shows the percentage of people scoring above the mean.
d) It measures the average deviation of scores from the mean.

11.5 SUMMARY

Evaluation in education is critical for assessing student achievement, relying on robust measurement tools characterized by reliability, validity, objectivity, usability, and practicability. Reliability, reflecting consistency in test results, can be estimated through methods like test-retest, parallel forms, internal consistency, and interrater reliability. Validity ensures the test measures what it purports to, while objectivity minimizes scorer bias. Various procedures develop test score norms, enhancing interpretative accuracy. Methods to ascertain reliability include the parallel form method, Kuder-Richardson formula, test-retest method, and split-half method, each with distinct advantages and limitations. These principles ensure that educational assessments are both accurate and fair.

11.6 GLOSSARY

Reliability: it is the degree of consistency of a measure.
Validity: Validity refers to how accurately a method measures what it is intended to measure.

11.7 ANSWERS TO SELF-CHECK EXERCISE 1, 2, 3 & 4.

Self-Check Exercise-1

1. b) The consistency of a measurement
2. c) The test produces similar scores on different occasions.

Self-Check Exercise-2

1. c) Content Validity
2. b) Predictive Validity

Self-Check Exercise-3

1. c) They eliminate examiner/rater bias.
2. b) High cost

Self-Check Exercise-4

1. b) The degree of dispersion of scores from the mean
2. b) It indicates how many standard deviations a score is from the mean.

11.8 REFERENCES/ SUGGESTED READINGS

- Education Measurement and Evaluation: J. Swarupa Rani, Discovery PublishingHouse.
- Measurement and Evaluation in Teaching: Norman Edward Gronlund Macmillian
- Measurement and Assessment in Teaching: Robert L. Linn Pearson Education India.
- Program Evaluation and performance measurement: James C. Me. David, Laura R.L. Hawthorn, Sage Publication

11.9 TERMINAL QUESTIONS

Dear learners, please check you progress by attempting the following questions:

1. Explain the concept of reliability in the context of educational evaluation tools. Provide an example to illustrate your explanation.

2. Discuss the different types of reliability mentioned in the text. Why is interrater reliability particularly important for performance-based assessments?

3. What are the methods for estimating the reliability of a test? Provide a brief description of one method.

4. Describe the 'Split-Half Method' for calculating the reliability of a test and mention one of its limitations**.**

UNIT-12

TYPES OF TESTS

Structure

12.1 INTRODUCTION

Dear Learner,

For measuring various characteristics of the students, we employ different types of instruments. Tests are the devices used to measure abilities, achievement or skills of the individuals. In various fields like education, psychology, healthcare, and employment, tests are used to measure knowledge, skills, attitudes, and behaviours. Different types of tests serve distinct purposes, and understanding these differences is crucial for appropriate test selection and interpretation.

12.2 LEARNING OBJECTIVES

After studying this unit, you will be able to:

- List down different types of tests.
- Differentiate between speed and power tests.
- Explain ability and achievement tests.
- Compare objective and subjective type of tests.
- Define mastery tests.
- Describe standardized test.

12.3 TYPES OF TESTS

On the basis of various dimensions, tests can be divided into different types. Different types of tests are explained below:

Speed Test and Power Test:

Some tests have very easy items but there is a limited amount of time to answer them. Such speed tests are used to see how quickly students can work on skills they have already mastered. One example is a test of keyboard skills. The teacher may wish to find out how fast students can maintain accurate work when typing data on a typewriter or computer keyboard.

In contrast, power tests are concerned with identifying skills which have been mastered. Power tests require adequate samples of student behaviour - so having sufficient time to attempt most of the items is an essential pre-requisite for such tests.

Aptitude or Ability Test and Achievement Test:

Achievement tests may be used to assess the extent to which curriculum objectives have been met in an educational programme. Such tests should have tasks which relate to the learning that students have to demonstrate. Since future learning depends to some extent on past learning, success on such achievement tests may provide evidence of future success (provided that other conditions such as good teaching, adequate health care, and stable family circumstances are maintained).

Tests which are constructed specifically to gather evidence about ability to learn are referred to as aptitude or ability tests. Results on such tests are used to predict future success on the basis of success on the specially selected tasks in the aptitude test. Often these tasks differ from the usual school learning requirements and depend to some extent on learning beyond the school curriculum. Of course, teaching students the test items and the corresponding answers may result in an increase in score without actually changing a student's (real) aptitude.

Objective and Subjective Test:

The term 'objective' can have several meanings when describing a test. It can mean that the score key for the test needs a minimum of interpretation in order to score an item correct or incorrect. In this sense, an objective test is one which requires task responses which can be scored accurately and fairly from the score key without having knowledge of the content of the test. For example, a multiple-choice test can be scored by a machine or by a clerical worker without either the machine or the clerical worker having had to reach a high level of expertise on the material being tested. A less common usage relates to the extent of agreement between experts about the correct answer. If there is less argument about the correct answer the item is regarded as more objective. However the choice of which items (whether objective in their answer format or not) are to appear on a test is subjective, in that it depends on the personal preferences and experiences of those constructing the test.

Objective tests are especially well suited to certain types of tasks. Because questions can be designed to be answered quickly, they allow lecturers to test students on a wide range of material. ... Additionally, statistical analysis on the performance of individual students, cohorts and questions is possible. The capacity of objective tests to assess a wide range of learning is often underestimated. Objective tests are very good at examining recall of facts, knowledge and application of terms, and questions that require short text or numerical responses. But a common worry is that objective tests cannot assess learning beyond basic comprehension. There are, however, limits to what objective tests can assess. They cannot, for example, test the competence to communicate, the skill pf constructing arguments or the ability to offer original responses. Tests must be carefully constructed in order to avoid the de contextualisation of knowledge and it is wise to use objective testing as only one of a variety of assessment methods within a module. However, in times of growing student numbers and decreasing resources, objective testing can offer a viable addition to the range of assessment types available to a teacher or lecturer.

Objective tests are appropriate when:

- The group to be tested is large and the test may be reused.
- Highly reliable scores must be obtained as efficiently as possible.
- Impartiality of evaluation, fairness, and free from possible test scoring influences are essential.

Essay tests are appropriate when:

- The group to be tested is small and the test is not to be reused
- You wish to encourage and reward the development of student skill in writing -
- You are more interested in exploring student attitudes than in measuring his/her achievement

Either essay or objective tests can be used to:

- Measure almost any important educational achievement a written test can measure
- Test understanding and ability to apply principles.
- Test ability to think critically.
- Test ability to solve problems.
- In general, question types fall into two categories:

1. Objective, which require students to select the correct response from several alternatives or to supply a word or short phrase to answer a question or complete a statement.

2. Subjective or essay, which permit the student, to organize and present an original answer. Examples: short-answer essay, extended-response essay, problem solving, performance test items.

Subjective items require students to write and present an original answer. It includes short - answer essay, extended - response essay, problem solving, and performance tasks. Advantages of subjective teste are higher learning skills are utilized by learners, for example synthesis, analysis and evaluation. Brevity and consciousness, precious of expression is developed among learners. It can quickly and easily constructed and eliminates guessing. Disadvantages of subjective tests are subjectivity the same piece of work can get different marks. Students with poor language prowess tend to fail. Time is consumed when answering these questions, usually it is limited in scope, and thus it doesn't cover much content. The examiner's judgment determines the final grade. The research recommends taking both subjective and objective teste for a classroom test. The universal adoption of a combination objective / subjective testing format would tend to sharpen writing and organizational skills.

Comparison between Objective and Subjective Tests:

Objective tests are so called according to their scoring which depends on personal judgments or opinions. Subjective tests are so called because their scoring depends on personal judgments or opinions, the techniques used in objective tests multiple - choice items (MCI), True / false items, matching items, transformation sentences, re-arrangement items and fill the blanks or gap filling. On the other hand, the techniques used in subjective tests include: essay writing, composition writing, letter writing, reading aloud, completion type and answer these - questions. To answer an objective test, the testee has to select his answers from two, three, four or even more alternatives while objective tests which [has] only one correct answer. Besides, to answer an objective test, the testee has to plan and write his own answer by using his own words and expressions. Furthermore, objective tests need much time and effort to write the questions because the examiner has to provide the answers as well as the question so that objective test requires more careful preparations than 46 other types of test. But in subjective test the examiner needs to write few questions without answers. In objective tests, it seems that kind is more, reliable because it gives a stable scoring. But in subjective test, it seems, it is not reliable because it doesn't give a stable scoring. Objective tests are used to test structures, vocabulary, comprehension, and sound discrimination. On the other hand, subjective tests are used to test ideas, culture, coherence and creativity. Objective tests encourage guessing and it is difficult to write simple to answer, easy to score, suit for a large number of testees, and this type of test can be scored by a machine. Besides, subjective test doesn't encourage guessing easy to write, difficult to score and suit for a small number of testee. This type of test can't be scored by a machine. Objective test can be used to test specific area of

language, while, subjective test can be used to evaluate overall achievements. Furthermore, objective tests require recognition more than production but subjective tests require production as well as recognition. An objective test is a type of discrete point test, but subjective test is a type of an integrative point test. Objective test is a type of close - ended atomistic, a system - referenced and it applicative test and replicative test. On the other hand, subjective test is a type open - ended test, a holistic and replicative and it isn't applicative. Objective test depends on students' knowledge. It is a valid test which student's need a short time to answer than subjective test. But subjective test depends on student's experience. It is invalid test and students need a long time to answer than objective test. Finally, a good classroom test should be contained both subjective and objective.

Self-check Exercise-1

1. What is the primary purpose of an aptitude or ability test?
    a) To assess curriculum objectives met in an educational program.
    b) To predict future success based on specific tasks.
    c) To evaluate the quality of teaching methods.
    d) To measure students' past achievements.

2. Which of the following is an advantage of objective tests?
    a) They can assess the skill of constructing arguments.
    b) They require extensive interpretation to score.
    c) They allow for testing a wide range of material efficiently.
    d) They are subjective and rely on the examiner's judgment.

12.4 STANDARDIZED TEST AND DIAGNOSTIC TEST

Standardized Test: The term 'standardized' also has a number of meanings with respect to testing. It can mean that the test has an agreed format for administration and scoring so that the task is as identical as possible for all candidates and there is little room for deviation in the scoring of candidate responses to the tasks. Another meaning refers to the way in which the scores on a tests are presented. For example, if scores are given as a raw score divided by some measure of dispersion like the standard deviation, the resulting score scale is said to be in terms of standardized scores (sometimes called standard scores). Finally, the term can refer (loosely) to a published test which was prepared by standard (or conventional) procedures. The usage of 'standardized' has become somewhat confused because published tests often present scores interpreted in terms of deviation from the mean (or average) and have a standard procedure for administering tests and interpreting results.

Diagnostic Test: This term refers to the use made of the information gained from administration of the test. The implication is that the test results will assist in identifying both the topics which are not known and in providing information on potential sources of the student's difficulty. Teachers may be expected to provide appropriate teaching for each difficulty exposed by the use of a diagnostic test. For example, a simple open-ended mathematics

question about area, given to junior secondary level classes provided a range of correct and incorrect answers.

Self-check Exercise-2

1. What is one meaning of a 'standardized' test?
    a) A test that has a flexible format for administration.
    b) A test that presents scores in terms of deviation from the mean.
    c) A test that has no specific scoring method.
    d) A test that is personalized for each candidate.
    e)
2. Which of the following best describes the purpose of a diagnostic test?
    a) To compare scores among a large group of students.
    b) To identify topics that are not known and sources of difficulty for students.
    c) To measure the overall achievement of students in a standardized manner.
    d) To provide a raw score without interpretation.

12.5 PRACTICAL TEST AND MASTERY TESTS

Practical Test: In some senses, an essay test is a practical task. The essay item requires a candidate to perform. This performance is intended to convey meaning in a practical sense by writing prose to an agreed format. However, the term 'practical test' goes beyond performance and other tasks used in traditional pencil-and-paper examinations. The term may refer to practical tasks in trade subjects (such as woodwork, metalwork, shipbuilding, and leather craft), in musical and dramatic performance, in skills such as swimming or gymnastics, or may refer to the skills required to carry out laboratory or field tasks in science, agriculture, geography, environmental health or physical education.

Mastery Tests: These tests are generally criterion-referenced tests with a relatively high score requirement. Students who meet this high score are said to have mastered the topic. It is assumed that the mastery test has sufficient items of high quality to ensure that the score decision is well founded with respect to the domain of interest. For example, in Mathematics the domain might be 'all additions of pairs of one-digit numbers where the total does not exceed 9'. A mastery test of this domain should have a reasonable sample of all possible combinations of one-digit numbers because the mastery decision implies that all can be added successfully even though all are not tested. This simple example should not be taken to imply that mastery testing is limited to relatively trivial skills. A more complex example is the regular testing of airline pilots. Safety requirements result in high standards being set for mastery in many areas. Failure to reach mastery will result either in further tuition under the guidance of an experienced tutor or in withdrawal of the permission to fly.

Self-check Exercis-3:

1. Which of the following is an example of a practical test?
    a) A multiple-choice test in mathematics.

b) An essay test in English literature.
c) A swimming skills assessment.
d) A standardized test in history.

2. What characterizes a mastery test?
   a) It is usually a norm-referenced test with moderate scoring requirements.
   b) It is a criterion-referenced test with a relatively high score requirement.
   c) It involves subjective evaluation of performance.
   d) It only tests trivial skills.

12.6 SUMMARY

In this lesson, we studied about various types of test. After going through this lesson, you would have understood the differences between various types of tests. You have also become clear about the concept of norm-referenced and criterion-referenced types of teste and tee differences between these two types.

12.7 GLOSSARY

Diagnostic Test: It is a form of pre-assessment or a pre-test where teachers can evaluate students' strengths, weaknesses, knowledge and skills before their instruction.

Practical Test: An assessment that evaluates an individual's ability to apply their knowledge and skills in a real-world setting.

12.8 ANSWER TO SELF-CHECK EXERCISE 1, 2 & 3.

Self-check Exercis-1

1. b) To predict future success based on specific tasks.

2. c) They allow for testing a wide range of material efficiently.

Self-check Exercis-2

1. b) A test that presents scores in terms of deviation from the mean.

2. b) To identify topics that are not known and sources of difficulty for students.

Self-check Exercis-3

1. c) A swimming skills assessment.

