Assessment for Learning

Unit – III

TOOLS AND TECHNIQUES FOR CLASSROOM ASSESSMENT

Assessment techniques

- Its purpose is to improve the quality of student learning
- It provides faculty with feedback about their effectiveness as teachers
- It gives students a measure of their progress as learners
- To provide faculty with information on what, how much and how well students are learning

Observation / Extro spection / Uncontrolled / Naturalistic / objective Observation Method

- Latin word
- Observare – to keep open
- Measurement without using any instrument or device
- Understanding the environment by sense organs

Observation

- External behaviour of persons in appropriate situations
- Controlled or uncontrolled
- Expert, purposive, systematic, carefully focused and thoroughly recorded
- Should be accurate, valid and reliable
- Tools such as check list and score-card, tape-recorder, thermometers, audiometer, stop-watch, binoculars etc.

Types

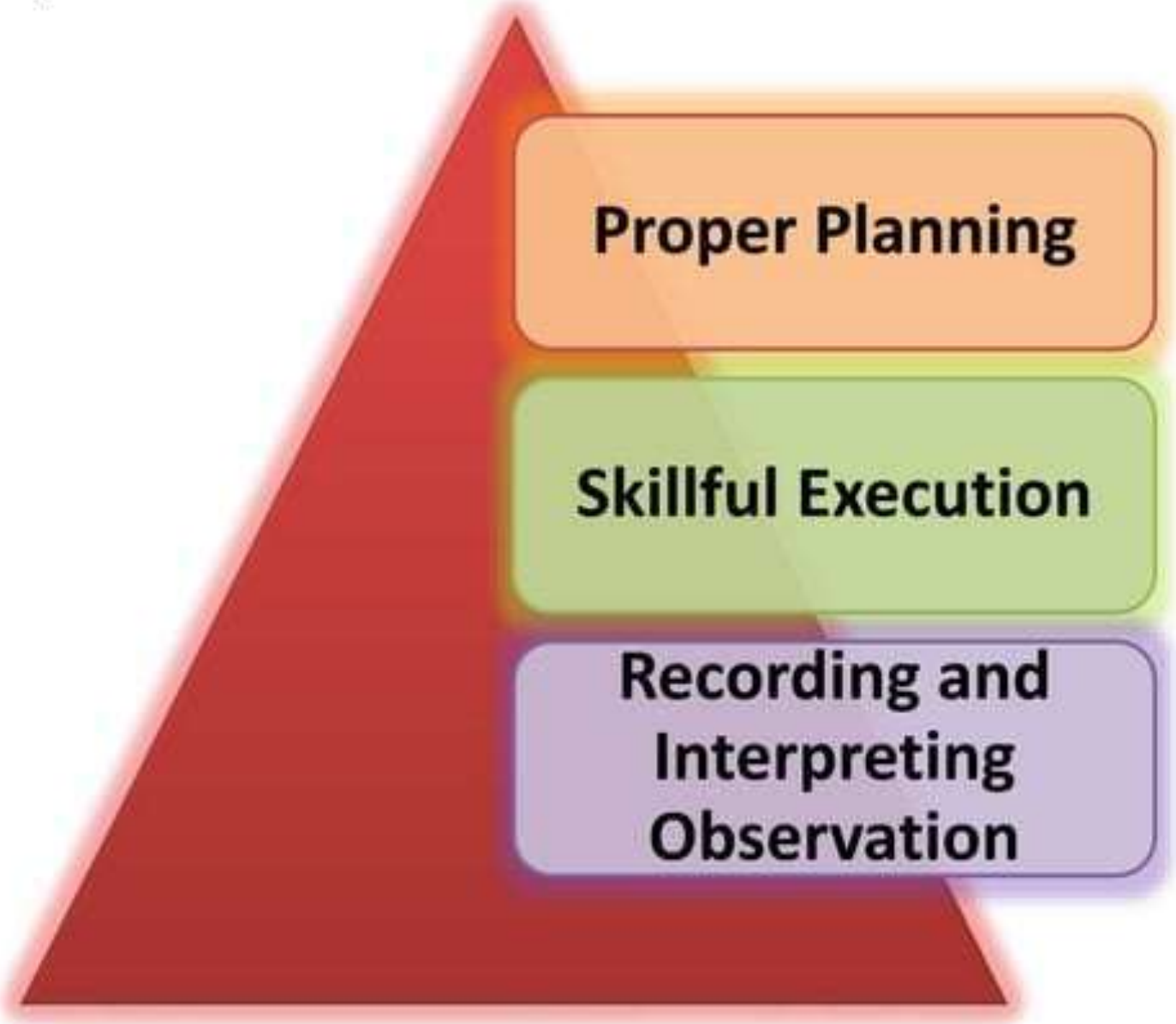
- Direct Observation
- Indirect Observation
- Natural Observation
- Artificial Observation
- Participant Observation
- Non-participant Observation
- Controlled Observation - laboratory
- Uncontrolled Observation - natural situation
- Known Observation
- Unknown Observation
- Group Observation
- Personal Observation

Types of Observation

Uses of Observation

- In descriptive research
- Significant aspects of personality which express themselves in behaviour
- Physical aspects of school buildings or students and teachers – through physical examination, measurement, assessment and comparison with fixed standards
- In classroom – learning behaviour
- Cumulative record – anecdotal evidence – research studies

Requisites of a Good Observation



Self reporting

- Type of survey, questionnaire, or poll
- Respondents read the question and select a response **by themselves** without researcher interference
- Any method which involves **asking a participant about their feelings, attitudes, and beliefs and so on**
- Examples - Questionnaires and interviews
- Used in **observational studies and experiments**

❖ Questionnaires

- Consists of a set of questions usually in a highly structured written form
- Open questions and closed questions

❖ Interviews

- **Spoken** questionnaire where the interviewer **records** the responses
- **Structured** - predetermined set of questions
- **Unstructured** - no questions are decided in advance

- Questionnaires and interviews - **to study large samples** of people fairly easy and quickly.
- Ask people **to reveal behaviour and feelings** which have been experienced in real situations.
- Participants may not respond **truthfully**
- **Social desirability** bias can be a big problem
- If questionnaires are sent out, say via **email** or through tutor groups, response rate can be **very low**

- Unstructured interviews can be **very time consuming** and difficult to carry out
- Structured interviews can **restrict the respondents' replies**
- **Psychologists** often carry out **semi-structured interviews** which consist of some pre-determined questions and followed up with further questions which allow the respondent to develop their answers

Anecdotal Record

- Anecdote – short account of some happening
- Anecdotal records are factual descriptions of the meaning incidents and events which the teacher has observe in the lives of his students
- An anecdotal record is a report of a significant episode in the life of a student

- Raths Louis

❖Meaning

- An anecdotal record is a **running description** of actual examples of **behavior of a student** is observed by teachers and the counselor.

❖Definition

- "Anecdotes are descriptive accounts of episodes or occurrences in a daily life of the student."

