

# Module for Certificate in Education Programme

2nd Semester  
April, 2023

**IoE/MoF/TUC/GHANA CARES TRAINING AND RETRAINING  
PROGRAMME FOR PRIVATE SCHOOL TEACHERS**



**Ministry of Finance**



**Trade Union Congress**



**Institute of Education, UCC**

## **EIN 055 - Introduction to Assessment in Schools**

**Prof. Eric Anane**

## UNIT 1: NATURE OF ASSESSMENT

The public uses the terms **assessment**, **test**, **measurement** and **evaluation** interchangeably, but is important for the student in assessment to distinguish among them. The meanings of the terms as applied to situations in schools are explained in the following paragraphs.

### Assessment

Assessment is the process of obtaining information that is used for making decisions about students, schools, curricula and programmes and educational policies. It includes the full range of procedures such as tests, observations, interviews, viva voce, laboratory work or quizzes used to gain information about student learning.

These procedures may be **formal** (pencil and paper tests; or computer adapted tests) or **informal** (observations, checklists and rubrics). Certain concepts and terms are associated with assessment. These are tests, measurement, and evaluation.

Assessment is the purposeful, systematic, and ongoing collection of information as evidence for use in making judgments of students' learning, curriculum, programmes,

### Test

A test is a task or series of tasks, which are used to measure specific traits or attributes in people. It is defined as an instrument or systematic procedure for observing and describing one or more characteristics of a student using either a numerical scale or a classification scheme.

In educational settings, tests include paper and pencil instruments, which questions help the test giver to obtain an estimate of the specific trait being measured. It answers the question, 'How well does the individual perform?' Two interpretations can be given to scores from tests. These are **norm-referenced** and **criterion-referenced** interpretations.

- Tests are instruments used to 'measure' students' characteristics or traits.
- In scoring a test, numbers are assigned to students based on the degree of correctness of the answers provided.
- A score is the product of testing.

### Measurement

The process of assigning numbers (scores/marks) to specified attributes or characteristics or traits possessed by a person, events, or a set of objects according to specific rules.

### **Scales of Measurement**

There are 4 types of measurement scales: Nominal, Ordinal, Interval and Ratio.

**Nominal:** Categorical data and numbers that are simply used as identifiers or names, represent a nominal scale of measurement. A nominal scale classifies persons or objects into two or more categories. Whatever the classification, a person can only be in one category, and purposes of identification, categories are numbered. E.g., for Sex: Male =1, Female =2.

**Ordinal:** An ordinal scale of measurement represents an ordered series of relationships or rank order. An ordinal scale not only classifies subjects but also ranks them in terms of the degree to which they possess a characteristics/attribute of interest. An ordinal scale puts subjects in order from highest to lowest or from most to least. With respect to height, 5 students from 1 to 5, the subject with rank 1 being the shortest and 5 being the tallest.

**Interval:** A scale that represents quantity and has equal units but for which zero represents simply an additional point of measurement is an interval scale. An interval scale has all the characteristics of both nominal and ordinal scales and in addition has equal intervals. The zero point is arbitrary and does not mean the absence of the characteristics/trait. Values can be added and subtracted to and from each other. But not multiplied or divided. Examples include Celsius temperature, academic achievement.

**Ratio:** The ratio scale of measurement is similar to the interval scale in that it also represents quantity and has equality of units. However, this scale also has an absolute zero (no numbers exist below zero). A ratio scale has all the advantages of the types of scales and in addition it has a meaningful **true zero** point. Height, Weight and time are examples. Values can be added, subtracted, multiplied and divided. Sixty minutes can be said to be 3 times as long as 20 minutes.

## **Evaluation**

Evaluation is defined as the process of making a value judgment about the worth of a student's product or performance. It is the systematic investigation of the worth or merit of an object (a person, programme or a book). For example, a teacher may judge a student's writing as exceptionally good for his grade placement.

**Formative evaluation** is the process of judging the worth of teaching and learning constantly. Formative evaluation of student's achievement means we are judging the quality of student's achievement while the student is still in the process of learning.

**Summative evaluation** is the process of judging the worth of teaching and learning at the end of the period of instruction. It is judgmental in nature. It attempts to determine the extent to which the broad objectives of teaching and learning have been achieved. In other words, it is judgment about the quality of students' achievement after instructional or learning process is completed.

## UNIT 2: PRINCIPLES AND PURPOSES OF ASSESSMENT

### Principles of assessment

Principles are fundamental truths and doctrines accepted by most authorities as characteristics of assessment. Some of them are:

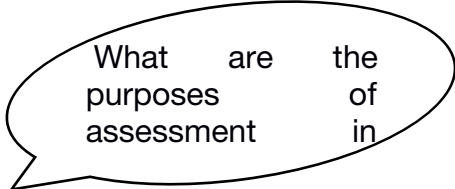
#### General Principles of Assessment

1. ***Test developer must be clear about the learning target(s) to be assessed.*** This involves clearly specifying the intended learning goals and helps to select the appropriate assessment technique.
2. ***The assessment technique selected must match the learning target.*** The main criterion is whether the procedure is the most effective in measuring the learning target. Assessment tasks should primarily reflect the nature of the discipline or subject but should also ensure that students could develop a range of generic skills and capabilities.
3. ***Assessment techniques must serve the needs of the learners.*** They should provide meaningful feedback to the learners about how closely they have approximated the learning targets.
4. ***Good assessments use multiple methods.*** Multiple indicators of performance provide a better assessment of the extent to which a student has attained a given learning target. Assessment needs to be comprehensive. Formative and summative assessment should be incorporated into the programmes to ensure that the purposes of assessment are adequately addressed.
5. ***Assessment is inherently a process of professional judgment.*** Proper use of assessment procedures requires that the user is aware of the limitations of each technique. In interpreting the results of the assessment, these limitations must be considered. Therefore, all those involved in the assessment of students must be competent to undertake their roles and responsibilities.
6. ***Assessment should be valid and reliable.*** Evidence needs to be provided that the interpretations and use of students' assessment results are appropriate and reliable. There is the need for assessment to be reliable and this requires clear and consistent processes for setting, marking, grading and moderation of assignments/tests.
7. ***Good assessment is fair and ethical.*** In assessing students, the rights and responsibilities of test takers, testing individuals of diverse linguistic backgrounds, and testing individuals with disabilities or special needs should be considered.
8. ***Good assessment appropriately incorporates technology.*** As technology advances and teachers become more proficient in the use

of technology, there will be increased opportunities for teachers and administrators to use computer-based techniques (e.g., item banks, electronic grading, computer-adapted testing, and computer-based simulations).

### Purpose(s) of Assessment

Educational assessment is conducted for a variety of reasons and the nature of the assessment often reflects the purpose for which it is being carried out.



Assessment provides information for decisions about students, curricula and programmes, and educational policy. These decisions are:

1. Instructional Management decisions
2. Selection decisions
3. Placement decisions
4. Counseling and Guidance decisions
5. Credentialing and Certification decisions

In all, purposes of assessment could be grouped under three main categories as Assessment of learning, as learning and for learning.