2. b) It is a criterion-referenced test with a relatively high score requirement

12.9 REFERENCES /SUGGESTED READING

- Ebel, Robert L.(1966) “Measuring Educational Achievement, Prentice Hall of India Pvt. Ltd. Pp.481
- Gronlund, N. E. (1976). Measurement and'Evaluation in Teaching. McMillan, USA.
- Hopkins, C.D. and Antes, R.L. (1990).Classroom measurement and evaluation. Itasca, Illinois: Peacock.
- Izard, J. (1991). Assessment of learning in the classroom. Geelong, Vic.: Deakin University.

12.10 TERMINAL QUESTIONS

Dear learners, please check you progress by attempting the following questions:

1. Name various types of tests.

2. What is the difference between speed and power tests?

3. Write down the advantages of objective type tests over subjective type tests.

4. What do you mean by standardized test?

5. Explain mastery tests.

****

UNIT-13

CRITERION-REFERENCED TESTS

AND NORM-REFERENCED TESTS

Structure

13.1 INTRODUCTION

Dear Learner,

Assessments and tests are essential tools in various fields, including education, employment, and certification. There are two primary categories of tests: Norm-Referenced Tests (NRTs) and Criteria-Referenced Tests (CRTs). Understanding the differences between these two types of tests is crucial for appropriate test selection, interpretation, and use. Norm-Referenced Tests (NRTs) compare individual performance to that of a larger group, known as the norm group. These tests aim to rank individuals relative to others, providing information on their relative standing within the group. NRTs are often used for: Selection and admission processes, ranking and comparison purposes, identifying strengths and weaknesses relative to others.

Criteria-Referenced Tests on the other hand, evaluate individual performance against specific criteria, standards, or learning objectives. These tests aim to determine whether individuals have met predetermined criteria, regardless of how others perform. CRTs are often used for: Evaluating mastery or proficiency, diagnosing strengths and weaknesses, Certifying competence or licensure

13.2 LEARNING OBJECTIVES

After studying this unit, you will be able to:

- Explain criterion-referenced tests.
- Explain norm-referenced test.
- Describe the features of CRT and NRT.
- Enlist the advantages and disadvantages of CRT and NRT.

13.3 MEANING OF CRITERION REFERENCED TEST (CRT)

Definition: A test in which questions are written according to specific predetermined criteria. A student knows what your standards are for passing and only competes against him or herself while completing.

A criterion-referenced test is designed to measure how well test takers have mastered a particular body of knowledge.

These tests generally have an established “passing” score. Students know what the passing score is and an individual’s test score is determined by knowledge of the course material.

He runs a second try out, having established that 10 seconds in the 100 yard dash is competitive in the event. He now picks those who run the dash in 10 seconds or less. This is criterion-referenced testing. He knows the runners he selected can compete. He gets the funds.

13.3.1 Features of Criterion Referenced Test (CRT)

Features of Criterion reference tests are as follows:-

(i) Criterion-referenced test place a primary focus on the content and what is being measured. Norm-referenced tests are also concerned about what is being measured but the degree of concern is less since the domain of content is not the primary focus for score interpretation. In norm-referenced test development, item selection, beyond the requirement that items meet the content specifications, is driven by item statistics. Items are needed that are not too difficult or too easy, and that are highly discriminating. These are the types of items that contribute most to score spread, and enhance test score reliability and validity.

(ii) With criterion-referenced test development, extensive efforts go into insuring content validity. Item statistics play less a role in item selection though highly discriminating items are still greatly valued, and sometimes item statistics are used to select items that maximize the discriminating power of a test at the performance standards of interest on the test score scale.

A good norm-referenced test is one that will result in a wide distribution of scores on the construct being measured by the test. Without score variability, reliable and valid comparisons of candidates cannot be made. A good criterion-referenced test will permit content-referenced interpretations and this means that the content domains to which scores are referenced must be very clearly defined. Each type of test can serve the other main purpose (norm-

referenced versus criterion-referenced interpretations), but this secondary use will never be optimal. For example, since criterion-referenced tests are not constructed to maximize score variability, their use in comparing candidates may be far from optimal if the test scores that are produced from the test administration are relatively similar.

Because the purpose of a criterion-referenced test is quite different from that of a norm referenced test, it should not be surprising to find that the approaches used for reliability and validity assessment are different too.

(iii) With criterion-referenced tests, scores are often used to sort candidates into performance categories. Consistency of scores over parallel administrations becomes less central than consistency of classifications of candidates to performance categories over parallel administrations. Variation in candidate scores is not so important if candidates are still assigned to the same performance category.

Therefore, it has been common to define reliability for a criterion-referenced test as the extent to which performance classifications are consistent over parallel-form administrations.

For example, it might be determined that 80% of the candidates are classified in the same Notes way by parallel forms of a criterion-referenced test administered with little or no instruction in between test administrations. This is similar to parallel form reliability for a norm referenced test except the focus with criterion-referenced tests is on the decisions rather than the scores. Because parallel form administrations of criterion-referenced tests are rarely practical, over the years methods have been developed to obtain single administration estimates of decision consistency that are analogous to the use of the corrected split-half reliability estimates with norm-referenced tests.

(iv) With criterion-referenced tests, the focus of validity investigations is on (1) the match between the content of the test items and the knowledge or skills that they are intended to measure, and (2) the match between the collection of test items and what they measure and the domain of content that the tests are expected to measure. The “alignment” of the content of the test to the domain of content that is to be assessed is called content validity evidence. This term is well known in testing practices. Many criterion-referenced tests are constructed to assess higher-level thinking and writing skills, such as problem solving and critical reasoning. Demonstrating that the tasks in a test are actually assessing the intended higher-level skills is important, and this involves judgments and the collection of empirical evidence. So, construct validity evidence too becomes crucial in the process of evaluating a criterion referenced test.

(v) Probably the most difficult and controversial part of criterion-referenced testing is setting the performance standards, i.e., determining the points on the score scale for separating candidates into performance categories such as “passers” and “failers.” The challenges are great because with criterion-referenced tests in education, it is common on state and national assessments to separate candidates into not just two performance categories,

but more commonly, three, four, or even five performance categories. With four performance categories, these categories are often called failing, basic, proficient, and advanced.

13.3.2 Advantages and Disadvantages of Criterion Referenced Test (CRT)

Advantages of CRT

1. Mastery of Subject Matter.

• Criterion-referenced tests are more suitable than norm-referenced tests for tracking the progress of students within a curriculum. Test items can be designed to match specific program objectives. The scores on a criterion referenced tests indicate how well the individual can correctly answer questions on the material being studied, while the scores on a norm-referenced test report how the student scored relative to other students in the group.

2. Criterion-Referenced Tests can be Managed Locally.

• Assessing student progress is something that every teacher must do. Criterion-referenced tests can be developed at the classroom level. If the standards are not met, teachers can specifically diagnose the deficiencies. Scores for an individual student are independent of how other students perform. In addition, test results can be quickly obtained to give students effective feedback on their performance. Although norm-referenced tests are most suitable for developing normative data across large groups, criterion-referenced tests can produce some local norms.

Disadvantages of CRT

• Criterion-referenced tests have some built-in disadvantages. Creating tests that are both valid and reliable requires fairly extensive and expensive time and effort. In addition, results cannot be generalized beyond the specific course or program. Such tests may also be compromised by students gaining access to test questions prior to exams. Criterion referenced tests are specific to a program and cannot be used to measure the performance of large groups.

Analysing Test Items

• Item analysis is used to measure the effectiveness of individual test items. The main purpose is to improve tests, to identify questions that are too easy, too difficult or too susceptible to guessing. While test items can be analyzed on both criterion-referenced and norm-referenced tests, the analysis is somewhat different because the purpose of the two types of tests is different.

Self-Check Exercise-1

1. What is the primary focus of a criterion-referenced test (CRT)?

a) Comparing a student's performance to that of other students.
b) Measuring how well students have mastered a particular body of knowledge.
c) Selecting items that maximize score variability.
d) Ensuring a wide distribution of scores.

2. Which of the following is an advantage of criterion-referenced tests (CRT)?
    a) They are inexpensive and quick to develop.
    b) They can be generalized beyond the specific course or program.
    c) They are more suitable for tracking student progress within a curriculum and diagnosing specific deficiencies.
    d) They are ideal for measuring the performance of large groups.

3. Fill in the blanks:

(i) Criterion reference tests have an established _____ scores.

(ii) _____ use criterion reference test to monitor student performance in their day to day activities.
(iii) Criterion reference tests are also used in the training programs to assess _____.

13.4 MEANING OF NORM REFERENCED TEST (NRT)

Definition: This type of test determines a student's placement on a normal distribution curve. Students compete against each other on this type of assessment. This is what is being referred to with the phrase, 'grading on a curve'. Norm-referenced tests allow us to compare a student's skills to others in his age group. Norm-referenced tests are developed by creating the test items and then administering the test to a group of students that will be used as the basis of comparison. Statistical methods are used to determine how raw scores will be interpreted and what performance levels are assigned to each score. Many tests yield standard scores, which allow comparison of the student's scores to other tests. They answer questions such as, "does the student's achievement score appear consistent with his cognitive score?" The degree of difference between those two scores might suggest or rule out a learning disability.

After the norming process, the tests are used to assess groups of students or individuals using standardized, or highly structured, administration procedures. These students' performance is rated using scales developed during the norming process.

Educators use norm-reference tests to evaluate the effectiveness of teaching programs, to help determine students' preparedness for programs, and to determine diagnosis of disabilities for eligibility for IDEA special education programs or adaptations and accommodations.

13.4.1 Features of Norm Referenced Test (NRT)

Norm-referenced tests (NRTs) compare a person's score against the scores of a group of people who have already taken the same exam, called the "norming group." When

you see scores in the paper which report a school's scores as a percentage -- "the Lincoln school ranked at the 49th percentile" -- or when you see your child's score reported that way -- "Jamal scored at the 63rd percentile" -- the test is usually an NRT.

Most achievement NRTs are multiple-choice tests: Some also include open-ended, short-answer questions. The questions on these tests mainly reflect the content of nationally-used textbooks, not the local curriculum. This means that students may be tested on things your local schools or state education department decided were not so important and therefore were not taught.

Creating the bell curve.

NRTs are designed to "rank-order" test takers -- that is, to compare students' scores: A commercial Norm-referenced test does not compare all the students who take the test in a given year. Instead, test-makers select a sample from the target student population (say, ninth graders). The test is "normed" on this sample, which is supposed to fairly represent the entire target population (all ninth graders in the nation). Students' scores are then reported in relation to the scores of this "norming" group.

To make comparing easier, test makers create exams in which the results end up looking at least somewhat like a bell-shaped curve :Testmakers make the test so that most students will score near the middle, and only a few will score low (the left side of the curve) or high (the right side of the curve).

Scores are usually reported as percentile ranks: The scores range from 1st percentile to 99th percentile, with the average student score set at the 50th percentile. If Jamal scored at the 63rd percentile, it means he scored higher than 63% of the test takers in the norming group. Scores also can be reported as "grade equivalents," "stanines," and "normal curve equivalents."

One more question right or wrong can cause a big change in the student's score : In some cases, having one more correct answer can cause a student's reported percentile score to jump more than ten points. It is very important to know how much difference in the percentile rank would be caused by getting one or two more questions right.

In making an NRT, it is often more important to choose questions that sort people along the curve than it is to make sure that the content covered by the test is adequate: The tests sometimes emphasize small and meaningless differences among test takers. Since the tests are made to sort students, most of the things everyone knows are not tested. Questions may be obscure or tricky, in order to help rank order the test takers.

Tests can be biased: Some questions may favor one kind of student or another for reasons that have nothing to do with the subject area being tested. Non-school knowledge that is more commonly learned by middle or upper class children is often included in tests. To help make the bell curve, test makers usually eliminate questions that students with low overall scores might get right but those with high overall scores get wrong. Thus, most questions which favor minority groups are eliminated.

NRTs usually have to be completed in a time limit: Some students do not finish, even if they know the material. This can be particularly unfair to students whose first language is not English or who have learning disabilities. This "speededness" is one way test makers sort people out.

13.4.2 Advantages and Disadvantages of Norm Referenced Test (NRT)

Advantages of NRT

To compare students, it is often easiest to use a Norm-Referenced Test because they were created to rank test-takers: It there are limited places (such as in a "Gifted and Talented" program) and choices have to be made, it is tempting to use a test constructed to rank students, even if the ranking is not very meaningful and keeps out some qualified children.

NRT's are a quick snapshot of some of the things most people expect students to learn: They are relatively cheap and easy to administer. If they were only used as one additional piece of information and not much importance was put on them, they would not be much of a problem.

Disadvantages NRT

The damage caused by using NRTs is far greater than any possible benefits the tests provide: The main purpose of NRTs is to rank and sort students, not to determine whether students have learned the material they have been taught. They do not measure anywhere near enough of what students should learn. They have very harmful effects on curriculum and instruction. In the end, they provide a distorted view of learning that then causes damage to Norm-Referenced Measures (NRM)

Most appropriate when one wishes to make comparisons across large numbers of students or important decisions regarding student placement and advancement. Norm-referenced measures are designed to compare students (i.e., disperse average student scores along a importance placed upon high scores, the content of a standardized test can be very influential in the development of a school's curriculum and standards of excellence.

The testing profession, in its Standards for Educational and Psychological Measurement, states, "In elementary or secondary education, a decision or characterization that will have a major impact on a test taker should not automatically be made on the basis of a single test score."

Any one test can only measure a limited part of a subject area or a limited range of important human abilities: A "reading" test may measure only some particular reading "skills," not a full range of the ability to understand and use texts. Multiple-choice math tests can measure skill in computation or solving routine problems, but they are not good for assessing whether students can reason mathematically and apply their knowledge to new, real-world problems.

Most NRTs focus too heavily on memorization and routine procedures: Tests like these cannot show whether a student can write a research paper, use history to help understand current events, understand the impact of science on

society, or debate important issues. They don't test problem solving, decision-making, judgement, or social skills.

Tests often cause teachers to overemphasize memorization and de-emphasize thinking and application of knowledge: Since the tests are very limited, teaching to them narrows instruction and weakens curriculum. Making test score gains the definition of "improvement" often guarantees that schooling becomes test coaching. As a result, students are deprived of the quality education they deserve.

Norm-referenced tests also can lower academic expectations: NRTs support the idea that learning or intelligence fits a bell curve. If educators believe it, they are more likely to have low expectations of students who score below average.