- Brown and Martin

A Specimen of an Anecdotal Record

- Name of the School
- Name of pupil observed _____ Class _____ Subject _____
- Observer Date and place
- _____
- Objective Description
- _____
- Comment of Observer

Values and Uses of Anecdotal Records

- They provide **specific and exact description** of personality and minimize generalizations.

- They are very helpful in **understanding the child's behavior** in diverse situations.

- They provide a **continuous record**.
- They provide data for pupils to **use in self-appraisal**
- A summary of these **records is valuable for forwarding** with a pupil when he is transferred from one school to another.
- The **new members of the staff** may use these records and acquaint themselves with the students.

Limitations

- Anecdotal record tends to be **less reliable** than other tools
- They are **time consuming** to write

Check List

- Consisting of prepared list of items
- Used to record the presence or absence of the item
- By checking 'yes' or 'no' or by inserting the appropriate word or number
- Matter of fact and not judgement or opinion
- In educational studies
- Educational appraisal studies of school buildings, text books, facilities available
- Recreation, laboratory, library etc.

Characteristics of Checklist

- Checklist is used for evaluation of **self and others.**
- It is used as an **instrument of observation**.
- It involves **questions and its answers**.
- It involves **signs by the respondent**.
- It involves the characteristics about a **particular subject** to be evaluated

Construction and Application of Checklist

- The first **horizontal line** of the check list is used to write the **name or number of the subject under observation**.
- **The characteristics of the subject or thing to be evaluated** are arranged in **vertical column** of the evaluation sheet with the corresponding **blank options to place** the tick mark in the adjacent columns.

- Then the characteristics present in the subjects under observation are decided and if that characteristic is present in the subject then the **tick mark** is placed in that column.
- Then after the frequency of **all tick mark is counted and marks are given to students** on the bass of predefined norms or standards.
- Then the **percentage, mean, median or correlation** is used.

- A check list is **simple tool** of evaluation.

- It is easy to make, easy to use, and yet can be constructed to cover **various aspects of an individual's behaviroural adjustment**.

- In preparing a checklist a teacher must keep in mind **what kinds of behaviours** are important to record and **what kind of objectives** are to be evaluated

Example

WEEKLY CHORES

Cleaning:	
	Vacuum
	Mop Floors
	Dust
	Clean Bathroom
	Empty Trash
	Clean out fridge

Laundry:	
	Adult Clothing
	Kids Clothing
	Towels
	Sheets
	Tablecloths

Meals:	
	Plan weekly menu
	Make grocery list
	Grocery Shopping

Financial:	
	Pay Bills
	File Papers

Errands:	

Seasonal:	
	Mow lawn
	Vacuum Pool
	Weed Garden

DAILY CHORES (Check when completed)

	Sun	Mon	Tues	Wed	Thurs	Fri	Sat
Dishwasher							
Open Mail							
Wipe down counters							
Straighten Rooms							
Make beds							

Notes for this week:

Uses of Checklist

- It is useful for **survey and research**.
- The **amount of characteristics or traits** of subjects can be known.
- It is helpful to **give the appropriate guideline** to the subjects.
- To know the **developmental direction of the specific behavior pattern** check list is used.
- It is useful for **self-evaluation and other's evaluation**.

Limitations of Checklist

- As only sign is used in checklist therefore no other options are found. It is **subjective and biased**

- It is **difficult to evaluate the personality of student or adjustment capacity** through checklist

Rating Scale

❖ Meaning

- A rating scale is a set of categories designed to elicit information about a quantitative or a qualitative attribute.
- Rating Scale is an important technique of assessment. 'Rating' is the assessment of a person by another person.
- Ruth Strang calls it 'directed observation'

Definition

- "Rating is a term applied to expression of opinion or judgement 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"

- A.S.Barr

Rating Scale

- Personality of a individual is not assessed by himself but by other persons who know the individual well
- Student – teacher
- Teacher rating, personality rating, testing the validity of many objective instruments like paper-pencil inventories of personality & School appraisal

- A rating scale is an improvement over check list.
- While a checklist simply records that something happened, a rating scale adds another dimension: how much or how well it happened.
- In a commonest type of rating scale, the columns opposite the list of traits may be captioned wither in quantitative terms such as - *always, sometimes, Never* or in qualitative terms such as *Good, Average, Poor.*

- **Example :**
- Limited number of items to which values on a scale have to be assigned
- The value be represented in the form of a number or one among a series of worded descriptions
- Usual to have 5 to 7 points on the scale for every item to be rated

Highly emotional	Occasionally emotional	Socially average	Very rarely emotional	Not at all emotional

Types of Rating Scales

❖**Numerical rating Scales:**

- In this numbers are assigned to each trait.
- If is a seven-point scale, the number 7 represents the maximum amount of that trait in the individual, and 4 represents the average.
- The rater merely enters the appropriate number after each name to indicate judgement of the person

❖Method of Paired Comparison :

- In this the rater compares each person being rated with respect to the trait of every other individual, being rated in the general terms of 'equal', 'better', or 'worse'.

❖**Graphical Rating Scales - are most commonly used**

Examples are:

How do you like the product?
(Please check)

Like very much	Like some what	Neutral	Dislike some what	Dislike very much

❖Score cards

- It is a type of scale in which whatever is being rated is analysed into its component parts
- An expert assigns each part a maximum score
- The rater assigns a value to each items and he passes judgement and these values are totaled and a final score is pronounced

❖Man to Man Scale

- In this case, an individual is asked to rate the person to be rated (the rate) by comparing him to a person already rated and assigned a position on the scale.
- The rate is assigned his position

Errors in rating

❖**Generosity Error.**

- This means, sometimes the rater does not want to run down the rate, the latter being his favourite, by giving him low ratings.

❖**The Error of Halo Effect.**

- Sometimes. the rating is done in terms of general impressions about the rates formed on the basis of some previous performance

❖The Error of Central Tendency.

- Sometimes the rater has a tendency to rate all or many of the rates near the mid-point on the scale

Limitations

- **Hallo – Effect**
- Rater frequently carry over one generalised impression of the person from one rating to another
- **Generosity Error**
- Rater develops a tendency to over estimate the desirable qualities of the rate whom he likes
- **Constant Error**
- There is a tendency on the part of the rater to see others as opposite to himself on a trait
- **Average Category**
- Rater have a tendency to play it safe and may mark all items in the centre

TEST

❖Definition

- "A test is a systematic procedure for observing and describing one or more characteristics of a person with the aid of either a numerical scale or a category system"

