
Module for Post-Graduate Diploma in Education Programme

EPS751A: ASSESSMENT IN EDUCATION

DR. RUTH ANNAN-BREW



REPUBLIC OF GHANA



INSTITUTE OF EDUCATION, UCC

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



Ministry of Finance



**Trade Union Congress
Coast**



University of Cape

DECEMBER, 2022

TABLE OF CONTENT

CONTENT	PAGE
ASSESSMENT PRACTICES IN SCHOOLS	1
UNIT 1: GOALS AND LEARNING OBJECTIVES OF INSTRUCTION	2
UNIT 2: PRINCIPLES OF EVALUATION	12
UNIT 3: PURPOSE OF ASSESSMENT	18
UNIT 4: TEST CONSTRUCTION, ADMINISTRATION, SCORING, AND GIVING FEEDBACK	28

ASSESSMENT PRACTICES IN SCHOOLS

This course focuses on the key ideas, principles, and practices regarding assessment in schools. It explains the principles of assessment which must be translated into practice in our classrooms to ensure that assessments are fit for purpose, meaningful, and intended to support learner learning and attainment. Assessment is one of the most important features in education and if not done well can lead to low achievement rates at the school level and attainment of needed qualifications to progress to the next stages of learning.

The course is organised into four main units which will be delivered over a four-week period as follows:

- The Nature, Principles and Qualities of Assessments
- Purpose of Assessment
- Forms of assessments in schools
- Test design, Administering, Scoring and Giving Feedback

The units cover general assessment practices which can be applied at any pre-tertiary level

Content Learning Outcomes

- Describe the nature of assessment and apply the principles of assessment in designing assessment tools/tasks
- Select appropriate assessment tools/formats for different forms of assessment based on the purpose of assessment.
- Evaluate the extent to which an assessment tool meets the qualities of good assessments
- Write, score, and critique test items that match given learning outcomes and give effective feedback

UNIT 1: GOALS AND LEARNING OBJECTIVES OF INSTRUCTION

Introduction

The assessment literature is full of terminologies such as *mission, goals, objectives, and outcomes*. However, there is lacking in a consensus on the precise meaning of each of these terms. Part of the difficulty stems from changes in approaches to education. For instance, shifts from objective-based, to competency-based, and to outcomes-based, education have taken place over the years with various champions of each espousing the benefits of using a different point of view.

Learning outcomes

After studying this session, you will be able to:

- a. explain the meaning of instructional goals.
- b. explain the meaning of instructional objectives.
- c. differentiate between instructional goals and objectives.

Concept of Instructional Goals and Objectives

- Let us begin with this: ask yourself this; what do I already know about instructional goals and objectives?

Goals and Objectives are similar in that they describe the intended purposes and expected results of teaching and learning activities and establish the foundation for assessment. Goals are broad, generalised statements of that to be learned. You should think of them as a target to be reached.

- Think of this: write down any instructional goal you will want to achieve at the end of a unit on rate of reaction. If you have finish, check out this example;

The unit on rate is designed to help students to *understand that chemical reactions at different rates depending on the conditions set for the react to occur*.

- You may ask yourself; what is an objective then? Do have anything to say? Compare your response to the following.

Instructional objectives are brief, clear statements that describe the desired learning outcomes of instruction (that is the specific skills, values, and attitudes students should exhibit that reflect the broader goals. They are stated desirable outcome of education or an intended learning outcome in terms of the types of performance students are able to 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.

Instructional objectives are the foundation upon which lessons and assessment are built; which can be used to prove the instructional goals are met. Check out instructional objectives as the arrows you will shoot towards your target.

- Think of this: write down any instructional objective you will want to achieve at the end of a lesson on rate of reaction. If you have finish, check out this example;

At the end of the lesson, the student will be able to

- *deduce the zero order reactions from any given data and graphs*
- *calculate the rate of any chemical reaction.*

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 discuss, write, read, state.

Learning objectives: These specify what the teacher likes the students to do, value, or feel at the completion of an instructional segment.