Assessment for learning	Assessment as learning	Assessment of learning
Assessment for learning is ongoing, diagnostic, and formative. It is for ongoing planning. It is not used for grading and Report Cards.	Assessment as learning actively involves students. It is ongoing, and it involves self and peer assessment. It provides students with the opportunity to use the feedback to improve learning. Allows time for self-edit.	Assessment of learning occurs at end of year or at key stages. It is summative. It is for grading and Report cards.
<ul style="list-style-type: none"> <li>• diagnostic and formative</li> <li>• teacher assessment, student self-assessment, and/or student peer assessment</li> <li>• criterion-</li> </ul>	<ul style="list-style-type: none"> <li>• self-assessment</li> <li>• the development of self-assessment skills</li> <li>• peer-assessment</li> <li>• the development of peer-assessment</li> </ul>	<ul style="list-style-type: none"> <li>• summative</li> <li>• teacher assessment</li> <li>• may be either criterion-referenced (based on prescribed learning outcomes) or</li> </ul>

<p>referenced – criteria based on prescribed learning outcomes identified in the provincial curriculum, reflecting performance in relation to a specific learning task</p> <ul style="list-style-type: none"> <li>• involves both teacher and student in a process of continual reflection and review about progress</li> <li>• teachers adjust their plans and engage in corrective teaching in response to formative assessment</li> </ul>	<p>skills</p> <ul style="list-style-type: none"> <li>• provides students with information on their own achievement and prompts them to consider how they can continue to improve their learning.</li> <li>• student-determined criteria based on previous learning and personal learning goals.</li> <li>• students use assessment information to make adaptations to their learning process and to develop new understandings.</li> <li>• provides the opportunity to respond to assessment feedback to improve a project.</li> </ul>	<p>norm-referenced (comparing student achievement to that of others)</p> <ul style="list-style-type: none"> <li>• information on student performance can be shared with parents/guardians, school and district staff, and other education professionals (e.g., for the purposes of curriculum development)</li> <li>• used to make judgments about students' performance in relation to provincial standards.</li> <li>• used for grading and Reporting on Cards</li> </ul>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

### UNIT 3: TYPES OF ASSESSMENT

#### Comparing formative and summative assessments

	FORMATIVE	SUMMATIVE
<b>Relation to Instruction</b>	<ul style="list-style-type: none"> <li>Occurs during instruction</li> </ul>	<ul style="list-style-type: none"> <li>Occurs after instruction</li> </ul>
<b>Frequency</b>	<ul style="list-style-type: none"> <li>Occurs on an ongoing basis (daily)</li> </ul>	<ul style="list-style-type: none"> <li>Occurs at a particular point in time to determine what students know</li> </ul>
<b>Relation to grading</b>	<ul style="list-style-type: none"> <li>Not usually graded – information is used as feedback to students and teachers, mastery is not expected when students are first introduced to a concept</li> </ul>	<ul style="list-style-type: none"> <li>Graded</li> </ul>
<b>Students' role</b>	Active engagement – self-assessment/Peer-assessment	<ul style="list-style-type: none"> <li>Passive engagement in design and monitoring</li> </ul>
<b>Requirements for use</b>	<ul style="list-style-type: none"> <li>Clearly defined learning targets that students understand.</li> <li>Clearly defined criteria for success that students understand.</li> <li>Use of descriptive versus evaluative feedback</li> </ul>	<ul style="list-style-type: none"> <li>Well-designed assessment blueprint that outlines the learning targets</li> <li>Well designed test items using best practices</li> </ul>
<b>Examples</b>	<ul style="list-style-type: none"> <li>Better thought of as a process rather than a thing. Examples include observations, interviews, evidence from work samples, paper, and pencil tasks</li> </ul>	<ul style="list-style-type: none"> <li>State assessments (WASSCE, BECE), interim assessments, end of unit assessments, common assessments</li> </ul>
<b>Purpose</b>	<ul style="list-style-type: none"> <li>Designed to provide information needed to adjust teaching and learning while they are still occurring</li> </ul>	<ul style="list-style-type: none"> <li>Designed to provide information about the amount of learning that has occurred at a particular point.</li> <li>Since summative assessments occur after instruction every few weeks, months, or once a year, they help to evaluate the effectiveness of programmes, school improvement goals, alignment of curriculum, or student placement in specific programs. Summative assessments</li> </ul>



		happen too far down the learning path to provide information at the classroom level and to make instructional adjustments and interventions during the learning process.
--	--	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------

#### **UNIT 4: ITEM FORMATS**

#### **Types of test item formats**

There are two major types of classroom achievement tests. These are the essay-type tests and objective-type tests.

<b>ESSAY TESTS (SUBJECTIVE)</b>	<b>OBJECTIVE TESTS</b>
1. Requires students to plan their own answers and to express them in their own words	Requires students to choose among several designated alternatives or write a short answer
2. Consists of relatively few items that call for extended answers	Consists of many items requiring only short answers
3. A lot of time is spent by students in thinking and writing when taking the test	A lot of time is spent by students in reading and thinking when taking the test.
4. Quality of test is determined largely by the skill of the test scorer	Quality of test is determined largely by the skill of the test constructor
5. Relatively easy to prepare but rather tedious and difficult to score	Relatively tedious and difficult to prepare but rather easy to score
6. Permits and encourages bluffing	Permits and encourages guessing.
7. Afford both the student and teacher the opportunity to be individualistic	Afford only the test constructor (teacher) the opportunity to be individualistic.
8. Score distribution varies from one scorer to another	Score distribution is determined largely by the test
9. Less amenable to item and statistical analysis	Amenable to item and statistical analysis
10. Scoring is subjective	Scoring is highly objective
11. Content validity could be low	Content validity could be high
12. Gives unreliable scores	Reliability of test scores could be high

## **OBJECTIVE-TYPE TESTS**

### **Description**

An objective test requires a respondent to provide a brief response which is usually not more than a sentence long. The tests normally consist of many items and the responses are scored objectively to the extent that competent observers can agree on how responses should be scored.

### **Strengths and advantages**

13. Scoring is easy and objective.
14. They allow an extensive coverage of subject content.
15. They do not provide opportunities for “bluffing” or “pooling”.
16. They are best suited for measuring lower-level behaviours like knowledge and comprehension and application.
17. They provide economy of time in scoring.
18. Student writing is minimized. Premium is not placed on writing but rather on reading.
19. They are open to item and statistical analysis.
20. Scores are not affected by extraneous factors such as the likes and dislikes of the scorer (e.g., no halo effect).

### **Weaknesses and disadvantages**

1. They are relatively difficult to construct.
2. Item writing is time consuming.
3. They are susceptible to guessing.
4. Higher-order mental processes like analysis, synthesis and evaluation are difficult to measure.
5. Places premium on student's reading ability and may encourage reproduction learning.

### **Some examples of Objective Tests**

#### **1. Multiple-Choice Questions (MCQs)**

A multiple-choice test is a type of objective test in which the respondent is given a stem and then is to select from among three or more alternatives (options or responses) the one that best completes the stem. The incorrect options are called foils or distracters.