Self-Check-Exercise-2

1. Which feature is associated with norm-referenced tests (NRTs)?

a) Items are selected to ensure they align with the curriculum being taught locally.
b) Scores are often reported as percentile ranks.
c) Focuses on ensuring all students meet a predetermined standard.
d) Emphasizes individual mastery of subject matter.

2. What is a significant disadvantage of norm-referenced tests (NRTs)?

a) They are expensive and time-consuming to develop.
b) They often cause teachers to overemphasize memorization.
c) They are not suitable for large-scale assessments.
d) They provide immediate feedback to students and teachers.

3. Fill in the blanks:

(i) Norm-referenced tests (NRTs) determine a student's placement on a _____ distribution curve.

(ii) Scores on NRTs are usually reported as _____ ranks, which range from 1st percentile to 99th percentile.

(iii) One disadvantage of NRTs is that they often cause teachers to overemphasize _____ and de-emphasize thinking and application of knowledge.

13.5 Summary

In summary, NRTs compare individuals to a norm group, while CRTs evaluate individuals against specific criteria or standards. NRTs focus on relative standing, while CRTs focus on absolute achievement. Both NRT's and CRT's used to evaluate the performance of learners and determine whether they have

failed or excelled in their tests. It is after this that the students can be held accountable and told to re-sit their tests.

13.6 Glossary

Norm Group: A large, representative sample of individuals used as a comparison group for NRTs.

Standards: Specific learning objectives or criteria used to evaluate performance on CRTs.

Mastery: Demonstrating a predetermined level of proficiency or competence on a CRT.

13.7 ANSWER TO SELF-CHECK EXERCISE 1 & 2.

Self-Check-Exercise-1

1. b) Measuring how well students have mastered a particular body of knowledge
2. c) They are more suitable for tracking student progress within a curriculum and diagnosing specific deficiencies.

3. (i) passing

   (ii) classroom teachers

   (iii) learning

Self-Check-Exercise-2

1. b) Scores are often reported as percentile ranks.
2. b) They often cause teachers to overemphasize memorization.
3. (i) normal
   (ii) percentile
   (iii) memorisation

13.8 REFERENCES/SUGGESTED READING

- Taiwo, Adediran A. (2005). Fundamentals of Classroom Testing. New Delhi: Vikas Publishing House Pvt. Ltd.
- Walter W. Cook (1958). Educational Measurement. Washington D.C.: American Council on Education.
- Withers, G. (1997).Item writing for Tests and Examinations. Paris: International Institute for Educational Planning.

13.9 TERMINAL QUESTIONS

Dear learners, please check you progress by attempting the following questions:

1. Discuss the primary differences between Norm-Referenced Tests (NRTs) and Criterion-Referenced Tests (CRTs).
2. Evaluate the advantages and disadvantages of Criterion-Referenced Tests (CRTs).
3. Explain the feature of criterion referenced test.
4. What are the features of norm referenced test?

5. Give the advantages and disadvantages of norm referenced test?

****

UNIT-14
QUESTIONNAIRE AND SCHEDULES

Structure

14.1 INTRODUCTION

Dear Learner,

A questionnaire is a research tool consisting of questions to collect data from respondents, using both close-ended and open-ended formats. A well-designed questionnaire features a limited number of questions, logical sequencing, simplicity, and clear instructions. Types include structured, unstructured, and open-ended questionnaires. Designing one involves defining goals, identifying respondents, developing questions, and conducting a pilot test. Advantages include standardization, efficiency, anonymity, cost-effectiveness, objectivity, and flexibility, but limitations are limited depth, response bias, and lack of context. In contrast, a schedule is filled out by researchers based on informant responses, offering more control and depth but with limited reach and requiring researcher involvement. Understanding these tools aids in selecting the appropriate method for effective data collection.

14.2 LEARNING OBJECTIVES

After studying this unit, you will be able to:

- Describe the characteristics of a good questionnaire.
- Explain various types of questionnaires.
- Explain in brief, the steps of designing a questionnaire.
- Discuss advantages and limitations of questionnaire.
- Explain Schedule.
- Compare between questionnaire and schedule.

14.3 QUESTIONNAIRE

A questionnaire is a research instrument that consists of a set of questions or other types of prompts that aims to collect information from a respondent. A research questionnaire is typically a mix of close-ended questions and open-ended questions.

Open-ended, long-form questions offer the respondent the ability to elaborate on their thoughts. Research questionnaires were developed in 1838 by the Statistical Society of London.

14.3.1 Characteristics of a well-designed questionnaire

Here are the key characteristics of a well-designed questionnaire:

1. **Limited Number of Questions**: Keep the number of questions as limited as possible, focusing only on those relevant to the inquiry.
2. **Proper Sequence of Questions**: Arrange questions logically, placing simple and direct ones at the start and more complex or indirect ones toward the end.
3. **Simplicity**: Use clear, easy-to-understand language with short questions. Avoid complexity.
4. **Clear Instructions**: Provide explicit instructions for filling out the form

14.3.2 Types of questionnaires

Let's explore the different types of questionnaires:

1. **Structured Questionnaire**:
    - This type of questionnaire has a fixed format with predetermined questions that the respondent must answer.
    - The questions are usually **closed-ended**, meaning that the respondent selects a response from a list of options (e.g., yes/no or multiple-choice questions).
2. **Unstructured Questionnaire**:
    - An unstructured questionnaire does not have a fixed format or predetermined questions.
    - Instead, the interviewer or researcher can ask **open-ended questions**, allowing respondents to provide their own answers without predefined options.
3. **Open-ended Questionnaire**:
    - In an open-ended questionnaire, respondents answer questions in their own words, without any pre-determined response options.
    - This format allows for more details and qualitative responses.

14.3.3 Steps of Questionnaire:

Designing a questionnaire involves several key steps:

1. **Define Goals and Objectives**:
    - Clarify what you want to study and the specific topics or experiences you aim to explore.
    - Consider the purpose of the questionnaire (e.g., research, feedback, assessment).
2. **Identify Target Respondents**:
    - Understand your audience: Who are the respondents? What characteristics do they have?
    - Tailor the questionnaire to their background, interests, and needs.
3. **Develop Questions**:
    - Create valid and reliable questions aligned with your research objectives.
    - Use clear and concise language.
    - Consider both closed-ended (predefined options) and open-ended questions.
4. **Choose Question Types**:
    - Closed-ended questions: Provide response options (e.g., yes/no, Likert scale).
    - Open-ended questions: Allow respondents to answer in their own words.
5. **Design Question Sequence and Layout**:
    - Arrange questions logically.
    - Start with simple, non-threatening questions.
    - Group related questions together.
6. **Run a Pilot**:
    - Test the questionnaire with a small group to identify any issues.
    - Revise as needed based on feedback.

14.3.4 Advantages and Limitations of Questionnaire

Some Advantage of Questionnaire are as follows:

- **Standardization:** Questionnaires allow researchers to ask the same questions to all participants in a standardized manner. This helps ensure consistency in the data collected and eliminates potential bias that might arise if questions were asked differently to different participants.
- **Efficiency:** Questionnaires can be administered to a large number of people at once, making them an efficient way to collect data from a large sample.
- **Anonymity:** Participants can remain anonymous when completing a questionnaire, which may make them more likely to answer honestly and openly.
- **Cost-effective:** Questionnaires can be relatively inexpensive to administer compared to other research methods, such as interviews or focus groups.
- **Objectivity:** Because questionnaires are typically designed to collect quantitative data, they can be analyzed objectively without the influence of the researcher's subjective interpretation.

- **Flexibility:** Questionnaires can be adapted to a wide range of research questions and can be used in various settings, including online surveys, mail surveys, or in-person interviews.

Limitations of Questionnaire:

Limitations of Questionnaire are as follows:

- **Limited depth:** Questionnaires are typically designed to collect quantitative data, which may not provide a complete understanding of the topic being studied. Questionnaires may miss important details and nuances that could be captured through other research methods, such as interviews or observations.
- R**esponse bias:** Participants may not always answer questions truthfully or accurately, either because they do not remember or because they want to present themselves in a particular way. This can lead to response bias, which can affect the validity and reliability of the data collected.
- **Limited flexibility:** While questionnaires can be adapted to a wide range of research questions, they may not be suitable for all types of research. For example, they may not be appropriate for studying complex phenomena or for exploring participants' experiences and perceptions in-depth.
- **Limited context:** Questionnaires typically do not provide a rich contextual understanding of the topic being studied. They may not capture the broader social, cultural, or historical factors that may influence participants' responses.
- **Limited control**: Researchers may not have control over how participants complete the questionnaire, which can lead to variations in response quality or consistency.

Self-Check Exercise-1

1. Which of the following is NOT a characteristic of a well-designed questionnaire?

a) Limited number of questions
b) Proper sequence of questions
c) Use of complex and lengthy language
d) Clear instructions

2. What type of questionnaire includes a fixed format with predetermined questions?

a) Structured questionnaire
b) Unstructured questionnaire
c) Open-ended questionnaire
d) None of the above

Fill in the blanks;

(i) Questionnaires are typically a mix of _____ and _____ questions, allowing for a range of responses.

(ii) One key advantage of questionnaires is their ability to ensure _____, asking the same questions to all participants consistently.

(iii) One limitation of questionnaires is their potential for _____ bias, where participants may not answer questions truthfully or accurately.

14.4 SCHEDULE, COMPARISON BETWEEN QUESTIONNAIRE AND SCHEDULE

Like the questionnaire, schedule is technique of data collection, which contains a list of questions. The difference between a questionnaire and schedule, however, is that while the former is filled by respondents, the latter is filled by the researcher. The researcher goes to the informants with the schedule, and asks them the questions. Researcher plays an important role in the collection of data, through schedules. They explain the aims and objects of the research to the respondents and interpret the questions to them when required. Most common example of data collection through schedule is population census. The main advantage of schedule is the presence of the researcher. In simple terms, the researcher could explain the question in detail, seek additional information (i.e., information beyond the questions listed in the schedule), obtain clarification on the response, may change the sequence, language and style of questions. While framing a schedule, the researcher has to take many aspects into consideration. In fact, it is appropriate to identify the aspects on which the schedule needs to be prepared. These aspects are logically arranged and relevant questions are framed. It is likely that more than one question is asked on an aspect with the purpose of obtaining complete information.

Comparison between Questionnaire and Schedule

There are certain differences between questionnaire and schedule as listed below.

A questionnaire is filled by the respondents, while the researcher fills the schedule.

Questionnaire is more rigid in structure than schedule. Researchers have no control over response rate in case of questionnaires as many people do not respond and/or often return them without answering all the questions. On the contrary, researchers have control over the response rate of schedules since they collect data themselves.

- While questionnaire has a larger reach since it can be distributed to a large number of people at the same time, schedule has a limited reach.
- Identity of respondents is protected when data is collected using questionnaire technique while identity of informants is revealed when data is collected using schedule technique of data collection.
- The success of the questionnaire depends much on the quality of the questionnaire while the research acumen and experience of the researcher determines the success of a schedule.
- The questionnaire can be employed only when the respondents are literate while schedule can be used for data collection from both literate and illiterate informants. Possibility of obtaining incomplete and imprecise information is relatively more when data is collected through

questionnaires than through schedules since the researcher is present in the field situation to verify and corroborate data there and-then.

Self-Check Exercise-2

1. Which of the following statements is true about schedules?

a) They are filled out by the respondents themselves.
b) They allow the researcher to seek additional information and obtain clarifications on responses.
c) They have a larger reach compared to questionnaires.
d) The identity of the informants is protected.

2. One of the main advantages of using a schedule over a questionnaire is that:

a) It ensures the anonymity of the respondents.
b) It can be distributed to a large number of people simultaneously.
c) It allows the researcher to control the response rate by collecting data themselves.
d) It requires respondents to be literate.

14.5 Summary

The questionnaire method presents researchers with a versatile and potent tool for data collection across diverse research domains. Its structured format enables standardized data collection, organization, and analysis, particularly beneficial for quantitative research endeavors. With advantages such as cost-effectiveness, accessibility, and the capacity to reach a broad and diverse population, questionnaires offer researchers an efficient means of gathering comprehensive insights. However, it is essential to acknowledge the disadvantages associated with this method, including low response rates, potential bias due to non-response, and challenges in ensuring the representativeness of respondents. Despite these drawbacks, the questionnaire remains a valuable instrument in the research arsenal, providing a structured approach to gathering insights and contributing to the advancement of knowledge in various fields. In contrast, schedules have limited reach, reveal informant identity, and depend on the researcher's expertise. Questionnaires require literate respondents and may yield incomplete information, while schedules can be used with both literate and illiterate informants and ensure more accurate data collection.

14.6 Glossary

Survey: The process of studying a phenomenon. A questionnaire is designed to generate data that measures various attributes related to the phenomenon.

Administration: The process of managing the survey, including selecting the audience, extending invitations, collecting responses, and loading data.

14.7 Answers to Self-Check Exercise 1 & 2.

Self-Check Exercise-1

1. c) Use of complex and lengthy language
2. a) Structured questionnaire

3. (i) close-ended, open-ended
   (ii) standardization
   (iii) response

Self-Check Exercise-2

1. b) They allow the researcher to seek additional information and obtain clarifications on responses.
2. c) It allows the researcher to control the response rate by collecting data themselves.

14.8 REFERENCES/SUGGESTED READINGS

- Izard, J. (1997). Content Analysis and Test Blueprints. Paris-international Institute for Educational Planning.
- Mehrens, WA. and Lehmann, I.J. (1984). Measurement and evaluation in education and psychology.(3rd Ed.) New York: Holt, Rinehart and Winston.
- Nandra, I.D.S.(2011). Learning Resources and Assessment of Learning.Patiala, 21s* Century Publications.
- Ebel, Robert L.(1966) “Measuring Educational Achievement, Prentice Hall of India Pvt. Ltd. Pp. 481

14.9 TERMINAL QUESTIONS

Dear learners, please check you progress by attempting the following questions:

1. Explain the characteristics of a good questionnaire.

2. What things will you keep in mind while designing a questionnaire?

3. Discuss about advantages and disadvantages of questionnaire.
4. Explain the main differences between questionnaires and schedules as techniques of data collection. Provide examples to illustrate these differences.