Comparison of Instructional Goals and Objectives

- Think of this; is there any difference between instructional goals and objectives?

A goal is a broad statement of expected learning outcome of a course or subject. For example, at the end of the acids and bases unit, students will be able to: *understand the behaviour of acid and base solutions in relation to their physical and chemical properties.*

An instructional objective is a specific statement of observable behaviours that can be evaluated and contributes to reaching the instructional goal. For example, at the end of the lesson, the student will be able to: *discuss the general properties of acids and bases.*

A goal has many instructional objectives. For example, at the end of the plant structure and physiology course, students will be able to: *relate the structure of plant parts to their functions.*

The instructional objectives for this goal will include the following. By the end of the lesson, the student will be able to:

- *describe the external features of monocotyledonous and dicotyledonous plants;*
- *outline three functions each of the major parts of plants;*
- *distinguish between monocotyledonous and dicotyledonous plants;*
- *identify at least two functions of modified parts of plants;*
- *describe the internal structure of root, stem and leaf; and*
- *distinguish between the different types of tissues of the root, stem and leaf.*

Importance of learning objectives (targets) for classroom assessment

1. Learning objectives 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 judgment.

Importance of stating instructional

Learning outcomes

After studying this session, you will be able to:

- a. discuss five important instructional objectives to the teacher.
- b. discuss three important instructional objectives to the student.

Now read on ...

Think of this; are instructional goals and objectives important to classroom practices?

Importance of Instructional Goals and Objectives

The purpose of instructional objectives is not to restrict spontaneity or constrain the vision of education in the discipline. However, the purpose of objectives is to ensure that learning is focused clearly enough that both students and teacher know the direction of lessons, and so learning can be objectively measured. Different archers have different styles, so do different teachers. Thus, you can shoot your arrows (objectives) many ways. Thus, stating clear instructional objectives is important because objectives:

Teacher

1. provide the teacher with a solid foundation for designing relevant activities and assessment. Activities, assessments, and grading should be based on the objectives.
2. help the teacher to develop a learning object, course, lesson, or learning activity. You must determine what you want the students to learn and how you will know that they learned. Instructional objectives (also called behavioral objectives) are requirements for the high-quality development of instruction.
3. help teacher to identify critical and non-critical instructional elements.
4. help to remove teacher subjectivity from the instruction.
5. help teacher to design a series of interrelated instructional topics.
6. Help teacher in assessing student's performance and whether the lesson was effective.

Student

7. help students to better understand expectations and the link between expectations, teaching and grading. This will help in selecting the content of instruction.
8. help student by making them aware where they stand- self assessment.
9. tell student what is expected from them and at what level – promote student responsibility for own learning.
10. make clear to student what they can gain from a particular course/lecture.

In general, the stated instructional objectives are important to both teachers and students in the teaching and learning processes. Thus, the selection of content, development of instructional strategy, development and selection of instructional materials, and evaluation of instruction.

Taxonomies of educational objectives

Different instructional objectives are written by teachers for each teaching and learning process. It is expected that the instructional objectives address the variety of learning needs of students. This session will introduce you to the taxonomies of instructional objectives, specifically, Bloom's taxonomy and Quellmalz's taxonomy. Have fun as you learn this another important aspect of our job as teachers.

Learning outcomes

After studying this unit, you will be able to:

- a. identify the three types of objectives.
- b. discuss the levels of Bloom's taxonomy of writing instructional objectives.
- c. Describe the nature of the assessment
- d. Apply the principles of assessment in designing assessment tools
- e. Determine the quality of assessment results.

Now read on ...

To look at taxonomies of instructional objectives, first there is the need to consider types of instructional objectives.

- What do you already know about the types of objectives? Write it down and compare your guess with the following.

Types of Objectives

There are three types of learning objectives, which reflect different aspects of student learning:

Cognitive objectives: "What do you want your students to know?"

Affective objectives: "What do you want your students to feel or care about?"