Multiple-choice questions are probably the most widely used of objective tests. Such questions are normally composed of **four** parts:

- ❖ **Stem**-question or incomplete statement
- ❖ **Options**- suggested answers or completions
- ❖ **Distracters/foils**- incorrect responses
- ❖ **Key**- correct responses

#### **Advantages of MCQs**

1. Scoring is easy, accurate and efficient.
2. Highly objective measurement of student achievement.
3. They allow an extensive coverage of subject content.
4. They do not provide opportunities for bluffing.
5. They are best suited for measuring lower-level behaviours like knowledge and comprehension.
6. They provide economy of time in scoring.
7. Student writing is minimized. Premium is not placed on writing.
8. They are amenable to item and statistical analysis.
9. Scores are not affected by extraneous factors such as the likes and dislikes of the scorer.

### **Disadvantages of MCQs**

1. They are relatively difficult to construct.
2. Item writing is time consuming.
3. They are susceptible to guessing.
4. Higher-order mental processes like analysis, synthesis and evaluation are difficult to measure.
5. Places premium on student's reading ability.

### **Guidelines for constructing multiple-choice tests**

1. The central issue of the item should be in the stem.
2. Item should be concise, easy to read and unambiguously worded.
3. The options should be plausible. Distracters must be plausibly attracted to the uninformed.
4. All options for a given item should be homogeneous in content.
5. Sentences should not be copied from textbooks or from others (colleagues, friends and so on) and past test items. Original items should be prepared. This builds capacity in item writing.
6. All options for a given item should be homogeneous or similar in grammatical structure.
7. Repetition of words in the options should be avoided.
8. All options must follow syntax and punctuation rules.
9. Specific determiners which are clues to the best/correct option should be avoided.
10. Vary the placement of the correct options. No discernible pattern of the correct/best responses should be noticed.
11. The responses/options in agreement must be in alphabetical/sequential order. This reduces unnecessary searching on the part of the testees.
12. Items measuring opinions should not be included. One option should clearly be correct or the best.
13. The responses in agreement must be itemized vertically and not horizontally.
14. The responses in agreement must be parallel in form i.e. sentences must be about the same length.
15. Each option must be distinct. Overlapping alternatives should be avoided.

16. Avoid using “all of the above” as an option but “None of the above” can be used sparingly. It should be used only when an item is of the correct answer type and not the best answer type.
17. Avoid double negatives. Do not use combinations of these words in the same question: not, no, nor the -un prefix, etc.
18. Stems and options should be stated positively. However, a negative stem could be used sparingly, and the word should not be emphasized either by underlining it, be made BOLD or writing it in CAPITAL form.
19. Create independent items. The answer to one item should not depend on the knowledge of the answer to a previous item.
20. The expected response should not be put at the beginning of the stem.
21. Read through all items carefully to ensure that the answer to one question is not revealed in another.

## ESSAY-TYPE TESTS

### Description:

An essay type test is a test that gives freedom to the respondent to compose his own response using his own words. The tests consist of relatively few items but each item demands an extended response.

There are **two types** of essay-tests. These are the

- **restricted response**
- **extended response.**

❖ **The restricted response type** limits both the content and the form that the student's answer may take. The respondent is also limited to a *specified length and scope* of the response.

For example, "in not more than 200 words, explain the causes of the COVID-19 outbreak in Ghana".

❖ **The extended response type** does not limit the student in the *form and scope of the expected answer*.

For example, "Discuss the factors that led to an outbreak of Cholera in a community".

### Advantages of Essay Tests

1. They provide the respondent with *freedom* to organize his own ideas and respond within unrestricted limits.
2. They are easy to prepare.
3. They **eliminate guessing** on the part of the respondents.
4. Skills such as the ability to organise material and ability to write and arrive at conclusions are improved.
5. They encourage good **study habits** as respondents learn materials in wholes.
6. They are best suited for testing **higher-order** behaviours and mental processes such as analysis, synthesis/creativity, and evaluation.
7. Little **time** is required to write the test items.

8. They are practical for testing a ***small number of students***.

### **Disadvantages of Essay Tests**

1. They are difficult to score objectively.
2. They provide opportunities for bluffing (i.e., where students write irrelevant and unnecessary material).
3. Limited aspects of student's knowledge are measured as students respond to few items only.
4. The items are an inadequate sample of subject content. Several content areas are omitted.
5. A premium is placed on writing. Students who write faster, all things being equal are expected to score higher marks.
6. They are time-consuming to both the teacher who scores the responses and the student who writes the responses.
7. They are susceptible to the **halo effect** where the scoring is influenced by extraneous factors such as the relationship between *scorer* and *respondent*.
8. A critical reader as well as a competent scorer can only effectively score responses.

## Guidelines in constructing good classroom essay tests

**1. Plan the test.** Give adequate time and thought to the preparation of the test items. The test items must be constructed from a Test Specification Table and well in advance (at least two weeks) of the testing date. This allows editing to be done.

**2.** The items should be based on *novel situations* and *problems*. Be *original*. Do not copy directly from textbooks or colleagues/others' past test items.

**3.** Test items should require the students to show adequate command of **essential knowledge**. The items should not measure rote memorization of facts, definitions and theorems but must be restricted to the measuring of *higher mental processes* such as application, analysis, synthesis and evaluation.

Examples of items include:

- **Application:** You oversee a youth camp of 100 campers. Prepare a menu chart which shows a balanced diet taking into consideration cost and nutritional value.  
Here, the student uses knowledge learnt in school to deal with a concrete situation.
- **Analysis:** A Form 1 student girl was severely and unfairly punished. Describe the feelings such treatment aroused in her.
- **Evaluation:** Evaluate the function of the United Nations Organization as a promoter of world peace.
- **Creativity/Synthesis:** You are the financial secretary of a society aimed at raising money to build a fishpond in your community. Plan and describe a promotional campaign for raising the money.

**4.** The length of the response and the difficulty level of items should be adapted to the *maturity level of students (age and educational level)*. An item like: "Discuss the implications of the Addis Ababa Convention on the economy of Ghana" would be too difficult for a first-year senior secondary school student.

**5.** Optional items should not be provided when content is relevant. They may be necessary only for large external examinations and when the purpose of the test is to measure writing effectiveness. If students answer different questions, an analysis of the performances on the test items is difficult.

**6.** All items should be of equal difficulty if students are to select from a given number of items.

**7.** Prepare a scoring key (marking scheme) at the time the item is prepared. Decide in advance what factors will be considered in evaluating an essay response. Determine the points to be included and the weights to be

assigned for each point. The preparation of a model answer will help disclose ambiguities in an item.

8. Establish a framework and specify the limits of the problem so that the student knows exactly what to do.

The following item for example does not give any framework for the student to operate:

Write **brief notes** on the following:

- a. United Nations Organization (UNO)
- b. African Union (AU)
- c. European Union (EU)

9. Present the student with a problem which is **carefully worded** so that only ONE interpretation is possible. The questions/items must not be ambiguous or vague.