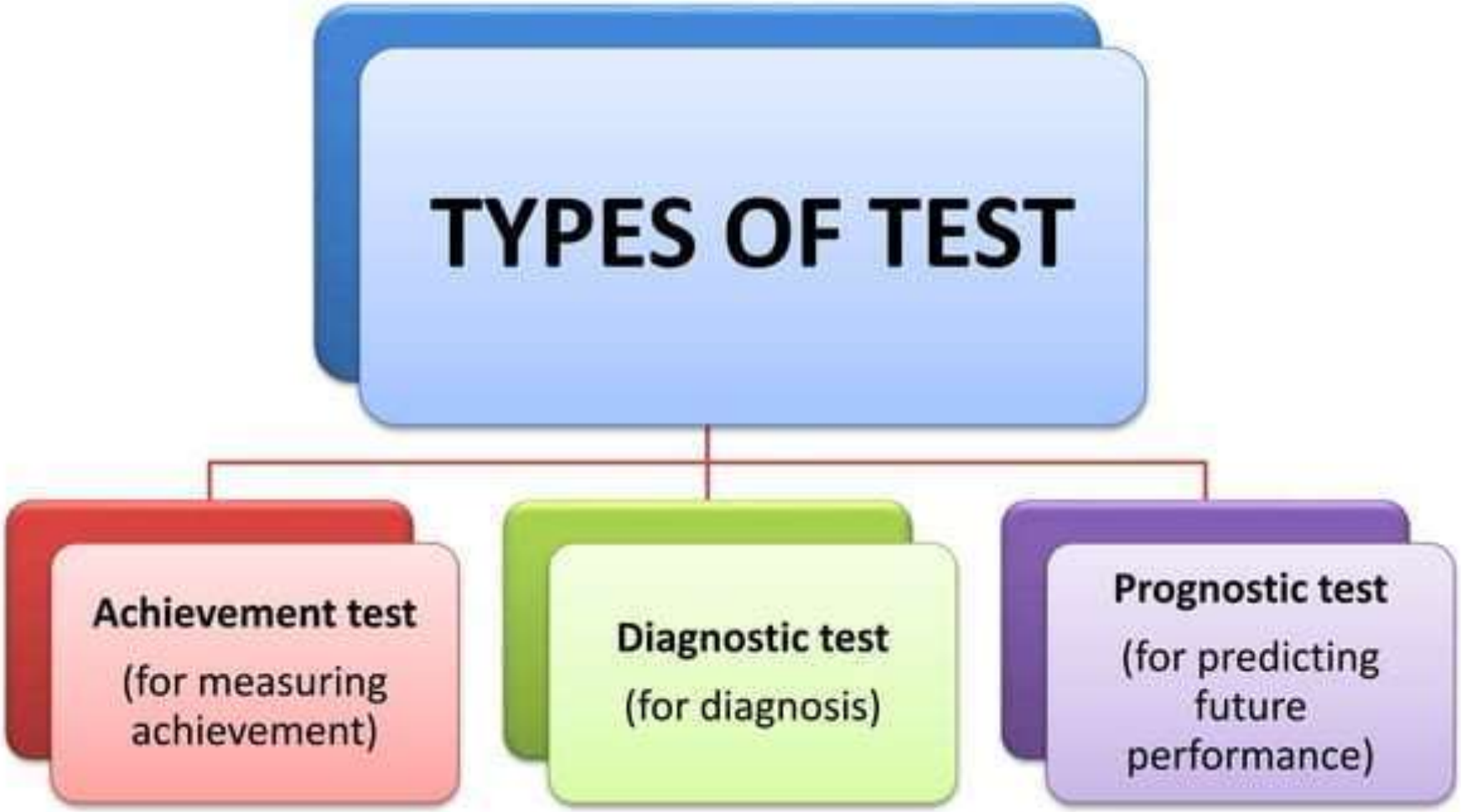
- L..J. Cronback

- "A test is a compact task or series of tasks designed to ascertain the merit or quantity of something. Educational tests constitute a series of items for which a score is obtained. Depending on how they are constructed, they can serve a purpose"

- A Dictionary of Education

❖Purposes:

- For the **grading** of students,
- For **individual attention** required by any student,
- For creating **merit** and so on.
- It tells how to **shape the learning methodologies.**
- There are a number of forms of tests that can be taken in a classroom. **Give insight of each and every student of the classroom**.


TYPES OF TEST
Achievement test
(for measuring achievement)
Diagnostic test
(for diagnosis)
Prognostic test
(for predicting future performance)

TYPES OF ACHIEVEMENT TESTS

STANDARDIZED TESTS

TEACHER-MADE TESTS

TYPES OF STANDARDIZED TESTS

APTITUDE TESTS

ATTITUDE TESTS

INTELLIGENCE TESTS

INTEREST TESTS

PERSONALITY TESTS

TYPES OF TEACHER-MADE TESTS

WRITTEN

- ESSAY TYPE
- SHORT ANSWER TYPE
- OBJECTIVE TYPE

Rubrics

Rubric

	Levels of Achievement			
Criteria	Unacceptable	Below Average	Acceptable	Above Average
Punctuality of Participation	0 Points Student did not participate in this forum	1 Points Student had an initial post to some questions by due date	2 Points Student had an initial post to all questions by due date.	3 Points Student had an initial post to all questions by due date and responded to one classmate by the due date.
Quality of Initial Posting	0 Points No initial posting	1 Points Response was not on topic, the message was unrelated to assignment, and post was less than 150 words.	2 Points The initial posting did not adequately address the question posed in the forum; superficial thought and preparation.	3 Points Initial posting demonstrates sincere reflection and answers most aspects of the forum; full development of concepts is not evidenced.
Responses to Classmates	0 Points Did not post any responses to other students in this forum	1 Points Posted a response to only one other student's initial posting.	2 Points Responds to at least two other students' initial posting; posts shallow contributions to discussion; responses neither promoted nor enriched further discussion.	3 Points Responds to at least two other students' initial postings; provides comments and new information to other posts; not all responses promote further discussion of the topic.
Language and Grammar	0 Points No postings for which to evaluate language and grammar	1 Points Poorly written initial posting and responses including frequent spelling, structure, and/or grammar errors and/or the use of slang.	2 Points Communicates in friendly, courteous, and helpful manner with some spelling, grammatical, and/	3 Points Contributes valuable information with minor grammatical or structural errors.

Meaning

- A rubric is typically an **evaluation tool** or **set of guidelines** used to **promote** the consistent application of learning expectations, learning objectives, or **learning standards** in the classroom, or to **measure** their **attainment** against a consistent **set of criteria**.

- The word rubric comes from the **Latin word** for **red.**

- The **online Merriam-Webster dictionary** lists the **first meaning** of *rubric* as "an authoritative rule" and the **fourth meaning** as "a guide listing specific criteria for grading or scoring academic papers, projects, or tests"

- A rubric is a coherent set of criteria for students' work that includes descriptions of levels of performance quality on the criteria.

- It should be clear from the definition that rubrics have **two major aspects**: coherent sets of criteria and descriptions of levels of performance for these criteria.

Steps to Create a Scoring Rubrics

- Example
- Explaining the scoring rubrics by taking the **scoring of writing skill of student**
- **Step· 1**
- Teacher uses the **assignment** of students for **scoring for assessing the writing skill**, and he gives a clear **guideline** for writing the assignment.

- **Step· 2**
- Teacher **discusses the qualities** for writing assignment with the students and **fixes criteria** for the good qualities of an assignment.