Behavioural Objectives: "What do you want your students to be able to do?"

Objectives can also reflect different levels of learning as:

Mastery objectives are typically concerned with the minimum performance essentials. That is those learning tasks or skills that must be mastered before moving on to the next level of instruction.

Developmental objectives are concerned with more complex learning outcomes – those learning tasks on which students can be expected to demonstrate varying degrees of progress.

Instructional objectives describe in detail the behaviours that students will be able to perform at the conclusion of a unit of instruction such as a class, and the conditions and criteria which determine the acceptable level of performance.

Taxonomies of Instructional 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.

Educators and psychologists concerned with learning theory have given considerable thought to the various types of learning that take place in schools and colleges. Probably the most comprehensive and widely known analysis of instructional objectives is the Taxonomy of Educational Objectives by Benjamin Bloom and others.

Nature, Principles, and Qualities of Assessment

This session explains what we mean by assessment and the principles which guide good assessment practices. It is important that teachers and learners understand how the assessment process works and understand why assessment must provide meaningful data and information for improving learner outcomes

What is Assessment?

Assessment is the systematic collection of information about learner learning using the time, knowledge, expertise, and resources available to inform decisions about how to improve teaching and learning. Assessment can also be seen as the process of gathering and discussing information from multiple and diverse sources to develop a deep understanding of what students know, understand, and can do with their knowledge because of their educational experiences; the process culminates when assessment results are used to improve subsequent learning.

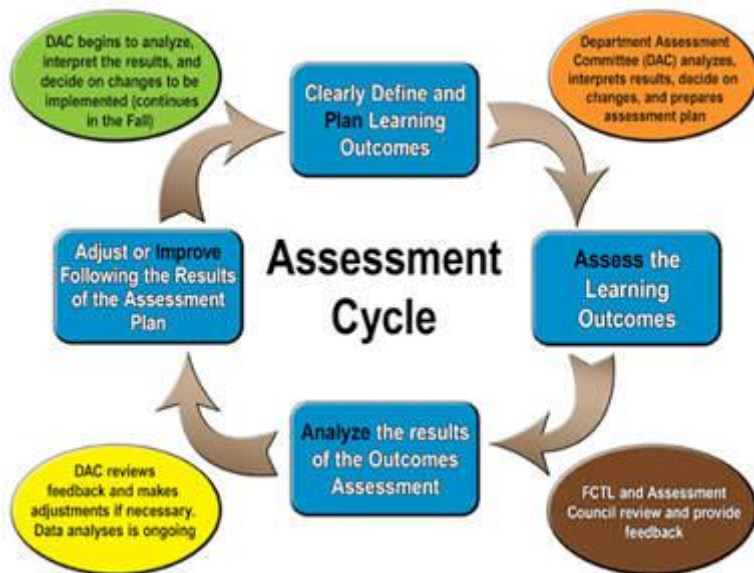


Figure 1: The Assessment Process

The Four Steps of the Assessment Cycle

- Step 1: Clearly define and identify the learning outcomes.
- Step 2: Select appropriate assessment measures and assess the learning outcomes. ...
- Step 3: Analyze the results of the outcomes assessed.
- Step 4: Adjust or improve programs following the results of the learning outcomes assessed.

The Classroom Assessment Process

Assessment involves learners, teachers, the school, parents, and the managers of the curriculum such as GES, NACCA, and WAEC. This is because the data collected whether formal or informal must be used either by the learner or teacher to determine the learning progression of the learner. Schools use this data in line with School-Based Assessment requirements and the feedback is given to learners and parents. The school must be able to meet national requirements for assessment.