****

UNIT-15
RATING SCALE, ATTITUDE SCALE AND PERFORMANCE TEST

Structure

15.1 INTRODUCTION

Dear Learner,

Rating scales quantify subjective judgments in research and assessment, using a continuum of values for traits, performances, or phenomena. Major approaches include paired comparison, ranking, and rating scales. Attitude scales, such as Thurstone and Likert, convert qualitative attitudes into quantitative data, offering insights in educational and social research. Performance tests require individuals to demonstrate skills through tasks, emphasizing problem-solving, critical thinking, and real-world application, making them essential for evaluating and guiding students' proficiency in educational settings.

15.2 LEARNING OBJECTIVES

After studying this unit, you will be able to:

- Explain meaning and purpose of a rating scale.
- Understand precautions while constructing a rating scale.
- Discuss the uses and limitations of a rating scale.
- Describe Likert method and Thurston method of measuring attitudes.
- Explain purpose and limitations of attitude scales.

15.3 RATING SCALE

Rating scale is one of the enquiry form. Form is a term applied to expression or judgment regarding some situation, object or character. Opinions are usually expressed on a scale of values. Rating techniques are devices by which such judgments may be quantified. Rating scale is a very

useful device in assessing quality, especially when quality is difficult to measure objectively. For Example, —How good was the performance? It is a question which can hardly be answered objectively. Rating scales record judgment or opinions and indicates the' degree or amount of different degrees of qualify which are arranged along a line is the scale. For example: How good was the performance?

Excellent Very good Good Average Below average Poor Very poor

_____

This is the most commonly used instrument for making appraisals. It has a large variety of forms and uses. Typically, they direct attention to a number of aspects or traits of the thing to be rated and provide a scale for assigning values to each of the aspects selected. They try to measure the nature or degree of certain aspects or characteristics of a person or phenomenon through the use of a series of numbers, qualitative terms or verbal descriptions. Ratings can be obtained through one of three major approaches:

- Paired comparison
- Ranking and
- Rating scales

The first attempt at rating personality characteristics was the man to man technique devised during World-war-1. This technique calls for a panel of raters to rate every individual in comparison to a standard person. This is known as the paired comparison approach. In the ranking approach every single individual in a group is compared with every other individual and to arrange the judgment in the form of a scale. In the rating scale approach which is the more common and practical method rating is based on the rating scales, a procedure which consists of assigning to each trait being rated a scale value giving a valid estimate of its status and then comparing the separate ratings into an overall score.

Purpose of Rating Scale: Rating scales have been successfully utilized for measuring the following:

- Teacher Performance/Effectiveness
- Personality, anxiety, stress, emotional intelligence etc.
- School appraisal including appraisal of courses, practices and programmes.

Useful Hints on Construction of Rating Scale: A rating scale includes three factors like: i) The subjects or the phenomena to be rated, ii). The continuum along which they will be rated and iii) The judges who will do the rating. All

taken three factors should be carefully taken care by you when you construct the rating scale. 1) The subjects or phenomena to be rated are usually a limited number of aspects of a thing or of a traits of a person. Only the most significant aspects for the purpose of the study should be chosen. The usual may to get judgment is on five to seven point scales as we have already discussed. 2) The rating scale is always composed of two parts: i) An instruction which names the subject and defines the continuum and ii) A scale which defines the points to be used in rating. 3) Any one can serve as a rater where non-technical opinions, likes and dislikes and matters of easy observation are to be rated. But only well informed and experienced persons should be selected for rating where technical competence is required. Therefore, you should select experts in the field as rater or a person who form a sample of the population in which the scale will subsequently be applied. Pooled judgments increase the reliability of any rating scale. So employ several judges, depending on the rating situation to obtain desirable reliability.

Use of Rating Scale: Rating scales are used for testing the validity of many objective instruments like paper pencil inventories of personality. They are also advantages in the following fields like:

- Helpful in writing reports to parents
- Helpful in filling out admission blanks for colleges
- Helpful in finding out student needs
- Making recommendations to employers.
- Supplementing other sources of understanding about the child
- Stimulating effect upon the individuals who are rated.

Limitations of Rating Scale: The rating scales suffer from many errors and limitations like the following: As you know that the raters would not like to run down their own people by giving them low ratings. So in that case they give high ratings to almost all cases. Sometimes also the raters are included to be unduly generous in rating aspects which they had to opportunity to observe. If the raters rate in higher side due to those factors, then it is called as the generosity error of rating.

The Errors of Central Tendency: Some observes wants to keep them in safe position. Therefore, they rate near the midpoint of the scale. They rate almost all as average.

- **Stringency Error:** Stringency error is just the opposite of generosity of error. These types of raters are very strict, cautions and hesitant in rating in average and higher side. They have a tendency to rate all individuals low.
- The Hallo Error: When a rater rates one aspect influenced by other is called hallo effect. For if a person will be rated in higher side on his achievement because of his punctually or sincerely irrespective of his

perfect answer it called as hallo effect. The biasedness of the rater affects from one quality to other.

- **The Logical Error:** It is difficult to convey to the rater just what quality one wishes him to evaluate. An adjective or Adverb may have no universal meaning. It the terms are not properly understood by the rater and he rates, then it is called as the logical error. Therefore, brief behavioural statements having clear objectives should be used.

Self-Check Exercise-1

1. What is a rating scale?
a) A method of qualitative data collection
b) A method of quantitative data collection
c) A method of data analysis
d) A method of data visualization

2. Rating Scale are used to record judgements about
   a) oneself
   b) objects
   c) others
   d) All of the above

15.4 ATTITUDE SCALE

Attitude scale is a form of appraisal procedure and it is also one of the enquiry term. Attitude scales have been designed to measure attitude of a subject of group of subjects towards issues, institutions and group of peoples. The term attitude is defined in various ways:

The behaviour which we define as attitudinal or attitude is a certain observable set II organism or relative tendency preparatory to and indicative of more complete adjustment.

-L. L. Bernard

An attitude may be defined as a learned emotional response set for or against something. - Barr David Johnson

An attitude is spoken of as a tendency of an individual to read in a certain way towards a Phenomenon. It is y/hat a person feels or believes in. It is the inner feeling of an individual. It may be positive, negative or neutral. Opinion and attitude are used sometimes in a synonymous manner but there is a difference between two. You will be able to know when we will discuss about opinionnaire. An opinion may not lead to any kind of activity in a particular direction. But an attitude compels one to act either favourably or unfavourably

according to what they perceive to bo correct. We can evaluate attitude through questionnaire. But it is ill adapted for scaling accurately the intensity

of an attitude. Therefore, Attitude scale is essential as it attempts to minimise the difficulty of opinionrtaire and questionnaire by defining the attitude in terms of a single attitude object. All items, therefore, may be constructed with graduations of favour or disfavour.

Purpose of Attitude Scale: In educational research, these scales are used especially for finding the attitudes of persons on different issues like:

- Co-education
- Religious education
- Corporal punishment
- Democracy in schools
- Linguistic prejudices
- International co-operation etc.'

Characteristics of Attitude Scale: Attitude scale should have the following characteristics.

- It provides for quantitative measure on a uni-dimensional scale of continuum.
- It uses statements from the extreme positive to extreme negative position.
- It generally uses a five point scale as we have discussed in rating scale.
- It could be standardized and norms are worked out.
- It disguises the attitude object rather than directly asking about the attitude on the subject.

Examples of Some Attitude Scale: Two popular and useful methods of measuring attitudes indirectly, commonly used for research purposes are:

1. Thurstone Techniques of scaled values.
2. Likert's method of summated ratings. .

1. **Thurstone Technique:** Thurstone Technique is used when attitude is accepted as a uni-dimensional linear Continuum. The procedure is simple. A large number of statements of various shades of favourable and unfavourable opinion on slips of paper, which a large number of judges exercising complete detachment sort out into eleven plies ranging from the most hostile statements to the most favourable ones. The opinions are carefully worded so as to be clear and unequivocal. The judges are asked not express tier opinion but to sort them at their face value. The items

which bring out a marked disagreement between the judges un assigning a position are discarded. Tabulations are made, which indicate the number of judges who placed each item in each category. The next step consists of calculating cumulated proportions for each item and ogive are constructed. Scale values of each item are read from the Ogive, the values of each item being that point along the baseline in terms of scale value units above and below which 50% of the judges placed the item. It we'll be the median of the frequency distribution in which the score ranges from 0 to 11. The respondent is to give his reaction to each statement by endorsing or rejecting it. The median values of the statements that he checks establishes his score, or quantifies his opinion. He wins a score as an average of the sum of the values of the statements he endorses. Thurstone technique is also known as the technique equal appearing intervals.

2. **The Likert Scale:** The Likert scale uses items worded for or against the proposition, with five point rating response indicating the strength of the respondent's approval or disapproval of the statement. This method removes the necessity of submitting items to the judges for working out scaled values for each item. It yields scores very similar to those obtained from the Thurstone scale. It is an important over the Thurstone method. The first step is the collection of a member of statements about the subject in. question. Statements may or may not be correct but they must be representative of opinion held by a substantial number of people. They must express definite favourableness or unfavourableness to a particular point of view. The number of favourable and unfavourable statements should be approximately equal. A trial test maybe administered to a number of subjects. Only those items that correlate with the total test should be retained. The Likert’s calling techniques assigns a scale value to each of the five responses. All favourable statements are scored from maximum to minimum i. e. from a score of 5 to a score of one or 5 for strongly agree and so on 1 for strongly disagree. The negative statement or statement opposing the proposition would be scored in the opposite order. e. from a score of 1 to a score of 5 or 1 for strongly agree and so on 5 for strongly disagree. The total of these scores on all the items measures a respondent's favourableness towards the subject in question. It a scale consists of 30 items, Say, the following score values will be of interest.

Limitations of Attitude Scale: In the attitude scale, the following limitations may occur;

- An individual may express socially acceptable opinion conceal his real attitude.
- An individual may not be a good judge of himself and may not be clearly aware of his real attitude.
- He may not have been controlled with a real situation to discover what his real attitude towards a specific phenomenon was.

- There is no basis for believing that the five positions indicated in the Likert's scale are equally spaced.
- It is unlikely that the statements are of equal value in‘for’and‘against’.
- It is doubtful whether equal scores obtained by several individuals would indicate equal favourableness towards again position.
- It is unlikely that respondent can validity react to a short statement on a printed form in the absence of real like qualifying Situation.
- In spite of anonymity of response, individuals tend to respond according to what they should feel rather than what they really feel.
- However, until more precise measures are developed, attitude scale remains the best device for the purpose of measuring attitudes and beliefs in social research.

Self-Check Exercise-2

1) Another name for a **Likert Scale** is:

a) Interview protocol
b) Event sampling
c) Summated rating scale
d) Ranking

2) The easiest attitudinal scale, which is a summated rating scale, is the:
a) Guttman scale
b) Likert Scale
c) Thurstone scale
d) MLA scale

15.5 PERFORMANCE TEST

It is an assessment that requires an examinee to actually perform a task or activity, rather than simply answering questions referring to specific parts.it requires students to demonstrate that they have mastered specific skills and competencies by performing or producing something.

Performance test- is non-standardized test. (Non-standardized test is usually flexible in scope and format, variable in difficulty and significance)

•It can be used to determine the proficiency level of students, to motivate students to study and to provide feedbacks to the students

•It also allows teachers to observe achievements, habits of mind, ways of working and behavior of value in the real world. In many cases, these are outcomes that conventional tests may miss.

The following characteristics should be remembered when designing a performance task:

•It has various outcomes; it does not require one right answer.

•It is integrative, combining different skills.

•It encourages problem-solving and critical thinking skills.

•It encourages divergent thinking.

•It focuses on both product and process

•It promotes independent learning, involving planning, revising and summation.

•It builds on pupils' prior experience.

•It can include opportunities for peer interaction and collaborative learning.

•It enables self-assessment and reflection.

•It is interesting, challenging, meaningful and authentic.

•It requires time to complete.

How to Design and Assess a Performance Task

Step 1. List the specific skills and knowledge you wish pupils to demonstrate.

Step 2. Design a performance task that requires pupils to demonstrate these skills and this knowledge.

Step 3. Develop explicit performance criteria and expected performance levels measuring pupils' mastery of skills and knowledge (rubrics).

Example of Performance Test

•Oral presentation

•Dance/movement

•Science lab demonstration

•Athletic competition

•Dramatic reading

•Enactment

•Debate

•Musical recital

•Tableau

Self-Check Exercise-3

1. What is a primary characteristic of a performance test?

a) It focuses on answering multiple-choice questions.

b) It requires examinees to perform tasks or activities.

c) It is highly standardized with a fixed format.

d) It provides only written feedback.

2. What is the first step in designing a performance task?

a) Develop explicit performance criteria and expected performance levels.

b) List the specific skills and knowledge you wish pupils to demonstrate.

c) Design a performance task that requires pupils to demonstrate these skills and knowledge.

d) Administer the task to a trial group of students.

15.6 SUMMARY

After going through this lesson, you must have understood about the meaning of a questionnaire, its characteristics and designing process. You were also acquainted with rating scales, its purpose, uses and limitations. In the last part of the lesson, we discussed about attitude scales, its purpose, limitations and techniques of measuring attitudes. You also understood the concept of performance test and their examples.

15.7 GLOSSARY

Rating Scale: A rating scale is *a method that requires the rater to assign a value, sometimes numeric, to the rated object, as* a measure of some rated attribute.

Likert Scale: A Likert scale is a rating scale used to measure opinions, attitudes, or behaviors.

Performance test: A performance test, is an educational assessment that requires students to demonstrate their knowledge, skills, and abilities through tasks, projects, or presentations.

15.8 ANSWER TO SELF-CHECK EXERCISE-1, 2 & 3.

Self-Check Exercise-1

1. b) A method of quantitative data collection

2. d) All of the above

Self-Check Exercise-2

1. c) Summated rating scale

2. d) MLA scale

Self-Check Exercise-3

1. b) It requires examinees to perform tasks or activities
2. b) List the specific skills and knowledge you wish pupils to demonstrate.

15.8 REFERENCES/SUGGESTED READINGS

- Mehrens, W A. and Lehmann, I.J. (1984).Measurement and evaluation in education and psychology.(3rd Ed.) New York: Holt, Rinehart and Winston.
- Nandra, I.D.S.(2011). Learning Resources and Assessment of Learning.Patiala, 21st Century Publications.
- Taiwo, Adediran A. (2005). Fundamentals of Classroom Testing. New Delhi: Vikas Publishing House Pvt. Ltd.
- Walter W.-Cook. (1958). Educational Measurement. Washington D.C.: American Council on Education.
- Withers, G. (1997).Item writing for Tests and Examinations. Paris: International Institute for Educational Planning.