For example: Family Planning in Ghana is a "**mixed bag**". Discuss Deferent interpretations could be given to the term '-mixed bag' if it was not mentioned.

10. Indicate the **value** of the question and the time to be spent in answering it.

11. Structure the test item such that it will elicit the type of behaviour you really want

12. The test items must be based on *the instructional objectives/outcomes/content standards/indicators* for each content unit.

13. Give preference to many items that require brief answers. These provide a broader sampling of subject content and thus better than a few items that require extended responses.

14. Statements and sub-questions for each item should be **clearly related**. For example

15. Start essay test items with words that are clear and as simple as possible, and which requires the student to respond to the stimulus expected. Avoid words such as: **what, list, who**, as much as possible

### **Commonly used words to start essay questions**

1. **Explain:** to make plain or clear; to make known in detail. To tell what an activity/process is and how it works and why it works the way it works
2. **Describe:** to tell or depict (a picture) in written words.
3. **Analyze:** to determine elements or essential features; examine in detail to identify causes, key factors, possible results.
4. **Assess:** to estimate or judge the value, character etc of
5. **Examine:** to inspect or scrutinize carefully; to inquire into or investigate
6. **Discuss:** to consider or examine by argument, or comment; give points for and against the content of the question



7. **Evaluate:** to judge or determine the significance worth or quality of. Involves discussion and making a judgment.
8. **Give an account of:** to describe a process/activity, and giving reasons, causes, effects etc

### Scoring essay tests

Essay tests can be scored by using the **analytic scoring** rubrics (also known as the point-score method) **holistic scoring** rubrics (also called **global-quality** scaling or rating method).

In analytic scoring, the main elements of the ideal answer are identified, and points awarded to each element. This works best on restricted response essays.

In analytic scoring, the main elements of the ideal answer are identified, and points awarded to each element. This works best on restricted response essays.

### Principles for scoring essay tests

1. Prepare a form of **scoring guide**, either an analytic scoring rubric or a holistic scoring rubric.
2. Test must be kept anonymous as possible. Score without knowing the one whose paper is being scored. This reduces the **halo effect**.
3. Grade the responses item by item and not script by script. Score all responses to each item before moving to the next item. This reduces the **CARRY OVER EFFECT**. The carryover occurs when the mark for a question is influenced by the performance on the previous questions.
4. Keep scores of previously graded items out of sight when evaluating the rest of the items.
5. Periodically **rescore** previously scored papers.
- 6 Before starting to score each set of items the script should be **shuffled**.
7. Score the essay test when you are **physically sound, mentally alert** and in an environment with very little or no distraction.
8. **Constantly follow the scoring guide** as you score. This reduces the *Rater Drift* which is the tendency to either not paying attention to the scoring guide (marking scheme) over time or interpreting it differently as time passes.
9. Score a particular question on all papers at one sitting. Break when fatigue sets in.
10. *Arrange for an independent scoring* of the responses or at least a sample of them where grading decision is crucial.
11. Comments could be provided, and errors corrected on the scripts for class tests to facilitate
12. Avoid being influenced by the *first few papers* read. These could make you either too harsh or too lenient.
- 13 The mechanics of writing such as *correct grammar usage*, **paragraphing**, flow of expression, quality of handwriting, orderly presentation of material and spelling should be judged separately from the content.

## PERFORMANCE-BASED ASSESSMENT

*performance assessment as testing that requires a student to create an answer or a product that demonstrates his or her knowledge or skills.*

Examples of performance assessments include:

- **Individual or Group projects** enabling several students to work together on a complex problem that requires *planning, research, internal discussion*, and group or individual presentation.
- **Essays** assessing students' understanding of a subject through a written description, analysis, explanation, or summary.
- **Experiments** testing how well students understand scientific concepts and can carry out scientific processes.
- **Demonstrations** giving students opportunities to show their mastery of subject-area content and procedures.
- **Portfolios** allowing students to provide a broad portrait of their performance through files that contain collections of students' works assembled over time.

### Some Characteristics of Performance-Based Assessment

1. Performance-based assessment is **authentic**, connected to *everyday life*.
2. Performance-based assessment provides opportunities for students to **show what they can do as well as what they know**.
3. Performance-based assessment **involves students in the process of evaluation**.
4. Performance-based assessment **integrates tasks involving multiple skills and knowledge of culture**.
5. Performance-based assessment helps all learners to **do their best and aims to improve overall student performance**.

## UNIT 5

### GOALS AND LEARNING TARGETS OF INSTRUCTION

#### What are learning outcomes /standards/indicators?

*A learning outcome is the knowledge, skill, or behaviour that a student is expected to exhibit after a period of study.*

Learning outcomes reflect a nation's concern with the level of knowledge acquisition among its student population. Measuring learning outcomes provides information on what knowledge (cognitive), skill or behavior (affective) students have gained after instruction is completed.

#### Definition of terms

*Instructional Objective/learning indicators:* A stated desirable outcome of education or an intended learning outcome in terms of the types of performance students can demonstrate at the end of instruction to show that they have learned what was expected of them. By the end of the lesson, students should be able to define the term, taxonomy.

*Behavioural objectives:* A statement that specifies what observable performance the learner should be engaged in when the achievement of the objective is evaluated. Behavioural objectives require action verbs such as discusses, write, read, state.

*Learning objectives/Content standards:* These specify what the students should do, value, or feel at the completion of an instructional segment.

#### Importance of learning objectives (targets) for classroom assessment

1. Learning objectives/Content standards/indicators make the general planning for an assessment procedure easier through the knowledge of specific outcomes.
2. The selection, designing and construction of assessment instruments depend on knowing which specific outcome should be assessed.
3. Evaluating an existing assessment instrument becomes easier when specific outcomes are known.
4. They help to judge the content relevance of an assessment procedure. Specific learning outcomes provide information for the judgments.