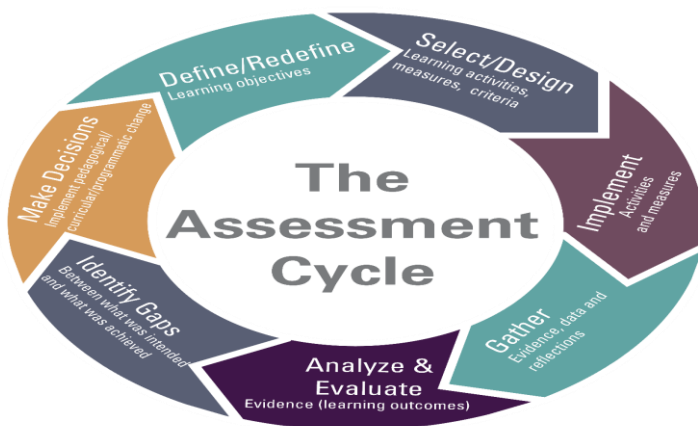


Figure 2: The Assessment Cycle

Essentially, the assessment process is to gather relevant information about student performance or progress, or to determine student interests to make judgments about their learning process.

Measurement, Assessment, and Evaluation

Evaluating learner learning or performance involves some sort of measurement using a range of assessment types that can be evaluated.

To explain this even further, think back to your own experiences as a learner. A teacher wants to check how well learners who have been taught antonyms and synonyms have gained the needed knowledge and understanding. The teacher decides to:

1. decides to test learners, using a series of worksheets over a period
2. assesses their need for reteaching or advancing to new learning based on the reliable test results
3. evaluates the effectiveness of the lessons by determining the degree to which learners can demonstrate knowledge and understanding in new contexts.

NB: These terms are usually associated with assessment. Figure 1 depicts these terms and their respective interpretations.

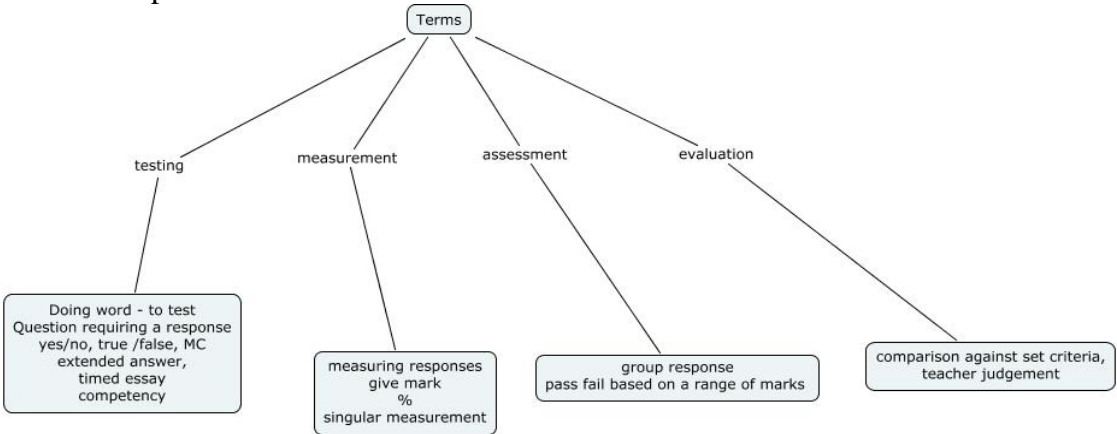


Figure 3: Interrelations among test, measurement, assessment, and evaluation.

TIPS:

This means that we should be careful about how we assign numbers/ values scores to the data we gather from our learners. To be able to assign these scores appropriately, we need to know the nature(scales) of the data, thus scales of measurement.

Relationship among Measurement, Assessment, and Evaluation

There's often a misconception about identifying the difference between assessment, evaluation, and measurement. Measurement, assessment, and evaluation mean very different things, and yet most of the learners were unable to adequately explain the differences

Measurement Is Always Numerical

Measurement refers to the process by which the attributes or dimensions of some physical object are determined. The verb measure means "to discover the exact size, amount, etc., of something, or to be of a particular size." So it means that measurement is how we rate and determine the performance of a student numerically.

- It is used to categorize and or quantify variables.
- It helps to quantify individual's achievement, personality, attitudes, habits, and skills.
- It involves the quantification of observable phenomena, and it is performed on the physical world by observer. Here, information collected about the characteristics and attributes of things can be determined and differentiated.