15.9 TERMINAL QUESTIONS

Dear learners, please check you progress by attempting the following questions:

1. What do you mean by rating scale?
2. Discuss the uses and limitations of a rating scale.
3. What do you mean by attitude scale?
4. Describe Likert method of measuring attitudes.
5. Write down limitations of attitude scales.
6. Describe Performance test with examples.

UNIT-16

ACHIEVEMENT TESTS

Structure

16.1 INTRODUCTION

Dear learner,

Achievement tests are designed to measure the knowledge, skills, and abilities that individuals have acquired in a specific area of study. They are crucial tools in educational and training settings, providing valuable information on how well students have understood and can apply what they have learned. This unit explores the meaning, characteristics, advantages, disadvantages, and types of achievement tests.

16.2 LEARNING OBJECTIVES

After studying this unit, you will be able to:

- Define achievement tests.
- Describe the characteristics of achievement tests.
- Discuss the advantages and disadvantages of achievement tests.
- Identify and explain different types of achievement tests.

16.3 MEANING AND CHARACTERISTICS OF ACHIEVEMENT TESTS

A test is an instrument or a tool. It follows a systematic procedure for measuring a sample of behaviour by posing a set of questions in a uniform manner. It is an attempt to measure what a person knows or can do at a

particular point in time. Furthermore, a test answers the question ‘how well’ does the individual perform either in comparison with others or in comparison with a domain of performance tasks? A test designed to appraise what the individual has learned as a result of planned previous experience or training is an Achievement Test. Since it relates to what has been learnt already its frame of reference is on the present or past. Achievement tests attempt to measure what a person knows or can do at a particular point in time. Furthermore, our reference is usually to the past; that is we are interested in what has been learned as a result of a particular course or experience or a series of experiences. An achievement test is designed to evaluate a person's knowledge, skills, and proficiency in a specific area of study or subject matter. It aims to measure how well individuals have learned and can apply what they have been taught. Achievement tests are used teachers to measure or test the achievements and success achieved in any particular field by the students. Whatever the student learns in school is called his achievement and examinations conducted to test that achievement are called achievement tests.

Characteristics of Achievement tests

Achievement tests have several key characteristics that distinguish them from other types of assessments:

1. **Standardization**: Achievement tests are standardized, meaning they are administered and scored in a consistent manner to ensure fairness and comparability of results.
2. **Objective Measurement**: These tests aim to provide an objective measurement of students' knowledge and skills, reducing the influence of subjective factors.
3. **Content Specificity**: The content of achievement tests is aligned with the curriculum or training program, focusing on specific knowledge areas and skills.
4. **Reliability**: Achievement tests are designed to be reliable, providing consistent results over repeated administrations.
5. **Validity**: These tests measure what they are intended to measure, ensuring that the test content accurately reflects the learning objectives.

Self-Check Exercise –1

1. What is an achievement test designed to measure?

   a) Future potential

   b) Innate intelligence

   c) Past learning and proficiency

   d) Social skills

2. Which of the following is NOT a characteristic of achievement tests?

   a) Standardization

   b) Subjectivity

   c) Content specificity

   d) Reliability

16.4 ADVANTAGES AND DISADVANTAGES OF ACHIEVEMENT TESTS

Achievement tests offer several **advantages** in educational and professional settings:

1. **Objective Assessment**: Provides an objective evaluation of students' knowledge and skills, reducing biases in assessment.
2. **Standardized Comparison**: Enables comparison of performance across different groups of students or individuals, providing benchmarks for achievement.
3. **Feedback for Improvement**: Offers valuable feedback to students, teachers, and administrators on areas of strength and areas needing improvement.
4. **Motivation**: Encourages students to study and perform well, as they know their achievements will be measured.
5. **Accountability**: Holds educators and institutions accountable for students' learning outcomes.

Disadvantages of Achievement tests

Despite their advantages, achievement tests also have some limitations:

1. **Limited Scope**: May not capture all aspects of a student's abilities or knowledge, particularly skills that are difficult to measure objectively.
2. **Test Anxiety**: Can cause anxiety and stress among students, which may affect their performance.
3. **Teaching to the Test**: May encourage teaching practices that focus primarily on test content, neglecting broader educational goals.
4. **Cultural Bias**: Standardized tests may contain cultural biases that disadvantage certain groups of students.
5. **Resource Intensive**: Developing, administering, and scoring achievement tests can be resource-intensive in terms of time and cost.

Self-Check Exercise-2

1. How do achievement tests benefit educators and institutions?

a) By increasing the overall number of tests administered annually.
b) By offering financial rewards for high test scores.
c) By holding educators and institutions accountable for students' learning outcomes.
d) By reducing the need for curriculum development.

2. What is a major disadvantage of achievement tests?
   a) Objective assessment
   b) Test anxiety
   c) Feedback for improvement
   d) Standardized comparison

16.5 TYPES OF ACHIEVEMENT TESTS

There are several types of achievement tests, each serving different purposes and assessing different aspects of learning:

1. **Diagnostic Tests**: Used to identify students' strengths and weaknesses in specific areas before instruction begins. They help teachers plan targeted interventions and support.
2. **Formative Tests**: Administered during the instructional process to monitor students' progress and provide ongoing feedback. They help teachers adjust their teaching strategies to improve learning outcomes.
3. **Summative Tests**: Given at the end of an instructional period to evaluate students' overall learning and achievement of course objectives. Examples include final exams and standardized tests.
4. **Criterion-Referenced Tests**: Measure students' performance against a fixed set of criteria or learning standards. They determine whether students have mastered specific skills or knowledge areas.
5. **Norm-Referenced Tests**: Compare a student's performance to that of a larger group (norm group). These tests rank students and provide information on how they perform relative to others.
6. **Performance-Based Tests**: Require students to perform tasks or produce work that demonstrates their knowledge and skills. Examples include projects, presentations, and portfolios.

Self-Check Exercise-3

1. Which type of test is used to identify students' strengths and weaknesses before instruction begins?

   a) Summative Tests
   b) Formative Tests
   c) Diagnostic Tests
   d) Performance-Based Tests

2. Which of the following tests is designed to measure students' performance against a fixed set of criteria or learning standards?

a) Norm-Referenced Tests
b) Criterion-Referenced Tests
c) Diagnostic Tests
d) Formative Tests

16.6 SUMMARY

Achievement tests are standardized assessments designed to measure the knowledge, skills, and proficiency individuals have acquired in specific areas of study. They are characterized by their standardization, objectivity, content specificity, reliability, and validity. While they offer several advantages, such as objective assessment and feedback for improvement, they also have limitations, including limited scope and potential for test anxiety. Understanding the different types of achievement tests helps in selecting the appropriate test for specific assessment needs.

16.7 GLOSSARY

- **Achievement Test**: A standardized test designed to measure the knowledge, skills, and proficiency that an individual has acquired in a specific subject area.
- **Standardization**: The process of administering and scoring a test in a consistent manner to ensure fairness and comparability of results.
- **Reliability**: The extent to which a test provides consistent results over repeated administrations.
- **Validity**: The degree to which a test measures what it is intended to measure.

16.8 ANSWERS TO SELF-CHECK EXERCISE

Self-Check Exercise-1

1. c) Past learning and proficiency

2. b) Subjectivity

Self-Check Exercise-2

1. c) By holding educators and institutions accountable for students' learning outcomes.

2. b) Test anxiety

Self-Check Exercise-3

1. c) Diagnostic Tests

2. b) Criterion-Referenced Tests

16.9 REFERENCES/SUGGESTIVE READINGS

- Aggarwal, J. C. (2007). *Educational Measurement and Evaluation*. Delhi: Vikas Publishing House.
- Dandekar, W. N. (2004). *Measurement, Evaluation, and Assessment in Education*. Pune: Vidya Prakashan.
- Verma, J. P. (2013). *Essentials of Educational Measurement*. New Delhi: Sage Publications.
- Kamte, V. (2015). *Educational Testing and Measurement*. Mumbai: Himalaya Publishing House.

16.9 TERMINAL QUESTIONS

Dear learners, please check you progress by attempting the following questions:

1. Define achievement tests and explain their significance in educational settings.
2. Discuss the key characteristics that distinguish achievement tests from other types of assessments.
3. Describe the advantages and disadvantages of using achievement tests.
4. Identify and explain different types of achievement tests and their purposes.

****

UNIT-17

ACHIEVEMENT TEST CONSTRUCTION

Structure

17.1 INTRODUCTION

Dear learner,

Test construction is based upon practical and scientific rules that are applied before, during, and after each item until it finally becomes a part of the test. Construction of tests is an important part of assessing students' understanding of course content and their level of competency in applying what they are learning. In this lesson, we will learn about the procedure and principles of achievement test construction. We will also learn about item analysis and desirable attributes of a good achievement test.

17.2 LEARNING OBJECTIVES

After studying this unit, you will be able to:

- Discuss the steps of constructing an achievement test.
- Prepare a blueprint of an achievement test.
- Perform item analysis to evaluate test items.
- Write clear instructions for test administration and scoring.
- Identify desirable attributes of a good achievement test.

17.3 STEPS, PLANNING AND PREPARATION OF AN ACHIEVEMENT TEST

Any test designed to assess the achievement in any subject with regard to a set of predetermined objectives involves several major steps:

- **Planning and Preparation of a Design for the Test**

- Determine the objective of the test.
- Determine the maximum time and maximum marks.
- Test Length: The number of items that should constitute the final form of a test is determined by the purpose of the test or its proposed uses, and by the statistical characteristics of the items.

- **Preparation of a Design for the Test**

Important factors to be considered in design for the test are as follows.

- Weightage to objectives
- Weightage to content
- Weightage to Type of questions
- Weightage to difficulty level.

▪ Weightage to objectives

Here's the data presented in a table format:-

Sr. No.	Objectives	Marks	Percentage
1.	Knowledge	6	24
2.	Understanding	8	32
3.	Application	11	44
Total		25	100

▪ Weightage to content

Sr. NO.	Contents	Marks	Percentage
1.	**Sub topic**	10	40
2.	**Sub topic**	15	60
Total		25	100

▪ Weightage to Type of questions

Sr.no.	Types of questions	No. of questions	Marks	Percentage
1.	Objective type	13	13	52

2.	Short answer type	1	2	8
3.	Essay type	1	10	40
4.	Total	15	25	100

Weightage to difficulty level

Sr. No.	Forms of questions	Marks	Percentage
1.	Easy	5	20
2.	Average	15	60
3.	Difficult	5	20
Total		25	100

Self-Check Exercise-1

1. Which objective is given the highest percentage of marks?

a) Knowledge
b) Understanding
c) Application
d) Analysis

2. What percentage of the total marks is allocated to "Average" difficulty level questions?

a) 20%
b) 60%
c) 40%
d) 52%

17.4 PREPARATION OF THE TEST BLUEPRINT AND WRITING OF TEST ITEMS

Test Blueprint:

A test blueprint is a crucial tool in the test development process. It serves as a detailed plan or a three-dimensional chart that outlines the structure and content of an assessment. The blueprint ensures that the test is comprehensive and aligned with the learning objectives, content areas, and question formats intended to be covered. Here's a detailed breakdown of the process and components involved in preparing a test blueprint and writing test items:

Objectives Form of Questions Content	Knowledge			Understanding			Application			Grand Total
	O	SA	E	O	SA	E	O	SA	O	
Sub- Topic -1	1(3)			1(6)			1(1)			10
Sub- Topic-2	1(3)				2(1)			10(1)		
Total Marks	6	0	0	5	3	0	2		10	
Grand Total	6			8			11			25

Table of Specifications: A table of specifications is a two-way table that represents along one axis the content area/topics that the teacher has taught during the specified period and the cognitive level at which it is to be measured, along the other axis. In other words, the table of specifications highlights how much emphasis is to be given to each objective or topic. After preparation of test blue print, the items are written. The details regarding writing of test items are given in next lesson. Assembling and Arranging Items in the Test:

Items after having written and selected they are organized in the form of a test. Item of the same format may be placed together. Each item type requires specific set of directions and a somewhat different mental set on the part of the examinee. So far as possible, within item type, items dealing with the same content may be grouped together.

The examinee will be able to concentrate on a single domain at a time rather than having to shift back and forth among areas of content. Furthermore, the examiner will have an easier job of analysing the results, as it will be easier to see at a glance whether the errors are more frequent in one content area than the other. Items may be so arranged that difficulty progress from easy to hard Items should be arranged in the test booklet so that answers follow no set pattern.

Self-Check Exercise-2

1. What is the primary purpose of a test blueprint in the test development process?

a) To increase the difficulty of the test
b) To outline the structure and content of an assessment
c) To reduce the number of questions in a test

d) To provide financial incentives to teachers

2. In the test blueprint provided, how many marks are allocated to the "Knowledge" objective for "Sub-Topic 1"?

a) 3
b) 6
c) 5
d) 10

17.5 ASSEMBLING AND ARRANGING ITEMS IN THE TEST

Items after having written and selected they are organized in the form of a test. Items of the same format may be placed together. Each item type requires specific set of directions and a somewhat different mental set on the part of the examinee. So far as possible, within item type, items dealing with the same content may be grouped together. The examinee will be able to concentrate on a single domain at a time rather than having to shift back and forth among areas of content. Furthermore, the examiner will have an easier job of analyzing the results, as it will be easier to see at a glance whether the errors are more frequent in one content area than the other. Items may be so arranged that difficulty progress from easy to hard Items should be arranged in the test booklet so that answers follow no set pattern.

Self-Check Exercise-3

1. Items should be arranged in the test booklet so that difficulty progresses from _____ to _____, and answers follow no set pattern.