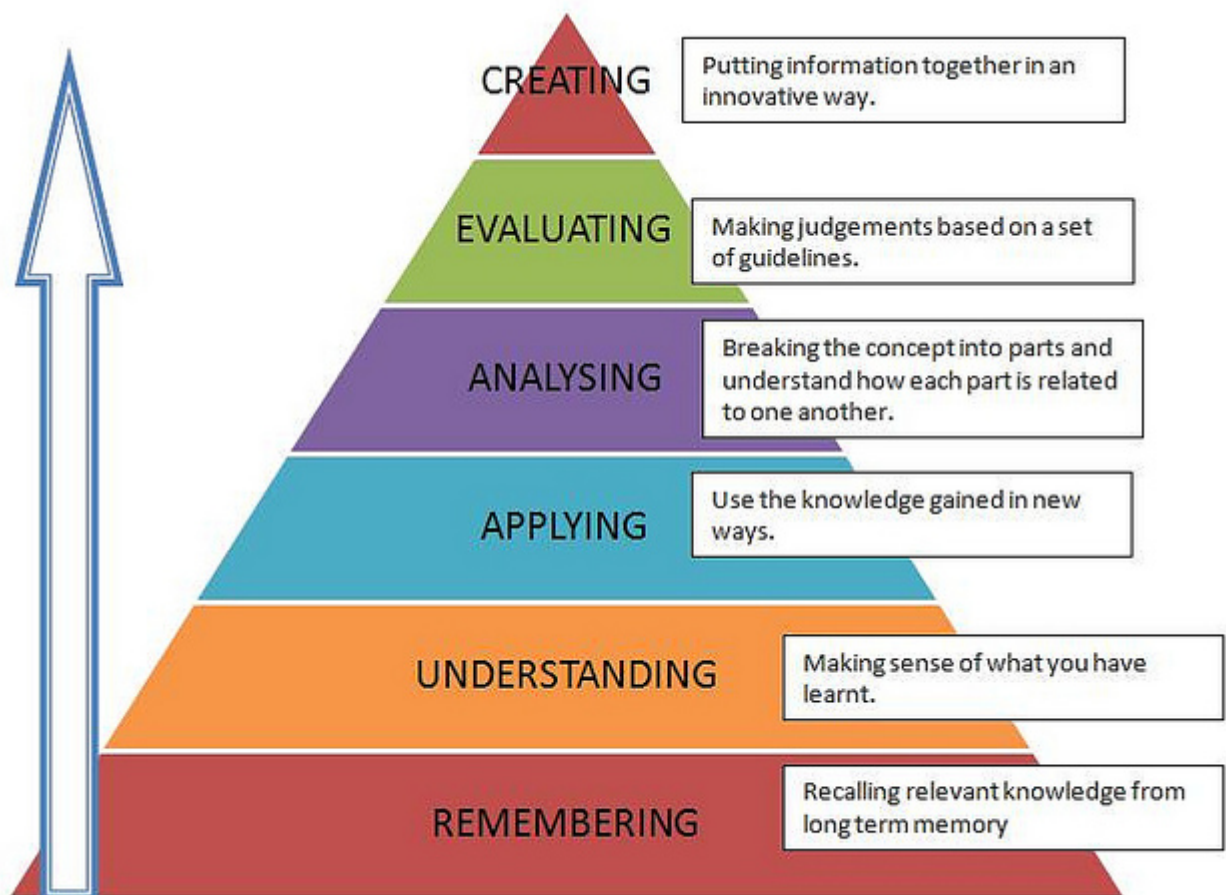
## TAXONOMIES OF EDUCATIONAL OBJECTIVES

Taxonomies are hierarchical schemes for classifying learning objectives into various levels of complexity. There are three main domains of educational objectives. These are (1) Cognitive, (2) Affective, (3) Psychomotor

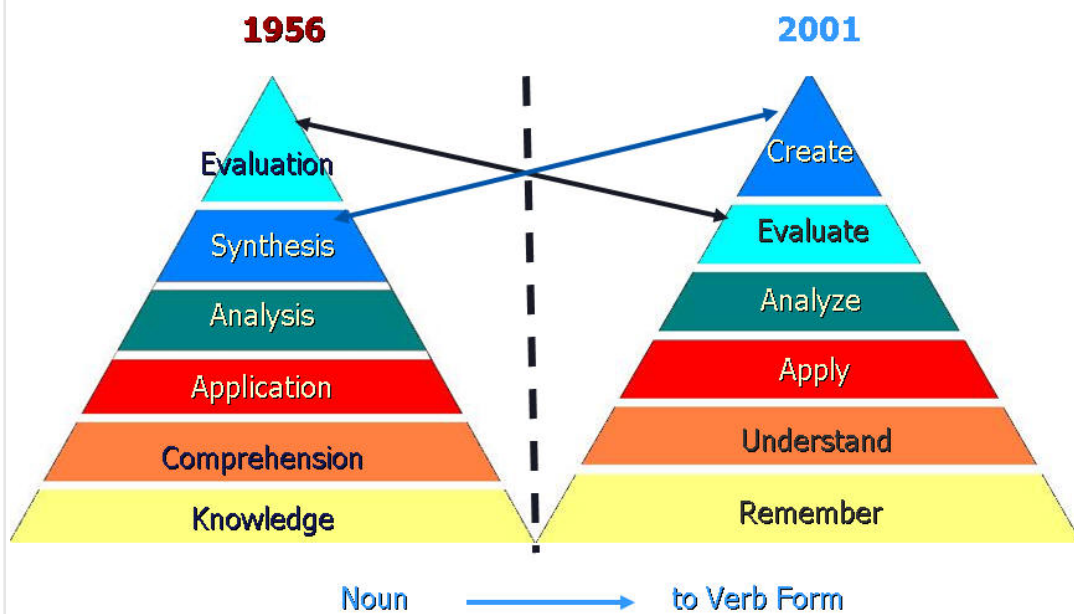
Cognitive domain objectives produce outcomes that focus on knowledge and abilities requiring memory, thinking, and reasoning processes.

Affective domain objectives produce outcomes that focus on feelings, interests, attitudes, dispositions and emotional states.

Psychomotor domain objectives produce outcomes that focus on motor skills and perceptual processes.



## Changes to Bloom's



### Higher order thinking skills

Creating	making, designing, constructing, planning, producing, inventing,
Evaluating	checking, hypothesizing, experimenting, judging, testing, monitoring
Analyzing	comparing, organizing, outlining, finding, structuring, integrating
Applying	implementing, carrying out, using
Understanding	comparing, explaining, classifying, exemplifying, summarizing
Remembering	recognizing, listing, describing, identifying, retrieving, naming, finding, defining

### Lower order thinking skills

## **Example and Key Words (verbs)**

**Knowledge:** Recall data or information.

Examples: Recite a policy or formulae. Quote prices from memory to a customer. Know the safety rules.

*Key Words:* An individual defines, describes, identifies, knows, labels, lists, matches, names, outlines, recalls, recognises, reproduces, selects, states.

**Comprehension:** Understand the meaning, translation, interpolation, and interpretation of instructions and problems. State a problem in one's own words.

Examples: Rewrites the principles of test writing. Explain in one's own words the steps for performing a complex task. Translate an equation into a computer spreadsheet.

*Key Words:* An individual comprehends, converts, defends, distinguishes, estimates, explains, extends, generalizes, gives an example, infers, interprets, paraphrases, predicts, rewrites, summarizes, translates.

**Application:** Use a concept in a new situation or unprompted use of an abstraction. Applies what was learned in the classroom into novel situations in the workplace.

Examples: Use a manual to calculate an employee's vacation time. Apply laws of statistics to evaluate the reliability of a written test.

*Key Words:* An individual applies, changes, computes, constructs, demonstrates, discovers, manipulates, modifies, operates, predicts, prepares, produces, relates, shows, solves, uses.

**Analysis:** Separates material or concepts into component parts so that its organizational structure may be understood. The student distinguishes between facts and inferences.

Examples: Troubleshoot a piece of equipment by using logical deduction. Recognize logical fallacies in reasoning. Gathers information from a department and selects the required tasks for training.