Measurement in education:

Today, measurement in education is much different and more advanced. With the development of many theories in education over time and assessment software systems that can be used in this process, various variables, related to students' marks and grades, are measured. These include:

- Intelligence, Interest, Students aptitude, and personality, Aims of education
- Effectiveness of the curriculum, Usefulness of teaching methods, Basis of educational policy
- The various educational activities of administrators and teachers as well

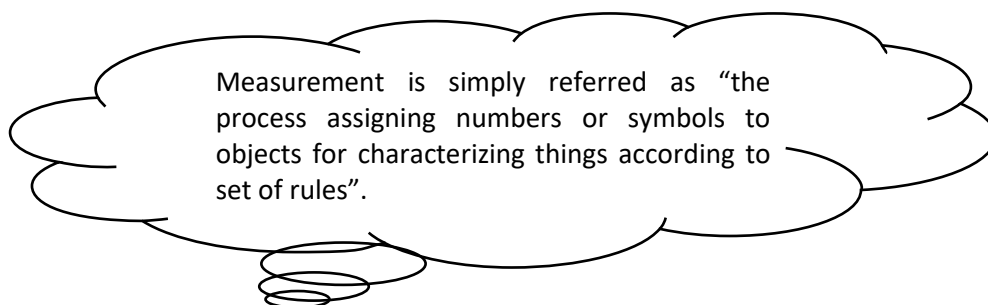
Types of measurement

Measurement can be of 3 types namely:

- Direct measurement
- Indirect measurement
- Relative measurement

External attributes or characteristics could be directly measured with high precision, using various tools and devices, Example height or length, Volume, and weight could be measured directly using appropriate tools. This type of measurement is known as **Direct measurement**. However, heat, light, and sound cannot be measured directly, but measured only through the effects they produce. Such kind of measurements is called **Indirect measurements**. In education and psychology, measurement is very complex. There is direct measurement is not possible. Students' achievements, comments, elegance, etc. are measured only by comparing them with those of their peers. These are called **Relative measurements**.

TIPS

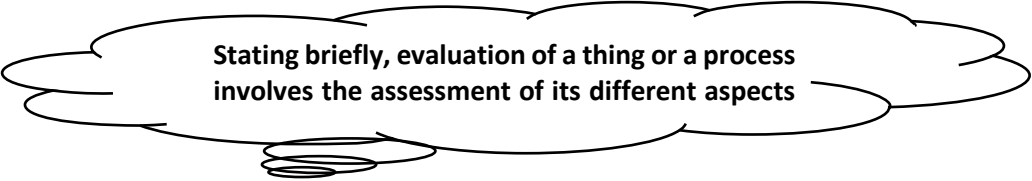


Assessment: Assessment Is the Detection and Analysis

Assessment is “a systematic procedure for collecting information that can be used to make inferences about the characteristics of people or objects. Assessment is referred as “a process of collecting evidence and making judgments relating to outcomes”. It is said that assessment has a narrower meaning than evaluation but a broader meaning than measurement. In its derivation, the word assess means “to sit beside” or “to assist the judge”, it reflects that assessment is a process of gathering data and fashioning them into an interpretable form; judgment can be made based on this assessment. It’s detecting, analysing, and interpreting students’ learning and progress. It is “the various methods used by educators to measure and document the academic achievement and skills of students during preschool adulthood. It is a process of inquiry to collect and synthesize evidence that concludes the status or quality of a program, product, person, policy, proposal, or plan.”

Evaluation: Evaluation Is Qualitative

Evaluation is when you start to interpret and judge the results of the data you’ve collected throughout the assessment process or elsewhere. It’s the phase for decision-making. Evaluation is “a systematic process of collecting and analyzing data in order to determine whether, and to what degree, objectives have been, or are being, achieved ” When we evaluate, what we are doing is engaging in some process that is designed to provide information that will help us make a judgment about a given situation. Evaluation is the process of determining the value of a thing or a process based on the assessment of its different aspects or components by employing observation, measurements, and tests. Evaluation includes the process of assessment and tools and methods of measurement.