Answer: easy, hard

17.6 WRITING INSTRUCTIONS OR GUIDELINES FOR TEST ADMINISTRATION AND SCORING

The directions should be simple but complete. They should indicate the purpose of the test, the time limits and the score value of each question. Write a set of directions for each item type that is used on the test specifying what the respondent is expected to do and how one is required to record the responses. All pupils must be given a fair chance to demonstrate their achievement. Physical and psychological environment be conducive to their best efforts. Control all factors that might interfere with valid measurement: Adequate workspace, quiet, proper light and ventilation are important. Pupils must be put at ease, tension and anxiety should be reduced to the minimum. Separate answer sheets, which are easier to score, can be used at high school level and beyond. If the pupils' answers are recorded on the test paper, the teacher may make a scoring key by marking the correct answers on a blank copy of the test. When separate answer sheets are used, a scoring stencil is a blank answer sheet with holes punched where correct answer should appear. Before scoring procedure is used, each test paper should also

be scanned to make sure that only one answer was marked for each item. Any item containing more than one answer should be eliminated from scoring. In scoring objective tests, each correct answer is usually counted as one point. When pupils are told to answer every item on the test, a pupil's score is simply the number of items answered correctly. Short answer questions may sometime require awarding partial credit and may pose some problem in scoring. However, a detailed key may be prepared in advance to avoid confusion. For each question and for the test as a whole, the examiner may make a tally for each kind error that the examinees make. A summary of these errors could then be used to plan instructional activities.

Self-Check Exercise-4

1. The directions for a test should indicate the _____ of the test, the time limits, and the score value of each question.

2. When separate answer sheets are used, a scoring _____ is a blank answer sheet with holes punched where the correct answers should appear.

17.7 PERFORMING ITEM ANALYSIS

Often students judge, after taking the exam, whether the test was fair and good. Teacher is also usually interested about how the test worked for the students. One way to ascertain this is to undertake item analysis. It provides objective, external and empirical evidence for the quality of the items we have pre-tested. The objective of item analysis is to identify problematic or poor items which might be either confusing the respondents or do not have a clearly correct response or a distracter might well be competing with the keyed answer. Good test making requires careful attention to the principles of item evaluation. The basic methods involve are assessment of item difficulty and item discrimination. These measures comprise item analysis. Item analysis is about how difficult an item is and how well it can discriminate between the good and the poor students.

(i) Item Difficulty Index/Facility Index

Item difficulty is determined from the proportion (p) of students who answered each item correctly. Item difficulty can range from zero (none could solve it) to hundred (all persons solved it correctly). The goal is usually to have items of all difficulty levels in the test so that test could identify poor, average as well as good students. However, most of the items are designed to be average in difficulty levels for they are more useful. Item analysis exercise provides us the difficulty level of each item.

● Optimally difficult items are those that 50%-75% of students answer correctly.

● Items are considered low to moderately difficult if (p) is between 70% and 85%

● Items that only 30% or below solve correctly are considered difficult ones.

Item Difficulty Percentage can also be denoted as Item Difficulty Index by expressing it in decimals e.g. .40 for items which could be solved by 40 % of the test-takers. Thus index can range from 0 to 1. Items should fall in a variety of difficulty levels in order to differentiate between good and average as well as average and poor students. Easy items are usually placed in the initial part of the test to motivate students in taking the test and alleviating test-anxiety. The optimal item difficulty depends on the question type and number of possible distracters as well.

(ii) Item Discrimination

Another way to evaluate items is to ask “Who gets this item correct”- the good, average and the weak students? Assessment of item discrimination answers this query. Item discrimination refers to the percentage difference in correct responses between the poor and the high scoring students.

The discrimination index is a basic measure of the validity of an item. It is a measure of an items ability to discriminate between those who scored high on the total test and those who scored low. Though there are several steps in its calculation, once computed, this index can be interpreted as an indication of the extent to which overall knowledge of the content area or mastery of the skills is related to the response on an item. Perhaps the most crucial validity standard for a test item is that whether a student got an item correct or not is due to their level of knowledge or ability and not due to something else such as chance or test bias.

In a small class of 30 students, one can administer the test items, score them and then rank. Next, we separate the upper 15 students and the-low 15 into two groups: The UPPER and the LOW groups. Finally, we find how well each item was solved correctly (p) by each group. In other words, percentage of students passing (p) each item in each of the two groups is worked out. Discrimination (D) power of the item is then known by finding difference between the percentage of upper group and the low group. The higher the difference, the greater the discrimination power of an item.

D = (p of upper group - p of lower group)

In a large class of 100 or more students, we take the top 25% and the lower 25% students to form upper and lower groups, to cut short the labor or amount of work. The discrimination ratio for an item falls between -1.0 and +1.0. The closer the ratio is to +1.0, the more effectively that item distinguishes students who know the material (the top group) from those who don’t (the bottom group). An item with a discrimination of 60% or greater is considered a very good item, whereas a discrimination of less than 20%

indicates a low discrimination and the item needs to be revised. An item with a negative index of discrimination indicates that the poor students answer correctly more often than do the good students. Strange! Such items should be dropped from the test.

For example, ten students in a class have taken a ten items quiz. The students' responses are shown below from high to low. The top five students can be called the high score group and the bottom half as the low scoring group. The number "1" indicates a correct answer; a "0'' indicates an incorrect answer.

Student	Total score %	Items No.									
		1	2	3	4	5	6	7	8	9	10
1.	100	1	1	1	1	1	1	1	1	1	1
2.	90	1	1	1	1	1	1	1	1	0	1
3.	80	1	1	0	1	1	1	1	1	0	0
4.	70	0	1	1	1	1	1	0	1	0	1
5.	70	1	1	1	0	1	1	1	0	0	1
6.	60	1	1	1	0	1	1	0	1	0	0
7.	60	0	1	1	0	1	1	0	1	0	1
8.	50	0	1	1	1	0	0	1	0	1	0
9.	40	1	1	1	0	0	0	0	0	1	1
10.	30	0	1	0	0	0	1	0	0	1	0

Difficulty index and Discrimination Index are calculated below:

Items No.	Correct High Group	Low Correct Group	Difficulty %	Discrimination %
1	4	2	60	40
2	5	5	100	0
3	4	4	80	0
4	4	1	50	60
5	5	2	80	60
6	5	3	80	40
7	4	1	50	60
8	4	2	60	40
9	1	3	30	40
10	4	2	60	40

● Question no 2 was the easiest; no 9 was most difficult.

● Question 9 also had negative discrimination and should be removed from the test.

● 100% discrimination would occur if all those in the upper group answered correctly and all those in the lower group answered incorrectly.

● Zero discrimination occurs when equal numbers in both groups answer correctly.

● Negative discrimination, a highly undesirable condition, occurs when more students

in the lower group than the upper group answer correctly.

● Items with 25% and above discrimination are considered good.

(iii) Analysis of Response Options (Distracter Analysis):

In addition to examining the performance of an entire test item, teachers are often interested in examining the performance of individual distracters (incorrect answer options) on multiple-choice items. By calculating the proportion of students who chose each answer option, teachers can identify which distracters are “working” and appear attractive to students who do not know the correct answer, and which distracters are simply taking up space and not being chosen by many students. To eliminate blind guessing which results in a correct answer purely by chance (which hurts the validity of a test item), teachers want as many plausible distracters as is feasible. Analyses of response options allow teachers to fine tune and improve items they may wish to use again with future classes.

Interpreting Distracter Values:

Distracters should be ideally equally attractive, but not more than the answer. Minimum, it must be opted by at least 5% of the examinees. Weak or non-functional distracters may be substituted with new ones and make sure that they align with the stem as well as the objective of the item, well connected with the rest, and are grammatically correct.

Effectiveness of Distracters:

Difficulty and discrimination index are estimates about an item which overall comprises a stem and a set of distracters or options. The item analysis statistics reflects on the goodness of both distracters and the stem. Let us look at the guidelines which can help us improve them.

1. Most MCQs have 2-4 distracters; 3 is better, 4 is best at the college level Where it is difficult to think of more than one distracter, frame it as true/false item

2. Distracters that have less than 5 percent response rate are weak and may be changed / improved. Distracters which attracted no response are not working at all.

3. No distracter should be chosen more than the keyed response in the upper group.

4. Similarly, no one distracter should pull more than about half the students.

5. If students have respond about equally to all the options, they might be marking randomly or wildly guessing. Critically check contents of such items. They might have been written badly and the students seem to have no idea what you are asking. It could be very difficult items and students might be completely baffled.

6. If the low group gets the keyed answer as often as the upper group, all the distracters might be looked into again. Or drop the item if you have a large pool of items.

Self-Check Exercise-5

1. What is the primary objective of item analysis in test development?

    a) To increase the number of questions
    b) To identify problematic or poor items
    c) To reduce the test duration
    d) To improve student attendance

2. What should teachers do with distracters that attract less than 5% of the examinees?

    a) Keep them as they are.
    b) Substitute them with new ones.
    c) Increase their difficulty.
    d) Eliminate the correct answer.

17.8 SUMMARY

Test construction involves several key steps to ensure the reliability and validity of an achievement test. These steps include planning the test by determining objectives, time, and length; designing the test by assigning weightage to objectives, content, question types, and difficulty levels; preparing a test blueprint and writing test items; and arranging the test items logically. Additionally, clear instructions for test administration and scoring are essential. Item analysis, including evaluating item difficulty and discrimination, helps improve test quality by identifying problematic items and ensuring the test distinguishes between different levels of student performance. Analysis of response options also helps refine multiple-choice questions by examining the effectiveness of distracters.

17.9 GLOSSARY

Blueprint :- A test blueprint is a rubric, document, or a table that lists the learning outcomes to be tested , the level of complexity, and the weight for the learning outcomes in rubric.

Item Analysis: A process used to evaluate the effectiveness of individual test items by measuring their difficulty and discrimination indices.

Item Difficulty Index: A measure that indicates the proportion of students who answered a test item correctly, ranging from 0 to 1.

Item Discrimination: A measure of how well a test item differentiates between students who perform well overall and those who perform poorly.

17.10 ANSWER TO SELF-CHECK EXERCISE

Self-Check Exercise-1

1. c) Application

2. b) 60%

Self-Check Exercise-2

1. b) To outline the structure and content of an assessment

2. a) 3

Self-Check Exercise-3

1. easy, hard

Self-Check Exercise-4

1. purpose

2. stencil

Self-Check Exercise-5

1. b) To identify problematic or poor items

2. b) Substitute them with new ones.

17.11 REFERENCES/SUGGESTIVE READINGS

- Education Measurement and Evaluation : J.Swarupa Rani, Discovery Publishing house.
- Measurement and Evaluation in Teaching : Norman Edward Gronlund Macmillian
- Measurement and Assessment in Teaching : Robert L. Linn Pearson Education India.
- Program Evaluation and performance measurement : James C. Me. David, Laura

17.12 TERMINAL QUESTIONS

Dear learners, please check you progress by attempting the following questions:

1. Discuss the major steps involved in constructing an achievement test.
2. Explain the significance of preparing a test blueprint in the test construction process.
3. What is item analysis, and why is it important in test construction?
4. How does item discrimination contribute to the validity of a test item?

****

UNIT-18
CONSTRUCTION OF NORM-REFERENCED TEST

Structure

18.1 INTRODUCTION

Dear Learner,

Norm-referenced tests (NRTs) are designed to compare a student's performance to that of a group. These tests play a crucial role in educational assessment by identifying the relative standing of students. This unit will delve into the process of constructing NRTs, ensuring that participants gain a thorough understanding of each step involved.

18.2 LEARNING OBJECTIVES

After studying this unit, you will be able to:

- Explain the purpose and importance of norm-referenced tests.
- Outline the steps involved in constructing norm-referenced tests.
- Develop test items that align with curriculum content.
- Conduct and analyze item trials.
- Assemble and finalize a norm-referenced test

18.3 PURPOSE AND CHARACTERISTICS OF NORM-REFERENCED TESTS

The primary purpose of NRTs is to rank students and compare their performance to a norm group. This helps in:

- Identifying high and low achievers.
- Placing students in appropriate educational tracks or programs.
- Informing instruction and curriculum development.
- Evaluating the effectiveness of educational programs.

Characteristics of Norm-Referenced Tests

1. **Comparative Evaluation**: NRTs rank students based on their performance relative to a norm group.
2. **Standardization**: Tests are administered and scored in a consistent manner to ensure comparability.
3. **Statistical Analysis**: Data from NRTs are analyzed using statistical methods to establish norms and interpret scores.

Self-check Exercise-1

1. Which of the following is NOT a characteristic of Norm-Referenced Tests (NRTs)?

a) Comparative Evaluation
b) Standardization
c) Formative Assessment
d) Statistical Analysis

2. How do Norm-Referenced Tests (NRTs) assist in curriculum development?

a) By providing detailed individual feedback for each student
b) By ranking students and comparing their performance to identify high and low achievers
c) By offering personalized learning paths for each student
d) By assessing students' growth over the academic year

18.4 CONSTRUCTING NORM-REFERENCED TESTS

The steps for constructing norm-referenced tests are briefly discussed below:

1. Content analysis and test blueprints

A content analysis provides a summary of the intentions of the curriculum expressed in content terms. Which content is supposed to be covered in the curriculum? Are there significant sections of this content? Are there significant sub-divisions within any of the sections? Which of these content areas should a representative test include?

2. Item writing

Item writing is the preparation of assessment tasks which can reveal the knowledge and skill of students when their responses to these tasks are inspected. Tasks which confuse, which do not engage the students, or which offend, always obscure important evidence by either failing to gather appropriate information or by distracting the student from the intended task.

3. Item review

Writing assessment tasks for use in tests requires skill. Sometimes the item seems clear to the person who wrote it but may not necessarily be clear to others. Before empirical trial, assessment tasks need to be reviewed by a review panel (with a number of people) with questions like:

➢ Is the task clear in each item? Is it likely that the person attempting an item will know what is expected?

- Are the items expressed in the simplest possible language?
- Is each item a fair item for assessment at this level of education?
- Is the wording appropriate to the level of education where the item will be used?
- Are there unintended clues to the correct answer?
- Is the format reasonably consistent so that students know what is required from item to item?
- Is there a single clearly correct (or best) answer for each item?
- Is the type of item appropriate to the information required?
- Are there statements in the items which are likely to offend?
- Is there content which reflects bias on gender, racial, or other grounds?
- Are the items representative of the behaviours to be assessed?
- Are there enough representative items to provide an adequate sample of the behaviours to be assessed?

This review before the items are tried should ensure that we avoid tasks which are expressed in language too complex for the idea being tested, avoid redundant words, multiple negatives, and distracters which are not plausible. The review should also identify items with no correct (or best) answer and items with multiple correct answers. Such items may be discarded or re-written.