*Key Words:* An individual analyzes, breaks down, compares, contrasts, diagrams, deconstructs, differentiates, discriminates, distinguishes, identifies, illustrates, infers, outlines, relates, selects, separates.

**Synthesis:** Builds a structure or pattern from diverse elements. Put parts together to form a whole, with emphasis on creating a new meaning or structure.

Examples: Write a company operations or process manual. Design a machine to perform a specific task. Integrates training from several sources to solve a problem. Revises and process to improve the outcome.

Key Words: An individual categorizes, combines, compiles, composes, creates, devises, designs, explains, generates, modifies, organizes, plans, rearranges, reconstructs, relates, reorganizes, revises, rewrites, summarizes, tells, writes.

**Evaluation:** Make judgments about the value of ideas or materials.

Examples: Select the most effective solution. Hire the most qualified candidate. Explain and justify a new budget.

Key Words: An individual appraises, compares, concludes, contrasts, criticizes, critiques, defends, describes, discriminates, evaluates, explains, interprets, justifies, relates, summarizes, supports.

## UNIT 6 CHARACTERISTICS OF A GOOD TEST

Test validity and reliability

### Test Validity

**Definition:** Validity is the “soundness of the *interpretations* and *use* of students' assessment results". Validity emphasizes the interpretations and use of the results and **not the assessment instruments/tests**. Evidence needs to be provided that the interpretations and use are appropriate.

### Nature of Validity:

In using the term validity in relation to **there are** five points have to be borne in mind.

1. Validity refers to the ***appropriateness of the interpretations of the results*** of an assessment procedure for a group of individuals. It does not refer to the procedure or instrument itself.
2. Validity is a ***matter of degree***. Results have different levels of validity for different purposes and for different situations. Assessment results may have **high, moderate, or low** validity.
3. Validity is always specific to some particular use or interpretations. No assessment is valid for ***all purposes***.
4. Validity is a *unitary concept* that is based on various kinds of evidence.
5. Validity involves an **overall evaluative judgment**. Several types of validity evidence should be studied and combined.

### Categories of Validity evidence

There are 3 major categories of validity evidence.

1. ***Content-related evidence***
2. ***Criterion-related evidence***
3. ***Construct-related evidence***

### Factors affecting validity.

1. Unclear directions (or Rubrics). Validity reduces if students do not clearly understand how to respond to the items and how to record the responses or the amount of time available.
2. Too difficult reading vocabulary and sentence structure tends to reduce validity. The assessment may be measuring reading comprehension which is not to be measured.
3. Ambiguous statements in assessment tasks and items. This confuses students and makes way for different interpretations thus reducing validity.
4. Inadequate time limits. This does not provide students with enough time to respond and thus may perform below their level of achievement. This reduces validity.
5. Inappropriate level of difficulty of the test items. Items that are too easy or too difficult does not provide high validity.



6. Poorly constructed test items. These items may provide unintentional clues which may cause students to perform above their actual level of achievement. This lowers validity.
7. Test items being inappropriate for the outcomes being measured lowers validity.
8. Test being too short. If a test is too short, it does not provide a representative sample of the performance being interest in and this lowers validity.
9. Improper arrangement of items. Placing difficult items in the beginning of the test may put some students off and cause them to become unstable thereby performing below their level performance thus reducing validity.
10. Identifiable pattern answers. Placing the answers to test like multiple-choice and true/false types enables students to guess the correct answers more easily and this lowers validity.
11. Cheating, when students cheat by copying answers or helping friends with answers to test items, validity is reduced.
12. Unreliable scoring. Scoring of test items especially essay tests may lower reliability if they are not scored reliably.
13. Student emotional disturbances. These interfere with their performance thus reducing validity.
14. Fear of the assessment situations. Students can be frightened by the assessment situation and are unable to perform normally. This reduces their actual level of performance and consequently, lowers validity.

### **Test Reliability**

Reliability is the degree of consistency of assessment results.

It is the degree to which assessment results are the same when:

- (1) the same tasks are completed on two different occasions
- (2) different but equivalent tasks are completed on the same or different occasions, and
- (3) two or more raters mark performance on the same tasks.

### **Methods of estimating reliability**

1. **Test-retest method.**
2. Equivalent forms method.
3. Split –half method.

**Factors influencing reliability.**

1. Test length.
2. Group variability.
3. Difficulty of items.
4. Scoring objectively.
5. Speed of the Tests,
6. Sole making.
7. Testing conditions.

IOE.TUC.MOF

UNIT 7  
**PLANNING CLASSROOM TESTS AND ASSESSMENTS**

**Definition**

Achievement tests are tests that measure the extent of present knowledge and skills. In achievement testing, test takers are given the opportunity to demonstrate their acquired knowledge and skills in specific learning situations.

**Types of achievement tests**

There are two types of achievement tests.

(1) standardized achievement tests

(2) teacher-made/classroom achievement tests.

The major difference between these two types of tests is that standardized achievement tests are carefully constructed by test experts with specific directions for administering and scoring the tests. This makes it possible for standardized achievement tests to be administered to individuals in different places often at the same time.

**Characteristics of standardized achievement tests**

1. Standardized specific instructions are provided for test administration and scoring.  
Directions are so precise and uniform that the procedures are standard for different users of the test.
2. The test items are developed by test experts and specialists who follow well-formulated procedures for test development. The tests are thus of high quality. Reliability is often over 0.90.
3. They use test norms which are based on national samples of students in the classes/forms where the tests are intended for use.
4. Test content is determined by curriculum and subject-matter experts and involves extensive investigations of existing syllabi, textbooks, and programs.
5. Equivalent and comparable forms of the test are usually provided and administered.
6. A test manual is available as a guide for test administration and scoring. It provides information for evaluating the test for technical quality and interpretation and use of the results.
7. They are useful for measuring broader curriculum objectives and for school, district, regional and national comparisons.

## **TEACHER-MADE/CLASSROOM ACHIEVEMENT TESTS**

These tests are constructed by classroom teachers for specific uses in each classroom and are closely related to objectives. They are usually tailored to fit the teacher's instructional objectives. The content of the test is determined by the classroom teacher. The quality of the test is often unknown but usually lower than standardized tests.

### **Stages involved in teacher made/classroom achievement testing.**

The main goal of classroom assessment is to obtain valid, reliable, and useful information *concerning student achievement*. It is therefore important that good and quality tests and assessment tasks are constructed. Four principal stages are involved in classroom testing.

These are:

1. Constructing the test,
2. Administering the test,
3. Scoring the test,
4. Analyzing the test results.

#### **Stage 1: Constructing the test.**

There are eight steps in the construction of a good classroom test.

##### **Step 1: *Define the purpose of the test***

The basic question to answer is, "Why am I testing?". The tests must be related to the teacher's classroom instructional objectives. Several purposes are served by classroom tests and the teacher has to be clear on the purpose of the test. The teacher has to answer other questions such as 'Why is the test being given at this time in the course?', 'Who will take the test?', 'Have the test takers been informed?', 'How will the scores be used?'