- **Step· 3**
- The teacher discusses **how he grades** the quality of an assignment and he also discusses the **grading of the assignment** he is going to give them in the classroom.

- **Step· 4**
- The teacher shows the students **how he gives marks** to the assignment by using a **model assignment.**

- **Step· 5**
- The teacher asks the students **to create scoring rubrics for the sample assignment** using the guidance given by him.

- **Step· 6**
- The teacher **gives the scoring rubrics and gets feedback** from the students for improving it.

- **Step· 7**
- Finally he uses the **scoring rubrics for scoring the assignment** given to the students.

Rubrics Importance

- *The main purpose of rubrics is to* ***assess performances.***
- Rubrics help students and teachers define **"quality"**
- When students use rubrics regularly to judge their own work, they **begin to accept more responsibility** for the end product. It cuts down on the "am I done yet?" questions.

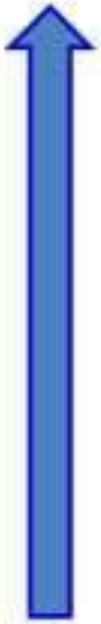
- Rubrics **reduce the time** teachers spend grading student work and **make it easier** for teachers to explain to students why they got the grade they did and what they can do to improve.
- **Parents** usually like the rubrics concept once they **understand** it, and they find rubrics **useful** when **helping with homework**. As one teacher says: **"They know exactly what their child needs to do to be successful"**

ASSESSMENT TOOL FOR AFFECTIVE DOMAIN

- Affective Domain
- Prepared in **1964** by **Dr.Bloom, Kruthwohl and Masin.**
- Emphasizes the **development of the heart**.
- It deals with **values, attitudes, interest, and appreciation, social and emotional adjustment**.
- If the objectives of the affective domain are attended to and achieved, the **evaluator** will be in a position to **predict the behavior of an individual**.

- Affective taxonomy is divided into **five major classes** arranged in a **hierarchical order** on the basis of the level of involvement.

- Characterization
- Organization
- Valuating
- Responding
- Receiving



- Receiving
- Learner should be **sensitized** to the existence of **certain phenomenon and stimuli** in his environment. This includes **awareness, willingness** to receive and give controlled or **selected attention**. (Ability to discriminate the stimuli)

- Responding (Interest objective)
- This is with response that goes beyond merely attending to phenomena. A person is **actively involved** in attending to them.
- This involves **responding, willingness to respond and satisfaction** in response. The person attaches emotional significance to the stimuli

- Valuating
- Stage of internalization-becomes part of the person.
- This includes **acceptance** of a value, **preference** for a value and **commitment** to a goal, idea or belief or a conviction in regard to a point of view.
- **Consistence in behavior**, which can be predicted. Forms principles of life and behavior is based on it.

- **Organisation**
- For situations where **more than one value** is relevant the need arises for the organization of the value into a system.

- The determination of the **interrelationship among them** and the establishment of the **dominant and pervasive value**.

- Characterisation by a value or value complex

- Organization becomes **part of the personality** and develops his **personality-characterisation**

- **Organization of values which control his behavior** is called **value complex**.

- Values like honesty, truthfulness, friendship develop through these levels of the affective domain.

- The values which are of **importance** to him will be **placed higher** up on the **hierarchy** and those which are of **less value go down**, thus creating an organization of values.

- The individual will then **behave in different situations according** to his/her value organization.

- So much so that **gradually** they become a **part of his personality** and **define his/her character.**

Attitude Scales or Opinionnaire

- Opinion and Attitude are allied but not Synonymous terms
- Attitude - Inner feeling or belief of a person towards a particular phenomenon
- Opinion – what a person says about his attitude towards some phenomenon

ATTITUDE SCALES

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

Measurement of attitude

- There are **numerous types** of attitude tests.
- The **most common** among them is **self-report inventory.**
- A large number of **statements** are given here and the person is **directly asked** what his attitude toward particular statement is.
- They are **technically** called **attitude scale**.

Two basic underlying assumptions

- An **individual's behavior** with respect to object or event will be **consistent** from one situation to another.

- Attitude **cannot** be measured **directly**. It is actually **inferred** from the **statements or actions** of the person. i.e., it is inferred from **verbal and non-verbal behavior** of the subject

Types of Attitude Scales

- Thurstone Technique of Scaled Values
- Likert Method of Summated Ratings

Thurstone scale

- **Five Steps**
- Step-I
- Several hundred **statements** expressing various degrees of **negative and positive attitude** towards the **objects or events** are collected. These statements can be obtained from **experienced people** or from popular literature on the issue

- Step-II

- Each statement is **written** on a **separate slip** of papers or cards. These statements are given to **different judges** to pile them on **11 point scale** ranging from **favourable to unfavourable.**

- Step-III

- This is the **scale construction stage**. These statements are **put into different scale points** on the basis on consensus of judges. On those statements where **judges do not agree** are **eliminated** here.

- Step-IV

- **Final statement** for the scale is **selected here**. Then the **subject** is **asked to mark** those statement with which he is **fully agreed** and he is then given **10 scored**

Thurstone Technique of Scaled Values

- 20 or more statements – express – groups, institution, idea or practice
- Submitted – panel of 50 or more judges
- 11 groups – assigning a position to an item
- Disagreement – discarded
- Median scale value – falls between 1 to 11
- Given to the subjects – check - agreement – responses - quantified

Likert scale

- The Likert Scale is the most **commonly used scale** in **quantitative research**.
- It is designed to **determine the opinion or attitude** of a subject.
- It contains a **number of statements** with a scale after each statement.
- The **original version** of the scale included **5 response categories,** and each response category was **assigned a value**.

- Usually, the **most negative response** is given a **numerical value of 1**, while the **most positive response** has a **numerical value of *5***.
- Obviously you can work out that the point in the **middle** would have a **numerical value of 3**.
- Response choices in a Likert Scale **usually address:**

➢ Agreement

➢ Evaluation

➢ Frequency

- **Example of a Likert Scale:**
- Statement:
- This method of teaching you about quantitative research methodology is useful
- Likert Scale:

➢ Strongly agree	5
➢ Agree	4
➢ Uncertain	3
➢ Disagree	2
➢ Strongly disagree	1

Likert Method of Summated Ratings

- Without the panel of judges
- Less time and efforts to construct
- Collecting a number of statements
- Express definite favourableness or unfavourableness
- Approximately equal statements
- Trial test – administrated to the subjects

- Eliminate – ambiguous
- SA, A, U, DA & SDA
- Favourable statements – 5, 4, 3, 2 & 1
- Unfavourable statements – 1, 2, 3, 4 & 5
- 50 * 5 = 250 – Most Favourable response
- 50 * 3 = 150 – Neutral attitude
- 50 * 1 = 50 – Most Unfavourable attitude

MOTIVATION SCALE

- The Motivation Assessment Scale (MAS) is a rating scale that assesses **functions of problem behavior in individuals** with developmental disabilities through informant responses.
- It includes **16 questions** and is comprised of **four subscales** that each represents a possible function of the behavior: **attention, escape, sensory, and tangible.**

- Each question has **six response options**
- (0 = never, 1 = almost never, 2 = seldom, 3 = half the time, 4 = usually, *5* = almost always, and 6 = always).