4. Trial of the Items

Item trial is sometimes called pilot testing - but in this context it does not mean testing those who fly aeroplane. As well as considering the best efforts of item writers and item reviewers as a means of eliminating faulty items and improving the quality of items, it is necessary to subject the proposed items to empirical trial with students similar to those who are going to use the final form of the test. It is usual to allocate the trial forms on a random basis within each trial examination room so that (on the average) each trial test is attempted by candidates of comparable ability. The same form of a test should not be given to candidates sitting in adjacent seats so as to ensure that candidates do not improve their scores by looking at another candidate's paper.

5. Processing Test Responses after Trial Testing

If the test needs to be scored before analysis; this scoring is done next. If there are essays to be scored, it is good practice to mark the first essay all the way through the stack of test papers. Then start the stack again to score the next essay. When all items have been marked, the scores on each item are entered into a computer file. If the test is multiple-choice in format, the responses may be entered into a computer file directly.

6. Item Analysis

Empirical trial can identify instances of confused meaning, alternative explanations not already considered by the test constructors, and (for multiple-choice questions) options which are popular amongst those lacking knowledge, and incorrect options which are chosen for some reason by very able students. The item analysis also provides an opportunity to collect

information about how each item performs relative to other items in the same test, and to judge the consistency of the whole test.

7. Amending the Test by Discarding/Revising/Replacing items

Items which do not perform as expected can be discarded or revised. However, discarding questions when there is a shortage of replacement questions can lead to distortions of the achieved test specification. If the original specification represents the best sampling of content, skills, and item formats, in the judgments of those preparing and reviewing the test, then leaving some cells of the grid vacant will indicate a less than adequate test To avoid this possibility, test constructors may prepare three or four times as many questions that they think they will need for each cell in the grid.

8. Assembling the Final Test (or a further trial test) and the Corresponding Score Key

After trial, tasks may be re-ordered to take account of their difficulty. Usually the easiest questions are presented first. This is to encourage candidates to proceed through the test and to ensure that the weaker candidates do not become discouraged before providing adequate evidence of their achievements and skills. Minor changes to items may have to be made for layout reasons (for example, to keep all of an item on one page of the test, or to avoid obvious patterns in the list of correct answers).

9. Other Practical Concerns in Preparing the Test

✓ How much time will students have to do the actual test? What time will be set aside to give instructions to those students attempting the test? Will the final number of items be too large for the test to be given in a single session? Will there be a break between testing sessions when there is more than one session?

✓ Will the students be told how the items are to be scored? Will they be told the relative importance of each item? Will they be given advice on how to do their best on the test?

✓ What test administration information will be given to those who are giving the trial test to students? Will the students be told that the results will be returned to them? Are the tests to be treated as secure tests (with no copies left behind in the venue where the test is administered)?

✓ Do students need advice on how they are to record their responses? If practice items are to be used for this purpose, what types of response should they cover? How many practice items will be necessary?

✓ Will the answers be recorded on a separate answer sheet (perhaps so that a test booklet can be used again)? Will this use of a separate sheet add to the time given for the trial test? What information should be requested in addition to the actual responses to the items? (This might include student name, school, year level, sex, age, etc.)

✓ Has the layout of the test (and answer sheet if appropriate) been arranged for
efficient scoring of responses? Are distracters for multiple-choice tests shown as
capital letters (easier to score than lower case letters)?

✓ Have the options in multiple-choice items been arranged in some logical order (for example, from smallest to largest)? Have the items been placed in order from easiest to most difficult (to encourage candidates to continue

through the test)? Has the layout of items avoided patterns in the correct answers such as 3 or more of the same letter in a row, or other patterns like ABCD or ABABAB (which might lead to 'correct' responses for the 'wrong' reasons)?

Developing Norms for Interpretation of Test Scores

Norm is average score of sample population. These are the level obtained by a particular group of persons on a test. There are many types of norms like age norms, grade norms, percentile norms and standard scores.

Self-check Exercise-2

1. Why is it important to conduct an empirical trial of test items?

a) To ensure the test is administered consistently
b) To identify instances of confused meaning and alternative explanations
c) To summarize the curriculum content
d) To prepare the final test score key

2. In preparing the final test, why are the easiest questions usually presented first?

a) To prevent students from becoming discouraged early in the test
b) To ensure that the test is scored efficiently
c) To avoid patterns in the list of correct answers
d) To cover the most important content areas first

18.5 SUMMARY

Norm-referenced tests (NRTs) compare a student's performance against a predefined group, highlighting their relative standing. These tests follow a systematic process: conducting content analysis, item writing, and review; piloting items; scoring and analyzing responses; and refining the test based on item performance. Norms, such as age or grade norms, are established using data from a representative sample to interpret scores. While NRTs are useful for ranking students and informing educational decisions, they can encourage teaching to the test and may not reflect individual progress. Ensuring fairness, validity, and security is crucial in constructing effective NRTs.

18.6 GLOSSARY

- **Norm-Referenced Test (NRT):** A test designed to compare a student's performance to that of a group.
- **Content Analysis:** The process of summarizing curriculum intentions in content terms.

- **Norms:** Average scores of a sample population, used for interpreting test scores.

18.7 ANSWERS TO SELF-CHECK EXERCISE

Self-check Exercise-1

1. c) Formative Assessment

2. b) By ranking students and comparing their performance to identify high and low achievers

Self-check Exercise-2

1. b) To identify instances of confused meaning and alternative explanations

2. a) To prevent students from becoming discouraged early in the test

18.8 REFERENCES/SUGGESTIVE READINGS

- Aggarwal, J. C. (2019). Essentials of Educational Evaluation. Vikas Publishing House.
- Sachdeva, K. (2014). Educational Measurement and Evaluation. Sterling Publishers Pvt. Ltd.
- Mishra, P. (2017). Educational Measurement and Evaluation. Himalaya Publishing House

18.9 TERMINAL QUESTIONS

Dear learners, please check you progress by attempting the following questions:

1. Explain the process of constructing a norm-referenced test from content analysis to final test assembly.
2. Describe the role of item analysis in the development of norm-referenced tests. What statistical methods are used, and how do they contribute to the overall quality of the test?

****

UNIT-19

WRITING TEST ITEMS-1

Structure

19.1 INTRODUCTION

Dear Learner,

Educational assessments play a pivotal role in evaluating students' understanding and mastery of learning objectives. At the heart of any assessment are the test items—questions or prompts designed to gauge students' knowledge, skills, and abilities. Effective test item construction is not merely about drafting questions but ensuring that these questions are valid, reliable, and aligned with educational goals.

This chapter delves into the principles and practices of constructing test items across various formats, such as multiple-choice questions, true-false statements, and short-answer items. Each type of item serves a distinct purpose in assessing different cognitive levels—from recalling facts to analyzing complex scenarios. By understanding the nuances of item construction, educators can design assessments that accurately measure student achievement and inform instructional decisions.

19.2 LEARNING OBJECTIVES

After studying this unit, you will be able to:

- Understand the principles of constructing different types of test items.
- Identify strategies to enhance the validity and reliability of test items.

- Demonstrate the ability to develop test items aligned with learning objectives.
- Evaluate the appropriateness of test items for assessing different cognitive levels.

19.3 WRITING TEST ITEMS

The next step after planning the test is preparing it in accordance with the plan. This step mainly deals with development of items and organizing them in the form of a test. The initially developed test draft or pool of items is termed as preliminary draft or rough draft of the test. Different types of questions can be devised for an achievement test, for instance, multiple choice, fill-in-the-blank, true-false, matching, short answer and essay.

Although each type of question is constructed differently, the following principles apply to constructing questions and tests in general:

1. Instructions for each type of question must be simple and brief.
2. Questions must be written in simple language. If the language is difficult or ambiguous, even a student with strong language skills and good vocabulary may answer incorrectly if his/her interpretation of the question is different from the author's intended meaning.
3. Test items must assess specific ability or comprehension of content developed during the course of study.
4. Write the questions as you teach or even before you teach, so that your teaching may be aimed at significant learning outcomes.
5. Devise questions that call for comprehension and application of knowledge skills.
6. Some of the questions must aim at appraisal of examinees' ability to analyze, synthesize, and evaluate novel instances of the concepts. If the instances are the same as used in instruction, students are only being asked to recall (knowledge level),
7. Questions should be written in different formats, e.g., multiple-choice, completion, true-false, short answer etc. to maintain interest and motivation of the students.
8. Prepare alternate forms of the test to deter cheating and to provide for make-up testing (if needed).
9. The items should be phrased so that the content rather than the format of the statements will determine the answer. Sometimes the item contains "specific determiners" which provide an irrelevant cue to the correct answer. For example, statements that contain terms like always, never, entirely, absolutely, and exclusively are much more likely to be false than to be true. On the other hand, such terms as may, sometimes, as a rule, and in general are much more likely to be true. Besides, care should be taken to avoid double negatives, complicated sentence structures, and unusual words.
10. The difficulty level of the items should be appropriate for the ability level of the group. Optimal difficulty for true-false items is about 75 percent, for five-option multiple choice questions about 60 percent, and for completion items approximately 50 percent. However, difficulty in itself is not an end, the item content should be determined by the importance of the subject matter. It is desirable to place a few easy items in the beginning to motivate students, particularly those who are of below average ability.

11. The items should be devised in such a manner that different taxonomy levels are evaluated. Besides, achievement tests should be power test, not speed test.
12. Items pertaining to a specific topic or of a particular type should be placed together in the test. Such a grouping facilitates scoring and evaluation. It will also be helpful for the examinees to think and answer the items, similar in content and format, in a better manner without fluctuation of attention and changing the mind-set.
13. Directions to the examinees should be as simple, clear, and precise as possible, so that even those students who are of below average ability can clearly understand what they are expected to do.
14. Scoring procedures must be clearly defined before the test is administered.
15. The test constructor, must clearly state optimal testing conditions for test administration.
16. Item analysis should be carried out to make necessary changes, if any ambiguity is found in the items.

Before we discuss preparing the test, it seems quite reasonable that we talk about different types of test items, their characteristics, use and limitations. Items commonly used for Tests of Achievement Two major types of items have been identified:
1. Constructed Response / Supply items
2. Structured Response / Select items

1. Constructed Response / Supply items: In the supply type items the question is so framed that the examinee has to supply or construct the answer on his own in his own words. They generally include the following type: Essay type, Short answer type, Completion type items.

2. Structured Response / Select items: In the select type items, as the name suggests the examinee is required to select the correct answer from amongst the given or structured options. They are often called objective items. They include: Alternate Response type, Multiple-choice type, Matching type.

Self-check Exercise-1

1. Which principle is NOT generally applicable to constructing questions and tests?
   a) Instructions for each type of question must be simple and brief
   b) Questions should require complex sentence structures to challenge students
   c) Test items must assess specific ability or comprehension of content developed during the course of study
   d) Write questions that call for comprehension and application of knowledge skills

2. Which principle is important for determining the difficulty level of test items?
   a) Test items should be extremely difficult to challenge students

b) The difficulty level should be appropriate for the ability level of the group
c) Items should always be at an optimal difficulty of 75% for all types
d) Difficulty should be determined without considering the importance of the subject matter

19.4 CONSTRUCTING OBJECTIVE TYPE TEST ITEMS

Construction of test items is a crucial step for the validity of a classroom test is determined by the extent to which performance to be measured is called forth by the test items. It is not enough to have knowledge of subject matter, defined learning outcomes, or a psychological understanding of the students' mental processes, although all of these are prerequisites. The ability to construct high-quality test items requires knowledge of the principles and techniques of test construction and skill in their application. Objective test forms typically measure relatively simple learning outcome.

Self-check Exercise-2

1. Constructing high-quality test items requires knowledge of the principles and techniques of test construction and _____ in their application.

19.5 ALTERNATIVE RESPONSE TYPE ITEMS

Alternative response item is the one that offers two options to choose from. They often consist of a declarative statement that the examinee is asked to mark true or false, right or wrong, correct or incorrect, yes or no, agree or disagree, or the like. Incomplete sentences providing two options to choose from to fill in the blank also fall in this category. The most common form it takes is True - False questions. Most common use of the true- false item is in measuring the examinee's ability to identify the correctness of statements of fact, definitions of terms, statements of principles, and the like, also to distinguish fact from opinion. Another aspect of understanding that can be measured by the true-false item is the ability to recognize cause-and-effect relationships. This type of item usually contains two true propositions in one statement, and the examinee is to judge whether the relationship between them is true or false. The true-false item also can be used to measure some simple aspects of logic. A common criticism of the true-false item is that an examinee may be able to recognize a false statement as incorrect but still not know what is correct.

Suggestions for Constructing True-False Items:

● Avoid trivial statements.
● Avoid broad general statements.
● Avoid the use of negative statements, especially double negatives.
● when a negative word must be used, it should be underlined or put in italics so that students do not overtook it.
● Avoid complex sentences. Avoid including two ideas in one statement, unless cause-effect relationships are being measured.
● Avoid using opinion that is not attributed to some sources, unless the ability to identify opinion is being specifically measured.
● Avoid using true statements and false statements that are unequal in length.
● Avoid using disproportionate numbers of true statements and false statements.

Self-check Exercise-3

1. Alternative response items often consist of a declarative statement that the examinee is asked to mark _____ or _____.
2.When constructing true-false items, it is recommended to avoid the use of negative statements, especially _____.

19.6 SHORT ANSWER/ COMPLETION TYPE ITEMS

The short answer item and the completion item both are supply-type test items. Yet, they are included here for their simplicity. They can be answered by a word, phrase, number, or symbol. The short-answer item uses a direct question whereas the completion item consists of an incomplete statement. Short-answer item is especially useful for measuring problem-solving ability in science and mathematics. Complex interpretations can be made When the short- answer item is used to measure the ability to interpret diagrams, charts, graphs, and pictorial data. When short-answer items are used the question must be stated clearly and concisely. It should be free from irrelevant clues, and require an answer that is both brief and definite.

Suggestions for Constructing Short Answer Items

● Word the item so that the required answer is both brief and specific. A direct question is generally more desirable than an incomplete statement.
● Do not take statements directly from textbooks to use as a basis for short-answer items.
● If the answer is to be expressed in numerical units, indicate the type of answer wanted.
● Blanks for answers should be equal in length and in a column to the right of the question.

● Do not include too many blanks.

Self-Check Exercise-4

1.The short-answer item uses a direct question, whereas the completion item consists of an _____.