##### **Step 2: *Determine the item format to use.***

Test items could either be essay, objective, -or performance, types. Objective-type tests include multiple-choice, short-answer, matching and true and false. The choice of format must be appropriate for testing topics and objectives. It is sometimes necessary to use more than one format in a single test.

Mehrens and Lehmann (1991) mentioned **eight (8)** factors to consider in the choice of the appropriate format. These include:

- 1) The purpose of the test
- 2) The time available to prepare and score the test,
- 3) The number of students to be tested,
- 4) Skill/outcome/standard to be tested,
- 5) Difficulty desired,
- 6) Physical facilities like reproduction materials,
- 7) Age of learners,
- 8) Skills in writing the different types of items.

**Step 3: Determine what is to be tested.**

The teacher asks himself or herself the question, 'What is it that I wish to assess?' The teacher must determine what the strands and sub-strands the test will cover as well as what knowledge, skills, and attitudes to measure. Instructional objectives or indicators must be defined in terms of student behaviour and linked to what has been stressed in class. A test plan made up of a table of specifications or blueprint must be made. The **Specification Table** matches the course content with the instructional objectives.

To prepare the table, specific topics and sub-topics covered during the instructional period are listed. The major course objectives or content standards are also specified, and the instructional objectives defined. The total numbers of test items are then distributed among the course content and instructional objectives or behaviours.

## **UNIT 8: ASSEMBLING, ADMINISTERING AND APPRAISING ACHIEVEMENT TESTS**

### **Guidelines for Assembling Achievement Tests**

#### **1. Review test items and assessment tasks.**

Where possible, allow fellow teachers or colleagues can review the test items or tasks. The following points should be considered.

- i. Test format should be appropriate for the learning outcome being measured.
- ii. Knowledge, understanding or thinking skill required by the item or task should match the specific learning outcome and subject-matter content being measured.
- iii. The item or task should not be excessively wordy.
- iv. The point of the item or task as well as the desired response should be clear.
- v. A scoring rubric or scoring guide should be available.
- vi. The item or task should be free from technical errors and irrelevant clues.
  - i. The item or task should be free from racial, ethnic and gender bias.

2. Decide on the total number of items and the length of time to constitute the test. The number of items and time to be used is dependent on institutional policy, number of credits for a course, test format, item difficulty level, examiner's experience and students' maturity level.

3. Test items should be typed or written neatly. Writing items on the chalkboard or dictating them must be done with utmost care since it may cause problems for students especially those with visual, listening comprehension or hearing problems.

#### **4. Arranging test items.**

- Items should be sequenced (especially objective-type tests) such that they appear in the order of difficulty with the easiest ones placed first.
- Items should also be arranged in sections by item-type. The sections should progress from easier formats to more difficult formats. Within each section, group items such that the easier ones come first. For example, all true-false items should be grouped together, then all matching items and so on.
- Items can also be arranged according to the order in which they are taught in class or the order in which the content appeared in the textbook.
- Sequencing is not necessary for essay-type tests where optional choices are made. All items of this nature should however equal difficulty levels.

5. Provide directions to students. Directions should include the amount of time allowed to complete the test, where and how answers should be written, number of points for each test item, what should be done about guessing (on selection-type items). Each item format should have a specific set of directions.

#### **6. Reproducing the test.**

- Items must be spaced and arranged so that they can be read and scored (for objective-type tests) with the least amount of difficulty. Cramming too many tests on to a page is poor economy.
- Multiple-choice items should have the alternatives listed vertically below the stem of the item rather than across the page.
- Items should not be split with parts of the item on two different pages. All items should be numbered consecutively.
- All illustrative material should be clear, legible and accurate.
- Proofread the entire test or assessment before it is finally reproduced.

### **Guidelines in Administering Achievement Tests**

1. Prepare students for the test. The following information is essential to students' maximum performance.
  - ✓ When the test will be given (date and time).
  - ✓ Under what conditions it will be given (timed or take-home, number of items, open book or closed book, place of test).
  - ✓ The content areas it will cover (study questions or a list of learning targets).
  - ✓ Emphasis or weighting of content areas (value in points).
  - ✓ The kinds of items on the test (objective-types or essay-type tests).
  - ✓ How the assessment will be scored and graded.
  - ✓ The importance of the results of the test.
2. Students must be made aware of the rules and regulations covering the conduct of the test. Penalties for malpractice such as cheating should be clearly spelt out and clearly adhered to.
3. Avoid giving tests immediately before or after a long vacation, holidays, or other important events where all students are actively involved physically or psychologically/emotionally.
4. Avoid giving tests when students would normally be doing something pleasant e.g., having lunch etc.
5. The sitting arrangement must allow enough space so that pupils will not copy each other's work.
6. Adequate ventilation and lighting is expected in the testing room.
7. Provision must be made for extra answer sheets and writing materials.
8. Pupils should start the test promptly and stop on time.
9. Announcements must be made about the time at regular intervals. Time left for the completion of the test should be written on the board where practicable.
10. Invigilators are expected to stand a point where they could view all students. They should once in a while move among the pupils to check on malpractices. Such movements should not disturb the pupils. He/she must be vigilant.
11. Invigilators should not be allowed to read novels, newspapers, grade papers or receive calls on mobile phones.
12. Threatening behaviours should be avoided by the invigilators. Speeches like 'If you don't write fast, you will fail' are threatening. Pupils should be made to feel at ease.

13. The testing environment should be free from distractions. Noise should be kept to a very low level if it cannot be eliminated or removed. Interruptions within and outside the classroom should be reduced. It is helpful to hang a "DO NOT DISTURB - TESTING IN PROGRESS" sign at the door.

14. Test anxiety should be minimized.

- Things that create excessive anxiety are (1) warning students to do their best 'because the test is important', (2) telling students that they must work fast to finish on time, (3) threatening dire consequences if they fail, and (4) threatening students with tests if they do not behave.

- Teachers and invigilators should not walk around looking over students' shoulders while they are responding to assessments.

- Before assessments, teachers should convey a sense of confidence about student's performance in the upcoming assessment.

15. Do not talk unnecessarily before letting students start working. Remarks should be kept to a minimum and related to the test.

16. Avoid giving hints to students who ask about individual items. Where an item is ambiguous, it should be clarified for the entire group.

17. Expect and prepare for emergencies. Emergencies might include shortages of answer booklets, question papers, power outages, illness etc.

### **Appraising Achievement Tests (Item Analysis)**

Item analysis is the process of collecting, summarizing, and using information from students' responses to make decisions about each test item. It is designed to answer the following questions:

1. Did the item function as intended?
2. Were the test items of appropriate difficulty?
3. Were the test items free of irrelevant clues and other defects?
4. Was each of the distracters effective (in multiple-choice items)?

### **Benefits of item analysis**

1 It helps to determine whether an item functions as intended. It provides information on whether an item assesses the intended learning targets, whether it is of the appropriate level of difficulty or whether it distinguishes between high achievers and low achievers and whether the options are working.