- Scores are calculated by **summing** the item ratings within a **particular subscale/function** and **calculating the mean** rating for that subscale.

- **High scores** for one or more of the subscales suggest that **those functions** may be maintaining the individual's problem behavior

- Never=O
- Almost Never=l
- Seldom=2
- Half the Time=3
- Usually=4
- Almost Always=5
- Always=6

Interest Inventories

- The tools used for describing and measuring interests of individuals – Interest Inventories or Interest Blanks
- Self-report instruments – in which the individuals note their own likes and dislikes
- Frequently used in educational and vocational guidance and in case studies
- Defined as eagerness, attention, curiosity, likes and dislikes

Measurement of Interest

- E.K.Strong's Vocational Interest Blank (SVIB)
- G.F.Kuder Preference Record (KPR)
- Occupational Interest Inventory

E.K.Strong's Vocational Interest Blank (SVIB)

- Designed by **E.K. Strong in 1919**.
- It consists of **420 items** concern with various occupations, school subjects, amusements, activities of people etc.
- It has been applied to persons in **various professions** such as law, medicine, teaching and engineering etc.,
- According to strong **some interests are common** to all these professions.
- It is **not concerned with ability**.
- It is **applicable for adults**.

- It is available in **four forms** for **men, women, students and those who have left school long back.**
- About **40-50 minutes** are given for taking the test.
- The **reliability** is about **0.80** validations it quite difficult.
- E.K. Strong validated after **16 years** who had taken test.
- The **norms** have been developed for the inventory.

Limitations

- The **accuracy of statements** made by the subject **cannot be tested**.
- There is the **problem of stability of in**terests as it is an acquired trait, the interests change at the different stages of development.
- It **does not indicate the success** in the occupation

E.K.Strong's Vocational Interest Blank (SVIB)

- Designed by E.K. Strong in 1919
- Classified for men, women, students and those who left the school long back
- Blank for men – 420 items with 8 divisions
- Subject will indicate – Like (L), Dislike (D) and Indifference (I) - Symbols
- 40 to 55 minutes
- 17 years of age and above

G.F.Kuder Preference Record (KPR)

G.F.Kuder Preference Record (KPR)

- The Kuder Preference Record has been developed for **high school and college students.**
- Each item of this inventory consists of **three preferences** such as-

(a) Study Physics

(b) Study of Musical Composition and

(c) Study Public Speaking.

- It consists of **198 items** in all, each item has three preferences. Preference is **measured in nine fields -** mechanical, scientific, computational, artistic, literacy, persuasive social service and clerical etc.

- It has **high reliability index 90**.

- Reference in this record can be **compared** with the strong vocational interest blank.
- Interest inventories are used in the **classification or selection of individual** for different occupations.
- It means that individual difference with regard to interests is **useful for vocational guidance as well as educational guidance and counseling purpose.**

- It provides the **basis for selection** of individuals for different jobs.
- It has the **administrative function**.
- Information about the interests of students is much **more useful for teachers** in school.
- In selecting the study subjects after delta class, **interests of the students are basic for their choices**.

G.F.Kuder Preference Record (KPR)

- High school and college 198 items
- Comprises of three preferences
- Ten fields – outdoor, mechanical, computational, scientific, persuasive, artistic, literacy, musical, social service and clerical

Occupational Interest Inventory

Occupational Interest Inventory

- Occupational Interest Inventory is designed to be **used in a wide spectrum of career guidance activities**.

- It helps candidates **choose an occupation, plan their career, and grow as professionals in the workplace.**

- The assessment, which is **based on the RIASEC** (Realistic, Investigative, Artistic, Social, Enterprising and Conventional) **model,** measures levels of interest in **12 domains** and **matches the candidate's profile** with a list of **80 occupations** across various sectors and fields.

Occupational Interest Inventory

- Orientation and mobility
- Skills assessment/training courses
- Recruitment

TYPES OF TEST ITEMS

- Multiple-choice tests
- Matching tests
- True-False tests
- Short-answer tests
- Problem Tests
- Oral Exams
- Essay Tests
- Performance tests

INTRODUCTION

- Achievement tests are conducted using different types of test items. Hence science teacher should master the skills of constructing test items. A constructor should take the following precautions while framing the test items.

The items should cover as far as possible, the whole range of topics prescribed in the syllabus.

No item or part of the item should be set which is outside the syllabus.

More items should be set to test higher objectives. For this purpose items should be in the context of new situations

Items should provide clear direction to the students regarding the scope and length of responses

The language of the items should be simple and within the easy grasp of students.

TYPES OF TEST ITEMS

1. Objective Type test item

A) True – false items (Alternate response type)

b) Multiple choice Test Items (Changing Alternative type)