Stating briefly, evaluation of a thing or a process involves the assessment of its different aspects

DISTINCTION AMONG MEASUREMENT, ASSESSMENT AND EVALUATION

Measurement	Assessment	Evaluation
Measurement is the method of giving score to learners after taking a test on any topic. E.g. 50 out of 100 in a Subject.	After analysing the score of measurement , the change in the syllabus, or enlargement or refinement is decided. This is called assessment.	The final judgement taken on the basis of assessment is called evaluation.
It is a very narrow concept	The concept of Assessment is broader than measurement.	And the concept of Evaluation is broadest among the three.
The first and foremost task of determining competency is carried out by Measurement.	Assessment is the middle stage of determining competency	Evaluation deals with the final task of determining competency among learners.

UNIT 2: PRINCIPLES OF EVALUATION

The following are some of the principles, which should be kept in view in evaluation.

1. Evaluation is a continuous process (continuity).
2. Evaluation should involve minimum possible costs (inexpensive).
3. Evaluation should be done without prejudice to day to day work (minimum hindrance to day to day work).
4. Evaluation must be done on a co-operative basis in which the entire staff and the board members should participate (total participation).
5. As far as possible, the agency should itself evaluate its program but occasionally outside evaluation machinery should also be made use of (external evaluation).
6. Total overall examination of the agency will reveal strength and weaknesses. (agency / program totality).
7. The result of evaluation should be shared with workers of the agency (sharing).

Figure 4: Basic Principles of Evaluation in Education

Principles of Assessment

Principles of assessment serve as guidelines to ensure that a test is useful, appropriate, effective, and credible. Since assessment is a crucial and integral part of learning and teaching, it is important that these general principles serve as guidance in determining the level of accomplishments of learners during the learning process and at the end of the learning. These principles, therefore, reinforce what we think is important in creation, selection, administering, measuring, assessing, and evaluation. The principles of assessment must therefore translate into our classroom practices. These principles include:

1. **Principle 1** - Assessment should be valid
 - Validity ensures that assessment tasks and associated criteria effectively measure learner attainment of the intended learning outcomes at the appropriate level.
2. **Principle 2** - Assessment should be reliable and consistent
 - There is a need for assessments to be reliable and this requires clear and consistent processes for the setting, marking, grading, and moderation of assignments.
3. **Principle 3** - Information about assessment should be explicit, accessible and transparent.
 - Clear, accurate, consistent, and timely information on assessment tasks and procedures should be made available to learners, staff, and other external assessors or examiners.
4. **Principle 4** - Assessment should be inclusive and equitable
 - As far as is possible without compromising academic standards, inclusive and equitable assessment should ensure that tasks and procedures do not disadvantage any group or individual.
5. **Principle 5** - Assessment should be an integral part of programme design and

- should relate directly to the programme aims and learning outcomes.
- Assessment tasks should primarily reflect the nature of the discipline or subject but should also ensure that learners can develop a range of generic skills and capabilities.
6. **Principle 6** - The amount of assessed work should be manageable
 - The scheduling of assignments and the amount of assessed work required should provide a reliable and valid profile of achievement without overloading staff or learners.
 7. **Principle 7** – Assessment must incorporate technology and use multiple tasks and methods.
 - That is, formative and summative assessments should be included in the assessment procedure to ensure that the purposes of assessment are adequately addressed. Assessment must also use technology.
 8. **Principle 8** - Timely feedback that promotes learning and facilitates improvement should be an integral part of the assessment process
 - Learners are entitled to feedback on submitted formative assessment tasks, and on summative tasks, where appropriate. The nature, extent, and timing of feedback for each assessment task should be made clear to learners in advance.
 9. **Principle 9** - Staff development policy and strategy should include an assessment
 - All those involved in the assessment of learners must be competent to undertake their roles and responsibilities.