2 Item analysis data provide a basis for efficient class discussion of the test results. Difficult items can be identified and discussed. Misinformation and misunderstanding of distracters can be corrected.

3. Item analysis provides feedback to the teacher about pupil difficulties. It brings to light general areas of weakness that require more attention.

4 Item analysis data provide a basis for the general improvement of classroom instruction. It assists in evaluating the appropriateness of the learning outcomes and course content of the students being taught.

5 Item analysis procedures provide a basis for increased skill in test construction. Item analysis reveals ambiguities, clues, ineffective distracters, and other technical defects that were missed during the test preparation. Information revealed provides experience for future writing of tests.

6. It helps to create item banks for use in future tests.



## UNIT 9 INTERPRETATION OF TEST SCORES

Scores obtained in classroom quizzes, tests and examinations are known as raw scores. They give truly little information about the performance or achievement of a student. For example, if John obtained 48 in a test, it is difficult to know his level of performance unless more information is provided. Such types of information include the maximum/total score, mean or median score the variability of the group, the difficulty level of the items, the number of test questions and the amount of time allowed for the test. To interpret and obtain meaning from the scores, they need to be referenced or transformed into other scores.

There are two popular ways of interpreting test scores so the meaning can be derived from the scores. These are:

1. Norm-referenced Interpretation
2. Criterion-referenced Interpretation

### **Norm-referenced interpretation**

These describe test scores or performance in terms of a student's position in a reference group that has been administered the assessment. In other words, it compares and individual's performance with others in the group who have taken the same test. The reference group is called the norm group.

In the earlier example, John's score of 48 can be compared with the mean score for the class. If the mean score is 40, then one could say that John's performance was above average. If the median score is 40, then one could also say that John's performance could be placed in the upper half of the class. The score of 40 can appropriately be called the norm and the class that provided the mean or median of 40 is called the norm group.

### **Types of norm-referenced scores**

The most popular norm-referenced scores are described below.

1. **Class raw score ranks.** Raw scores in a class are often ordered in merit form from the highest score (1st position) to the lowest score (last position). The ranks show how a student performs compared with the others in the group.
2. **Percentile and percentile ranks.** A percentile is a point in a distribution below which a certain percentage of the scores fall while a percentile rank is a person's relative position such that a given percentage of scores fall below the score obtained. If a raw score of 48 is the 60th percentile, it means that a student who obtains 48 in a test, has done better than 60 percent of all those in the group that took the test.
3. **Standard scores.** These are either linear Z or T scores. A linear Z scores is based on the normal distribution such that the mean is 0. Raw scores that are transformed to Z-scores use the formula:

$Z = \frac{x - \bar{x}}{s}$ , where  $x$  is the raw score,  $\bar{x}$  is the group mean and  $s$ , the group standard deviation. Negative values show that performance is below average and positive values mean that performance is above average.

T-scores are based on Z-scores and use the formula:  $T = 50 + 10Z$ . Scores above 50 show above average performance and scores below 50 show below average performance.

4. Stanines (Standard Nine). These are derived scores based on the normal distribution with a mean of 5 and standard deviation of 2. It uses the integers 1-9. The percentages of scores at each Stanine are: 9 (top 4%), 8 (next 7%), 7 (next 12%), 6 (17%) 5 (next 20%) 4 (next 17%) 3 (next 12%) 2 (next 7%) and 1 (lowest 4%).

### Uses of norm-referenced interpretations

1. **Selection decisions.** In selecting students for awards and prizes, a norm-referenced approach is often used. To award the prize for the best student, a ranking is done and the student at the top position is awarded the prize. Also the ranking of students allows the selection of those who meet a fixed quota. If an educational institution wants to admit 400 students out of a total of 600 based on an entrance examination, an order of merit is done and the top 400 students are selected.

2. **Comparison decisions.** Norm referenced interpretations allow the comparison of performance across subjects. For example performance in Mathematics and English can be compared by using Z and T scores. It may be necessary to compare performance between two classes, for example, Home Economics and Arts. Norm referenced scores provide the information needed for the comparisons. Mean or median scores would provide information as to which class performs better.

3. **Achievement testing.** Examination bodies such as the West African Examinations Council use norm-referenced scores sometimes in interpreting the results of students in examinations such as the BECE and WASSCE.

4. **Monitoring decisions.** Norm-referenced scores are useful in monitoring the general progress of individual students. A student who was at the 10th percentile in Mathematics in the first term but moved up to the 75th percentile in the third term has made much progress.

### Criterion-referenced interpretations

These describe test scores or performance in terms of the kinds of tasks a person with a given score can do. The performance can be compared to a pre-established standard or criterion.

For example a student may be able to solve 8 problems out of 10 concerning fractions. A level of performance can be established at 6. The criterion or standard can be used as a competency/mastery score, so that students who have obtained scores that are greater than 6 are termed competent or have mastered skills in a particular domain. Criterion-referenced interpretations generally indicate what an individual can or cannot do with respect to a specified domain of knowledge, attitudes or skills.

## **Types of criterion-referenced scores**

1. **Percent correct scores.** This is the percentage of items that a student got correct. For example, if a student obtained 8 marks out of 10, the percent correct is 80.
2. **Competency scores.** These are cut-off scores set to match acceptable performance. Students who obtained the cut-off scores are believed to have achieved a required level of competency. Cut-off scores should not be arbitrarily set. There should be a support or basis for them.
3. **Quality ratings.** This is the quality level at which a student performs a task. For example, a student can be rated as A for outstanding, B+ for excellent, etc.
4. **Speed of performance scores.** These indicate the amount of time a student uses to complete a task, or the number of tasks completed within a specified time. For example, a student may type 30 words in a minute, or an athlete may run 100 meters in 11.5 seconds.

## **Uses of criterion-referenced interpretations**

1. **Certification decisions.** Certificates are needed in several areas of work to demonstrate the acquisition of skills and knowledge. Criterion-referenced scores provide information about whether an applicant has the required level of skill or not and certificates of achievement attest to this.
2. **Minimum competency decisions.** Certain curricula are structured such that a student needs to achieve a certain level of competency before moving on to a higher level. Criterion-referenced scores are used to determine whether a student or a class can move on to a higher level of study.
3. **Diagnostic decisions.** Criterion-referenced scores help the teacher to discover the learning difficulties of the pupils. They help the teacher to analyze and know which topics or learning targets have not been grasped. It helps the teacher to provide individual or class learning activities that will best adapt to students' requirements and thereby maximize their opportunities to attain chosen learning targets.
4. **Placement decisions.** Criterion-referenced scores provide information as to whether a student can succeed in a programme or not. For example, to determine whether a person can be a medical doctor or not, a test can be given such that performance on the test can determine whether the individual has the pre-requisite skills to succeed in the medical programme.
5. **Programme evaluation.** Criterion-referenced scores provide information about national progress in education. The performance of students can indicate whether a particular curricular is successful in its implementation or not. In Ghana, criterion-referenced scores were used in the 1990s to assess the level of mathematics and English literacy.