c) Matching Type Test Item

d) Simple Recall Type Test Items

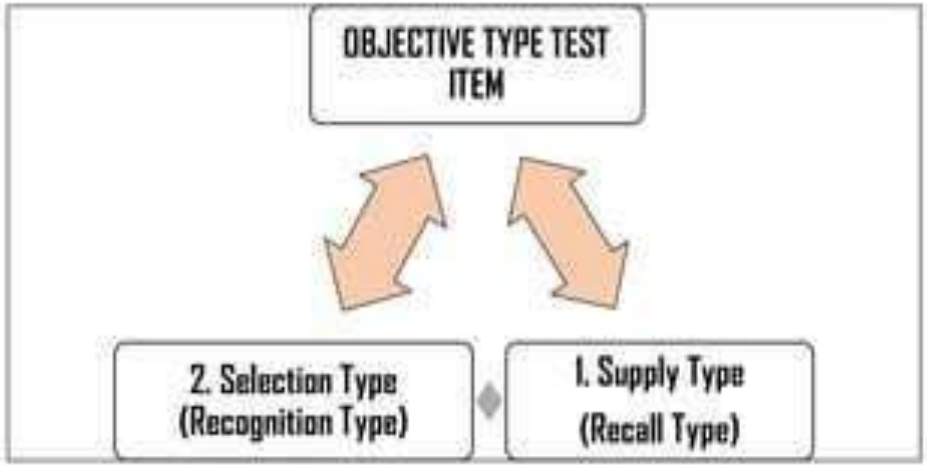
e) Completion Type Test Item

2. Short Answer Type

3. Essay Type

1. OBJECTIVE TYPE TEST ITEM

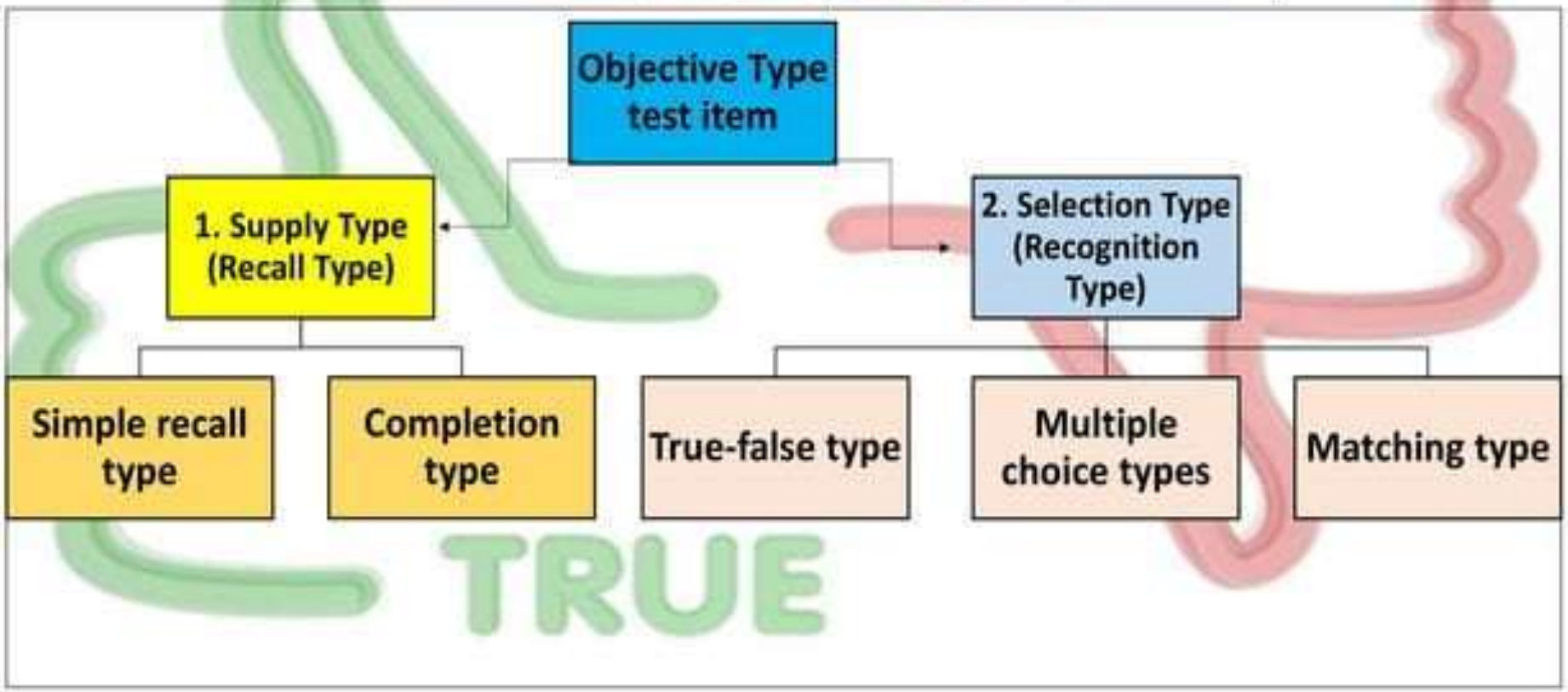
- An objective type test item is one in which the response will be objective. The responses are made fixed and hence the freedom of the respondent to deviate subjectively is restricted.



Selection type they have to select the responses from among the given responses.

Supply type test items the respondents have to supply the response

OBJECTIVE TYPE TEST ITEM



shutterstock.com · 428411002

A) TRUE – FALSE ITEMS (ALTERNATE RESPONSE TYPE)

- The respondent is asked to read a statement and indicate in some specific manner suggested, whether it is .....
- It tests the ability to discriminate between misconceptions and scientific truth.
- It is suitable for young children who have poor vocabulary.
- Large sample of subject matter can be covered within a short period.

- ✓ True / false
- ✓ Right / wrong
- ✓ Correct / incorrect
- ✓ Agree / disagree
- ✓ Yes / no

B) MULTIPLE CHOICE TEST ITEMS (CHANGING ALTERNATIVE TYPE)

- These are items presenting **four or more responses** in which **one is either correct or definitely better than the others.**
- The examinee has to **find this out and record** this in the manner required in the paper. Here **the chances of guess work are minimized**. Multiple choice items consist of two parts.

Multiple choice items

1. Stem

Presented in the form of a direct `question or incomplete statements

2. Options / Alternatives / Response

Usually four or five in number among the options one is the keyed response and others are called distracters or misleads or foils

Contd.,

- The stem gives data for the selection of the keyed response. The respondent has to read the stem and options and select the correct or best alternative.

- ✓ correct answer form
- ✓ best answer form
- ✓ multiple response form

C) MATCHING TYPE TEST ITEM

- This is a modified version of the multiple choice test items. In fact matching type is an economized form of combining a number of multiple choice items in the same question- a condensation of several multiple choice items. It consists of two parallel columns,

MATCH THE FOLLOWING	
CONTENT TO BE MATCHED	CONTENT TO BE MATCHED
with each phrase, word or number or symbol in one column	being matched to a word, phrase or sentence in the other column.

Contd.,

- The items in the column for which a match is sought are called premises or stem and items in the column from which selection is made is called responses or options.
- The respondent is required to make some sort of association between each premise and each response in the two columns

D) SIMPLE RECALL TYPE TEST ITEMS

This test requires the respondent to recall a response to a direct question.

The typical response should be short preferably a word, a number or a small phrase.

It eliminates the chance of guessing.

E) COMPLETION TYPE TEST ITEM

It consists of a series of sentences in which certain words are omitted and replaced by blanks.

The respondents are expected to fill in the blanks with a word or a number or at the most a phrase.

The probability of guess work is completely eliminated.

ADVANTAGES OF OBJECTIVE TYPE ITEM

1. Large number of question are set, a wide coverage of the syllabus is possible.
2. Questions can be set which are designed to assess one particular educational quality.
3. Marking of such tests is objective and can be done speedily
4. They are more valid and reliable, since the response/ answers are definite
5. There is greater administration use and control
6. They have higher diagnostic value
7. They are less time consuming

DISADVANTAGES OF OBJECTIVE TYPE ITEMS

1. Such tests do not encourage verbal fluency or a student's ability to development argument
2. Chancing of guessing are high
3. An objective test is difficult and expensive to construct
4. Emphasis on testing superficial knowledge
5. Inefficiency in testing complicated skill.
6. Objective type items are often ambiguous, particularly for the better students.
7. Such tests when over used can have a negative effect on teaching, since they encourage the student to learn bits of knowledge rather than the whole.

2. SHORT ANSWER TYPE

A question requiring value points at the most may be defined as a short answer question.

The term value points indicates a point to be given credit in the expected answer.

Thus the length of the answer expected from a short answer question becomes very short.

This diminishes subjectivity.