Table 1: Some General Principles

Validity	Assessments should measure what they purport to measure and should align with the programme and module's learning outcomes.
Reliability	Assessment tasks should generate comparable grades across time, across markers and across methods.
Effectiveness	Assessment tasks should be designed to encourage good quality 'deep' approaches to learning in the students and where appropriate, evidence of competence/fitness to practice.
Comparability and Consistency	There should be consistent and comparable approaches to the summative assessment requirements of awards of the same level across programmes and institutions.
Equity and Diversity	All students should have opportunity to effectively demonstrate their learning and should have opportunity to be assessed by different methods across a programme/subject major.
Practicability and Efficiency	Assessment tasks should be practical for both staff and students in terms of the time needed for completion and marking and they should be cost effective.
Transparency	Information, guidance, assessment criteria, rules and regulations on assessment should be clear, accurate, consistent and accessible to all students, staff and examiners.
Attribution	Assessment tasks should generate clear evidence that the work has been produced by the candidate.

Qualities of Assessment Results

An assessment is a process through which learners can share their educational experiences. For a test to be a good tool for measuring learners' knowledge and skills, it should have the following characteristics that are essential for the success of any test. A good assessment is supposed to show what we have truly learned. There are four qualities of good assessments. Educators should ensure

these qualities are met before assessing students. A quality assessment in education consists of four elements - reliability, standardization, validity, and practicality

What Are the Characteristics of a Good Test in Education?

What is a good test in education? It is an evaluation through which teachers measure learners' abilities and points of weaknesses and strengths. It gauges their knowledge in the field of study and provides both sides with real feedback. A good test should ensure that learners are ready to move to the next step whether this step is a high school, college, or even the military

An assessment is a process through which students can share their educational experiences. For a test to be a good tool for measuring students' knowledge and skills, it should have the following characteristics of examination that are essential for the success of any test.

Reliability or Consistency: The reliability or consistency of a test means that learners should perform the same or get the same score if they are exposed to different questions in different times and places. A test is considered reliable when the same result is achieved over different tests OR "The reliability of test scores is the extent to which they are consistent across different occasions of testing, different editions of the test, or different raters scoring the test taker's responses." He also mentions some statistics to describe how a test can be reliable.

Making Sure Your Test Is Reliable:

- Score Distribution: The percentage of test takers at each score level.
- Mean Score: The average score, computed by summing the scores of all test takers and dividing by the number of test takers.
- Standard Deviation: A measure of the amount of variation in a set of scores. It can be interpreted as the average distance of scores from the mean. (Actually, it is a special kind of average called a "root mean square," computed by squaring the distance of each score from the mean score, averaging the squared distances, and then taking the square root.)
- Correlation: A measure of the strength and direction of the relationship between the scores of the same people on two tests.

Reliability is the ratio of the true score and the observed score variance. To measure a test's reliability, we may administer a test to the same group more than once.

However, errors may occur as students may forget or have some physical problems. Thus, it is crucial to administer the same test in identical conditions to ensure that we will get the same results.

Validity: The validity of a test can be achieved when the use and interpretation of the test results are sound. Therefore, certain criteria must be selected. Validity is very important to gauge the quality of a given test as questions must be in line with the selected criteria and measures. Here are some of the different types of validity:

- Content Validity: A test should fairly represent the content of the course or the field of study.

- **Criterion Validity:** It is used to predict the performance of a job applicant or a student.
- **Convergent validity:** This is mostly used in the field of sociology or psychology.
- **Discriminant Validity:** Discriminant validity means that a test of a concept is not highly correlated with other tests that are set to measure theoretically different concepts.

Objectivity: According to Gronlund and Linn “Objectivity of a test refers to the degree to which equally competent scores obtain the same results,” the test should be away from any personal or subjective judgment. It should be based only on the evaluation of human development. For example, in an essay-type test, students answer differently as each one has his/her own style of writing. Hence, when more than one instructor checks the test, they may give different scores according to whether they like the style or not. So, here, the test is less objective.