Question 3:

2. The short-answer item is especially useful for measuring problem-solving ability in _____ and _____.

19.7 SUMMARY

This chapter emphasizes the importance of constructing effective test items to ensure educational assessments are valid, reliable, and aligned with learning objectives. Various types of test items, including true-false statements, and short-answer items are discussed along with their specific purposes in assessing different cognitive levels. The principles of test item construction, such as clarity, simplicity, alignment with learning outcomes, and consideration of students' abilities, are highlighted. The chapter also provides specific guidelines for writing true-false and short-answer items, ensuring they accurately measure students' knowledge and skills.

19.8 GLOSSARY

- **Educational Assessment**: A systematic process of documenting and using empirical data on the knowledge, skill, attitudes, and beliefs to refine programs and improve student learning.
- **Test Item**: A question or prompt designed to assess students' understanding, skills, and abilities in a given subject area.
- **True-False Statements**: Test items that offer two options (true or false) for the examinee to choose from.
- **Short-Answer Items**: Test items that require the examinee to provide a brief, specific response.

19.9 ANSWERS TO SELF-CHECK EXERCISE

Self-Check Exercise-1

1. b) Questions should require complex sentence structures to challenge students

2. b) The difficulty level should be appropriate for the ability level of the group

Self-Check Exercise-2

1. skill

Self-Check Exercise-3

1.true, false

2. double negatives

Self-Check Exercise-4

1. incomplete statement

2. science, mathematics

19.10 REFERENCES/SUGGESTIVE READINGS

- Nitko, A. J., & Brookhart, S. M. (2011). Educational Assessment of Students (6th ed.). Pearson.
- Palomba, C. A., & Banta, T. W. (1999). Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education. Jossey-Bass.
- Lane, S., Raymond, M. R., & Haladyna, T. M. (Eds.). (2015). Handbook of Test Development (2nd ed.). Routledge.

- Haladyna, T. M. (2004). Developing and Validating Test Items (3rd ed.). Routledge.

19.11 TERMINAL QUESTIONS

Dear learners, please check you progress by attempting the following questions:

1. **What are the key principles to consider when constructing test items to ensure they are effective in assessing students' knowledge?**
2. **Why is it important to include questions that assess higher-order thinking skills in a test?**
3. What are the guidelines for writing short-answer items to ensure they are clear and concise?

****

UNIT- 20
WRITING TEST ITEMS -2

Structure

20.1 INTRODUCTION

Dear learner,

In educational assessments, the design and construction of test items are crucial for accurately measuring students' understanding and abilities. Two prevalent types of test items are multiple-choice questions (MCQs) and essay-type questions. Each type has its unique structure, advantages, and challenges in construction, and serves distinct purposes in evaluating different cognitive levels. Multiple-choice questions are widely used due to their efficiency in assessing a broad range of content and their ease of scoring. They consist of a stem that presents a problem and a set of options, including one correct answer and several distractors. Essay-type questions, on the other hand, are designed to assess higher-order thinking skills and the ability to organize and express thoughts coherently. This unit focuses on the principles and guidelines for constructing multiple- choice and essay-type questions. By understanding the intricacies of these item types, educators can create assessments that are both valid and reliable, providing meaningful insights into student learning.

20.2 LEARNING OBJECTIVES

After studying this unit, you will be able to:

- Understand the structure and components of multiple-choice questions, including the stem and distractors.

- Identify the principles of constructing effective multiple-choice questions to minimize guessing and maximize validity.
- Develop multiple-choice questions that are clear, concise, and aligned with learning objective.
- Understand the purpose and advantages of essay-type questions in assessing higher-order thinking skills.
- Identify the key elements of well-constructed essay prompts, including clarity and specificity.
- Develop essay questions that effectively evaluate students' ability to organize and express their thoughts.

20.3 MULTIPLE CHOICE QUESTIONS, THEIR ADVANTAGES AND DISADVANTAGES, GUIDELINES FOR CONSTRUCTING MULTIPLE-CHOICE ITEMS

The multiple-choice item (MCQ) consists of two distinct parts: The first part that contains task or problem is called stem of the item. The stem of the item may be presented either as a question or as an incomplete statement. The form makes no difference as long as it presents a dear and a specific problem to the examinee. Second part presents a series of options or alternatives. Each option represents possible answer to the question. In a standard form one option is the correct or the best answer called the keyed response and the others are mis-leads or foils called distracters. The number of options used differs from one test to the other. An item must have at least three answer choices to be classified as a multiple-choice item. The typical pattern is to have four or five choices to reduce the probability of guessing the answer. A good item should have all the presented options look like probable answers at least to those examinees who do not know the answer. The multiple-choice items, despite having advantages over other items, have some serious limitations as well. It takes time to construct MCQ. They are susceptible to guessing and do not provide any diagnostic information. Multiple-choice items (MCI): is usually divided into three groups or parts:

1- The Stem
2- The correct choice or correct answer
3- The distracters.

The Stem, the initial part of each multiple - choice items is known as the stem. It can be complete statement, an incomplete statement and question. The answer can be a word or a group of words. The distracters can be two or three or four options. They are the options which surrounded the answer so that the students with inadequate knowledge cannot find the answer. Objective items require students to select the correct response from several alternatives or supply a word or short phrase to answer a question. It includes multiple - choice, True / false, matching and completion items.

Advantages of Multiple-Choice Questions (MCQs)

1. **Broad Content Coverage**: MCQs allow for a wide sampling of subject content, enabling educators to assess a broad range of knowledge within a limited testing period.

2. **Objective Scoring**: Scoring of MCQs is straightforward and objective, reducing potential biases and ensuring consistency.
3. **Efficiency**: MCQs can be scored quickly, especially with automated systems, making them suitable for large classes.
4. **Diagnostic Information**: Well-constructed MCQs can provide diagnostic information about students' understanding of specific concepts.
5. **Versatility**: MCQs can assess various cognitive levels, from basic recall of facts to higher-order thinking skills like analysis and application.

Disadvantages of Multiple-Choice Questions (MCQs)

1. **Time-Consuming Construction**: Developing high-quality MCQs is time-consuming and requires careful crafting to ensure clarity and avoid ambiguity.
2. **Guessing**: Students may guess the correct answer, which can undermine the validity of the assessment if not properly mitigated with plausible distractors.
3. **Limited Depth of Understanding**: MCQs may not adequately assess complex thinking or the ability to synthesize and evaluate information.
4. **Susceptibility to Test-Wiseness**: Students with test-taking skills (test-wise students) may perform better than their actual knowledge warrants.
5. **Lack of Diagnostic Depth**: While MCQs can identify what students know, they may not provide insights into how or why students think as they do, limiting the depth of diagnostic feedback.

Guidelines for Constructing Multiple-Choice Items

- Be sure that the stem clearly formulates a problem. The stem should be worded so that the examinee clearly understands the question being asked before he reads the answer choices.
- Stem should be written either in direct question form or in an incomplete statement form.
- The stem of the item should present only one problem. Two concepts must not be combined together to form a single stem.
- Include as much of the item in the stem and keep options as short as possible. This leads to economy of space, economy of reading time and clear statement of the problem.
- Unnecessary words or phrases should not be included in the stem. Such words add to the length and complexity of the stem but do not enhance meaningfulness of the stem. The stem should be written in simple, concise and clear form.
- Avoid the use of negative words in the stem of the item. There are times when it is important for the examinee to detect errors or to know exceptions. For these purposes, sometimes the use of ‘not’ or ‘except’ is justified in the stem. When a negative word is used in a stem it should be highlighted.

- Use novel material in formulating problems to measure understanding or ability to apply principles. Do not focus too closely on rote memory of the text that neglects measurement of the ability to use information.
- Use plausible distracters as alternatives. If an examinee who does not know the correct answer is not distracted by a given alternative, that alternative is not plausible and it will add nothing to the functioning of the item.
- Be sure that no unintentional clues
- The correct answer should appear at each position in almost equal numbers. While constructing multiple-choice item, some examiners have a tendency to place correct alternative at the first position. Some place it in the middle and others at the end. Such tendencies should be consciously controlled.
- Avoid using 'none of the above'. 'all of the above'. both a and b etc. as options for an MCQ.
- Alternatives should be grammatically consistent with the stem. Grammatical
  Inconsistency provides irrelevant clues.

Self-check exercise 1

1. Which of the following describes the stem in a multiple-choice question?

a) The correct answer
b) The question or problem
c) The distractors
d) The scoring guidelines

2. What is the purpose of including distractors in a multiple-choice question?

a) To confuse students
b) To increase the difficulty level
c) To ensure the correct answer is easily identifiable
d) To challenge students' understanding and knowledge

20.4 MATCHING TYPE QUESTIONS AND SUGGESTIONS FOR CONSTRUCTING MATCHING TYPE QUESTIONS

Matching type questions consists of two parallel columns with each word, number, or symbol in one column being matched to a word, sentences, or phrase in the other column. Items in the column for which a match is sought are called premises, and the items in the column from which the selection is made are called responses.

Suggestions for Constructing Matching Type Questions

- Use only homogeneous material in a single matching exercise.
- Include an unequal number of responses and premises and instruct the student that responses may be used once, more than once, or not at all.
- Keep the list of items to be matched brief, and place the shorter responses on the right.

- Arrange the list of responses in logical order. Place words in alphabetical order and numbers in sequence.
- Indicate in the directions the basis for matching the responses and premises.
- Ambiguity and confusion will be avoided. And testing time will be saved.
- Place all of the items for one matching exercise on they same page.
- **Self- check Exercise-2**

1. What is the main characteristic of matching type questions?
    a) They require short answers
    b) They involve selecting from columns
    c) They assess higher-order thinking
    d) They are subjective in nature

2. What is a potential benefit of using matching type questions?
    a) They allow for creative responses
    b) They measure deep understanding
    c) They are quick to grade
    d) They are easy to construct

20.5 ESSAY TYPE QUESTIONS AND GUIDELINES FOR CONSTRUCTING ESSAY TYPE QUESTIONS

There are two major purposes for using essay questions that address different learning outcomes. One purpose is to assess students understanding of subject-matter content. The other purpose is to assess students writing abilities: These two purposes are so different in nature that it is best to treat them separately. An essay question is “a test item which requires a response composed by the examinee, usually in the form of one or more sentences, of a nature that no single response or pattern of responses can be listed as correct, and the accuracy and quality of which can be judged subjectively only by one skilled or informed in the subject.”

An essay question should meet the following criteria:

1. Requires examinees to compose rather than select their response. Multiple-choice questions, matching exercises, and true-false items are all examples of selected response test items because they require students to select an answer from a list of possibilities provided by the test maker, whereas essay questions require students to construct their own answer.
2. Elicits student responses that must consist of one or more sentences.
3. No single response or single response pattern is correct.
4. The accuracy and quality of students’ responses to essays must be judged be subjectively by a competent specialist in the subject.

Guidelines for Constructing Essay Questions:

- Clearly define the intended learning outcome to be assessed by the item.
- Avoid using essay questions for intended learning outcomes that are better
  assessed with other kinds of assessment.
- Define the task and shape the problem situation.

- Helpful Instructions: Specify the relative point value and the approximate time limit in clear directions.
- Helpful Guidance: State the criteria for grading
- Use several relatively short essay questions rather than one long one.
- Avoid the use of optional questions. Students should not be permitted to choose one essay question to answer from two or more optional questions. The use of optional questions should be avoided for the following reasons. Students may waste time deciding on an option. Some questions are likely to be harder which could make the comparative assessment of student abilities unfair.

Self- check Exercise-3

1. **What distinguishes essay type questions from other types of questions?**

    a) They require short answers
    b) They are subjective in nature
    c) They have a single correct response
    d) They measure specific knowledge

2. **Which guideline is important when constructing essay type questions?**

    a) Include multiple questions within one essay
    b) Keep the questions vague to encourage creativity
    c) Provide a clear grading rubric
    d) Limit the word count for each response

20.6 SUMMARY

MCQs are structured with a stem presenting a problem or question and multiple options, one of which is correct and the rest are distractors. However, constructing effective MCQs can be time-consuming, they are susceptible to guessing, and may not fully assess deep understanding or higher-order thinking skills. Essay questions require students to construct detailed responses in their own words, demonstrating understanding and application of knowledge. They are valuable for assessing complex learning outcomes, encouraging critical thinking, and evaluating higher-order cognitive skills such as analysis, synthesis, and evaluation. While MCQs are efficient for assessing basic knowledge and can cover a wide range of content, essay questions provide deeper insights into students' comprehension and analytical abilities.

20.7 GLOSSARY

- **Stem:** -The main part of a multiple-choice question that presents the problem or task to the examinee.
- **Distractors:**- Incorrect answer choices in a multiple-choice question that are designed to challenge the examinee's knowledge and understanding.

20.8 ANSWERS TO SELF-CHECK EXERCISES

Self-check Exercise-1

1.b) The question or problem

2. d) To challenge students' understanding and knowledge

Self-check Exercise-2

1.b) They involve selecting from columns
2.c) They are quick to grade

Self-check Exercise-3

1. b) They are subjective in nature

2. c) Provide a clear grading rubric

20.9 REFERENCES/SUGGESTIVE READINGS

- Ebel, Robert L.(1966) “Measuring Educational Achievement, Prentice Hall of India Pvt. Ltd.
- Gronlund, N. E. (1976), Measurement and Evaluation in Teaching. McMillan, USA.
- Hopkins, C.D. and Antes, R.L. (1990). Classroom measurement and evaluation. Itasca,Illinois: Peacock.
- Mehrens, W.A. and Lehmann, I.J. (1984). Measurement and evaluation in education and psychology.(3rd Ed.) New York: Holt, Rinehart and Winston. Nandra, I.D.S.(2011). Learning Resources and Assessment of Learning.Patiala, 21st Century Publications.

20.10 TERMINAL QUESTIONS

Dear learners, please check you progress by attempting the following questions:

1. Describe the structure of a multiple-choice item (MCQ) and explain the function of each part.

2. Discuss the advantages and disadvantages of using multiple-choice questions (MCQs) in assessments. Provide specific examples to support your points.

3. What are some key guidelines for constructing effective multiple-choice items? Explain why each guideline is important for the validity of the assessment.

4. Explain the characteristics and suggestions for constructing matching type questions.

5. What are the purposes of using essay type questions in assessments? Discuss the guidelines for constructing effective essay questions.

****