Such question are of great helping having wide coverage of content and each item can be set to a test a definite objective. Because of this reason, a fair proportion of such questions should be included in a test

ADVANTAGES OF SHORT ANSWER TYPE

1. Questions of this form can be made stimulating

2. Students can be trained to select relevant information and present it in a few short, crisp Sentences

3. Short answers are easy to score

4. Reliability of scoring is high

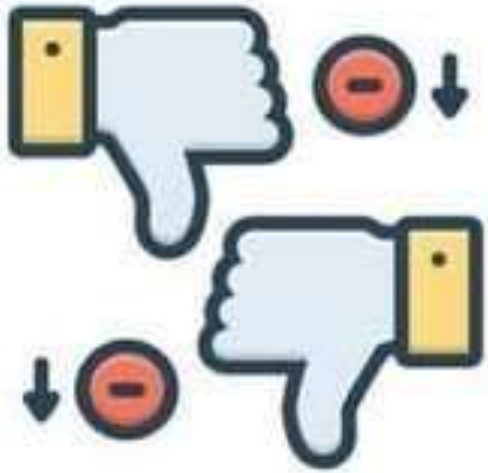
5. Questions can cover a wider content area.

6. It is possible to achieve a more expensive sampling in the short answer test than in the essay type test

7. The short answer test is especially useful in diagnosis. In part, this follows from the factor of extensive sampling

DISADVANTAGES OF SHORT ANSWER TYPE

- 1. It is more subjective than the objective type of items
- 2. Its excessive use may encourage a student to memorize facts and develop poor study habits
- 3. Mechanical scoring is not possible because of the subjectivity involved



shutterstock.com · 1647261361

3. ESSAY TYPE

- "Essay test is a type of examination in which the subject or examinee is asked to discuss, enumerate, compare, state, evaluate, analyze, summarize or criticize and involves writing at specific length on a given topic involving the processed listed above".

  - Good

- The essay type questions get its name from the manner in which the examinee responds.
- The term essay implies a written response which may consists of many sentences to several pages.
- The student is allowed freedom with respect to what his answer will include its wording length and organization.

ADVANTAGES OF ESSAY TYPE

They are easy to construct

They can be used to test the student's language mastery, expression and organizational ability of a student

Chances of copying are minimal

A student's ability to use knowledge effectively can be assessed.

It helps to develop a variety of skills.

It encourages good study habits.

Highlight important units, look for relationships and exercise judgement in deciding points of emphasis.

DISADVANTAGES OF ESSAY TYPE

1. Guessing creates few problems

2. Subjective bias could creep in as these test are based on the examiner's moods and whims

3. It encourages rote memory.

4. The higher levels of the cognitive domain cannot be completely assessed by this method.

5. Sampling is limited.

6. Time limitations make it impossible to achieve good sampling in an easy test, assuming that a large body of subject matter has been covered.

DISADVANTAGES OF ESSAY TYPE

7. There is danger of bluffing.

8. The "gift of gap" can be encountered in written as well as in oral communication. It requires a discerning teacher to realize that nothing much is been said.

9. Essay type test are difficult to score.

10. Besides no two teachers agree on the score given to a particular paper, the hand writing, presentation and so on. Thus score rating cannot be generalized.

Principles for constructing test items

PRINCIPLES FOR CONSTRUCTING TEST ITEMS.

- Test construction requires a systematic organized approach if positive results are to be expected. Firstly the objective must be well defined. There are numerous points which are common t all types of tests and items which must be observed in constructing a test. Some of the more important are given below.

1. Avoid obvious ,trivial, meaningless and ambiguous items.
2. Observe the rules of rhetoric, grammar and punctuation.
3. Avoid items that have no answer upon which all experts will agree.

Principles of test construction

1. **Measure all instructional objectives**
- Objectives that are communicated and imparted to the students.
- Designed as an operational control to guide the learning sequences and experiences.
- Harmonious to the teachers instructional objectives.
2. **Cover all learning tasks**
- Measure the representative part of the learning task.
3. **Appropriate testing strategies or items**
- Items which appraise the specific learning outcome.
- Measurements or tests based on the domains of learning.

Principles of test construction

4. Make test valid and reliable

- reliable when it produce dependant, consistent, and accurate scores.
- Valid when it measures what it purports to measure.
- Test which are written clearly and unambiguous are reliable.
- Tests with fairly more items are reliable than tests with less items.
- Tests which are well planned, covers wide objectives, & are well executed are more valid.

Principles of test construction

5. Use test to improve learning

- tests are not only an assessment but also it is a learning experience.
- Going over the test items may help teachers to reattach missed items.
- Discussion and clarification over the right choice gives further learning.

6. Norm referenced and Criterion referenced tests

- Norm Referenced: higher and abstract level of cognitive domain.
- Criterion Referenced: lower and concrete levels of learning.

Criteria for a good test

The following seven points are taken as criteria of a good test.

- A good test must possess a very high validity.
- It must have a high reliability.
- It must be very objective in nature .
- It must be very comprehensive
- It must pick out the good students from the poor i.e it must possess high discriminating power .
- It must be easy to use .its administration and scoring must be easy and there must be economy of item and effort.
- Norms established on the basis of its results must be satisfactory .

VALIDITY

A test should measure what it intends to measure and not what we do not want it intends to measure and not what we do not want to measure for example a test on general scientific knowledge should only measure scientific knowledge and not linguistic or arithmetical ability of the child.

RELIABILITY

A test is reliable when it functions consistently it must function similarly with similar groups it should rate the same candidates at the same score even if it is examined by the same or different examines at the same or different times the difference in score should be negligible.

OBJECTIVITY

It should yield the same or nearly the same score irrespective of the person who scores it the scorers personal judgement like and dislikes should not influence the scoring

PRACTICABILITY

Students should finish the test in the time allotted for it it should neither be too long nor too short it should keep the students busy all the time from disciplinary and administrative point of view some provision should be made for individual differences and some questions may be set for brighter students it should be manageable within the funds provided

SCORABILITY

It should be constructed in such a way that the boredom caused by the routine of scoring is cut down to the minimum it should allow the use of key in marking so that the test sheet is checked more or less automatically and without subjectivity it saves a great deal of time and simplifies the work of the teacher with the help of scoring key

clarity

The directions given should be brief and definite so that pupils are not handicapped with wrong performance due to misunderstanding of directions the language of the questions should be simple understandable definite and unambiguous

COMPREHENSIVE

It should cover the whole syllabus due importance should be given to each topic while setting the paper care should be taken that minimum choice is given .

GRADED

It should be according to the age and intelligence of the students.

EVALUATION

The evaluation of test results involving as it does all the problems of scoring and weighting of items and the assignment of relative weights to tests used in a battery , it surrounded with highly technical considerations.

Norm-Referenced Tests

- The norm-referenced test measures an **individual's achievement at a given period of time compare to students elsewhere**.
- The norm-reference test is **valuable** in **measuring higher and abstracts level of cognitive domain**
- The norm-referenced test is valuable for **heterogeneous groups** in which the **range of abilities is wide** and a test is intended to **measure a wide range of performance**.

Criterion-Referenced Tests

- **Criterion-referenced** tests and assessments are designed to **measure student performance against a fixed set of predetermined criteria** or learning standards—i.e., concise, written descriptions of what students are expected to know and be able to do at a specific stage of their education.

Criterion-Referenced Tests

- Scores from a criterion-referenced test do not indicate **a relative level of achievement or produce standards because no comparisons are made.**
- The criterion-referenced test is valuable for **measuring lower and concrete levels of learning.**
- The criterion-referenced test is more useful in **homogeneous groups** in which the **range of abilities is narrow** and a test is intended to **measure a limited range of objectives and outcomes**.

NORM-REFERENCED TESTS	CRITERION-REFERENCED TESTS
1.Norm-referenced tests are used to determine the achievements of individuals in comparison with the achievements of other individuals who take the same test.	1. Criterion-referenced tests are used to determine the achievements of individuals in comparison with criterion, usually an absolute standard.
2. In a norm-referenced test, the quality of achievements of a student is determined by the distance of his score from the mean or median	2. In a criterion-referenced test, the quality of achievement of a student is determined by the distance of his score from the criterion established.
3. Norm-referenced tests are designed to produce variability among individuals.	3. In criterion-referenced tests, variability is irrelevant.

Assessment Issues

Commercialization of Assessment

- Turning **student assessment into a profit-making activity**.
- Involves **private coaching centres, publishing companies, and test agencies**.
- Assessment becomes **market-driven**, not **learner-centered**.

Features of Commercialization

- High dependency on **private coaching**.
- Mushrooming of **test-preparation industries**.
- Selling **model question papers**, answer keys, and test series.
- Focus on **marks and ranks**, not learning outcomes.

Causes of Commercialization

- Overemphasis on **examinations and marks**.
- Competitive entrance tests (NEET, JEE, UPSC, etc.).
- Poor quality of school assessments.
- **Parental pressure** and social status.
- Lack of **trust** in public education assessment.

Effects on Students

- Stress and anxiety.
- Learning becomes **rote-based**, not conceptual.
- Inequity – **only the rich can afford coaching**.
- **Low confidence** in self-assessment and school exams.

Effects on Teachers and Schools

- Teachers lose control over **classroom-based assessments**.
- Focus shifts to **test practice** instead of holistic teaching.
- Demotivation due to reduced role in learner evaluation.

Effects on Education System

- Undermines the purpose of **Continuous and Comprehensive Evaluation (CCE)**.
- Ignores **skills, values, creativity, and co-curricular talents**.
- Encourages **corruption, favoritism**, and unfair means.

Ethical and Equity Issues

- Education becomes a **commodity**, not a right.
- Commercial tests may lack **validity and reliability**.
- Access denied to **marginalized and rural students**.

Countering Commercialization – What Can Be Done?

- Strengthen **school-based assessments**.
- Encourage **formative and diagnostic evaluation**.
- Promote **open educational resources (OERs)**.
- Raise **awareness among parents and students**.
- Government regulation on **coaching centers and test publishers**.

Poor Test Quality

- A test that **fails to measure what it is intended to measure**.
- It lacks **validity, reliability, clarity, and alignment** with learning objectives.
- Often includes **ambiguous, irrelevant, or poorly constructed items**.

Characteristics of Poor Test Quality

- **Ambiguous wording** in questions.
- **Unbalanced coverage** of syllabus or units.
- **Guessable options** in multiple-choice questions.
- Overuse of **memory-based** items.
- Poor or no **blueprint** used in test construction.
- **No clear marking scheme or rubrics**.

Causes of Poor Test Quality

- **Lack of teacher training** in test construction.
- Time pressure during test preparation.
- No use of **test blueprint** or table of specifications.
- Focus only on **summative evaluation**.
- Dependence on **readymade or recycled questions**.

Consequences of Poor Test Quality

- **Unfair evaluation** of student ability.
- Students may **lose interest or confidence** in learning.
- Teachers cannot make **data-driven instructional decisions**.
- May result in **grade inflation or deflation**.
- Hampers **diagnostic and remedial** planning.

- **Types of Common Poor Test Items**
- Example of Poor Quality
- MCQs Options are too obvious or confusing
- Essay -Vague instructions, no time limit
- Fill-in-the blank - Multiple correct answers possible
- True/False Statements that are ambiguous or double-barreled

Evaluation Dimensions Affected

- **Validity** – Does it measure what it's supposed to?
- **Reliability** – Is it consistent across time and scorers?
- **Usability** – Is it easy to administer and score?
- **Objectivity** – Is it free from bias?
- **Fairness** – Is it equally accessible to all learners?

Remedies to Improve Test Quality

- Use of **test blueprint** and specification table.
- Training in **item writing and validation**.
- Peer review and **moderation of question papers**.
- Inclusion of **higher-order thinking questions**.
- Use of **rubrics and scoring keys**.

Role of Teachers

- Plan tests as a part of teaching-learning.
- Ensure **content validity** and representation.
- Promote **formative and authentic assessments**.
- Use **feedback** from test results for improvement.

Measurement Issues

- Problems in **quantifying learning outcomes accurately**.
- Arise due to:
  - Faulty test construction
  - Improper tools and techniques
  - Human errors in scoring
  - Biases and inconsistencies

Types of Measurement Issues

- **Lack of Reliability**
- **Lack of Validity**
- **Subjectivity in Scoring**
- **Measurement Errors**
- **Bias and Fairness Issues**
- **Improper Use of Tools**

Consequences of Measurement Issues

- **Misleading results** about student learning.
- **Inappropriate feedback** and grading.
- **Reduced student motivation**.
- **Flawed decisions** in promotion, placement, or intervention.

Ways to Minimize Measurement Issues

- Use **blueprint and specification table**.
- Ensure tests are **valid and reliable**.
- Use **standardized scoring rubrics**.
- Provide **scorer training**.
- Use **triangulation of tools** (written + performance + self-assessment).
- Conduct **item analysis and test reviews**.

System Issues in Assessment

- Flaws at the level of:
  - **Curriculum design**
  - **Assessment policies**
  - **Teacher preparation**
  - **Administrative pressures**
  - Result in assessments that are **mechanical, high-stakes, and narrow in focus**.

Key System Issues in Assessment

- **Overemphasis on Summative Assessment**
- **Examination-driven Learning**
- **Lack of Continuous and Comprehensive Evaluation (CCE)**
- **Inadequate Teacher Training in Assessment**
- **Policy-Implementation Gap**
- **Misuse of Assessment for Ranking and Filtering**

Impact of System Issues

- **Narrow learning goals**
- **Inequality in assessment** outcomes
- **Stress and anxiety** in learners
- Teachers feel **powerless** to change assessment practices
- Hinders **real educational reform**

Addressing System Issues

- **Policy monitoring and accountability**
- **Teacher empowerment through regular training**
- Use of **flexible, formative, and diagnostic tools**
- Ensure **assessment aligns with learning objectives**
- Involve **students and parents** in the process