To avoid such bias, sharp rules should be set in evaluating such types of tests. There should be a unified guide for teachers to use while correcting such tests. Personal judgment does not occur in true or false or multiple-choice tests. Besides, teachers should receive training on how to score a test as untrained teachers may give wrong scores and not be able to maintain the required fairness and accuracy.

Comprehensiveness: A test should fully cover the entire field of study that students are exposed to during the course. Vague questions should not be included especially during online tests when students are confused and short in time.

Absence of Ambiguity: There must be no place for ambiguity, especially in online tests where examiners are absent. Students should not be left in confusion and all questions must be crystal clear. Ambiguous questions often result when instructors put off writing test questions until the last minute. Careful editing and an independent review of the test items can help to minimize this problem.”

To ensure the success of any test, instructors should take into consideration the following factors:

- Students must be well-prepared for the test through extensive revisions and discussions.
- There should not be any gaps between the revision period and the exam.
- Examiners should make it clear to students which topics are expected to be tackled in the exam.
- Students should be well-trained for the test type.

Appropriateness of Time: One of the top characteristics of a good test is when students have appropriate time to answer all questions. For example, essay questions require more time than multiple-choice or true/false questions. Some teachers take the test themselves first and then double or triple the time for students. A good test is supposed to be practical and comprehensive.

TIPS: Outstanding Characteristics of a Good Online Test

- No logistic setback
- Easy access from anywhere

- High speed
- Support essay questions, multiple-answer questions, short answers, & equation & scientific questions
- Built-in questions bank in quality online test systems
- Immediate students' results reports are generated
- Highly detailed, error-free analytics reports on students' performance as well as test and questions quality

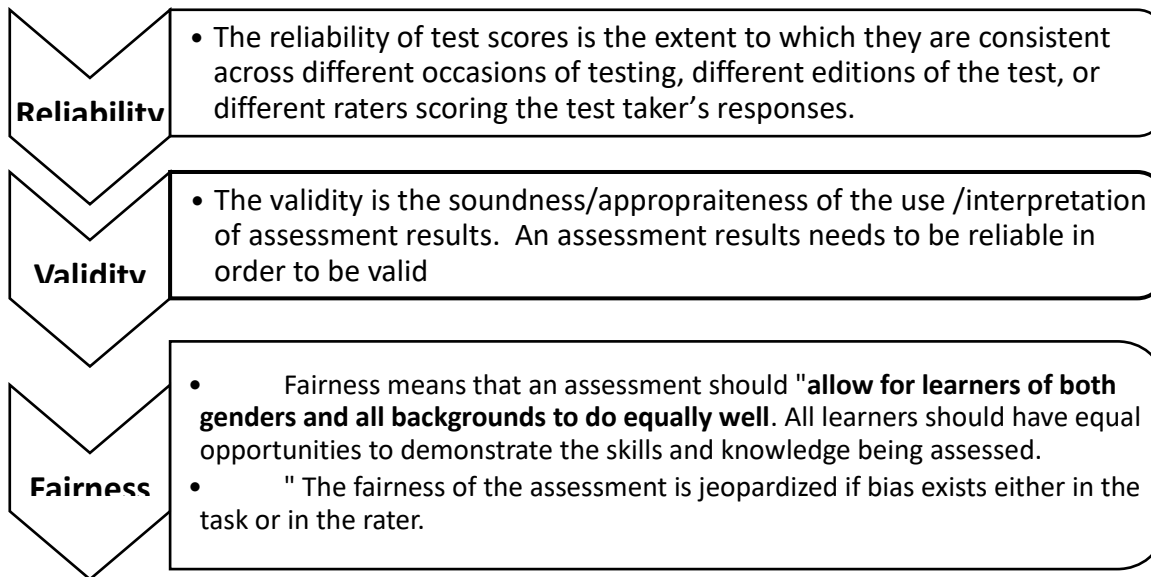
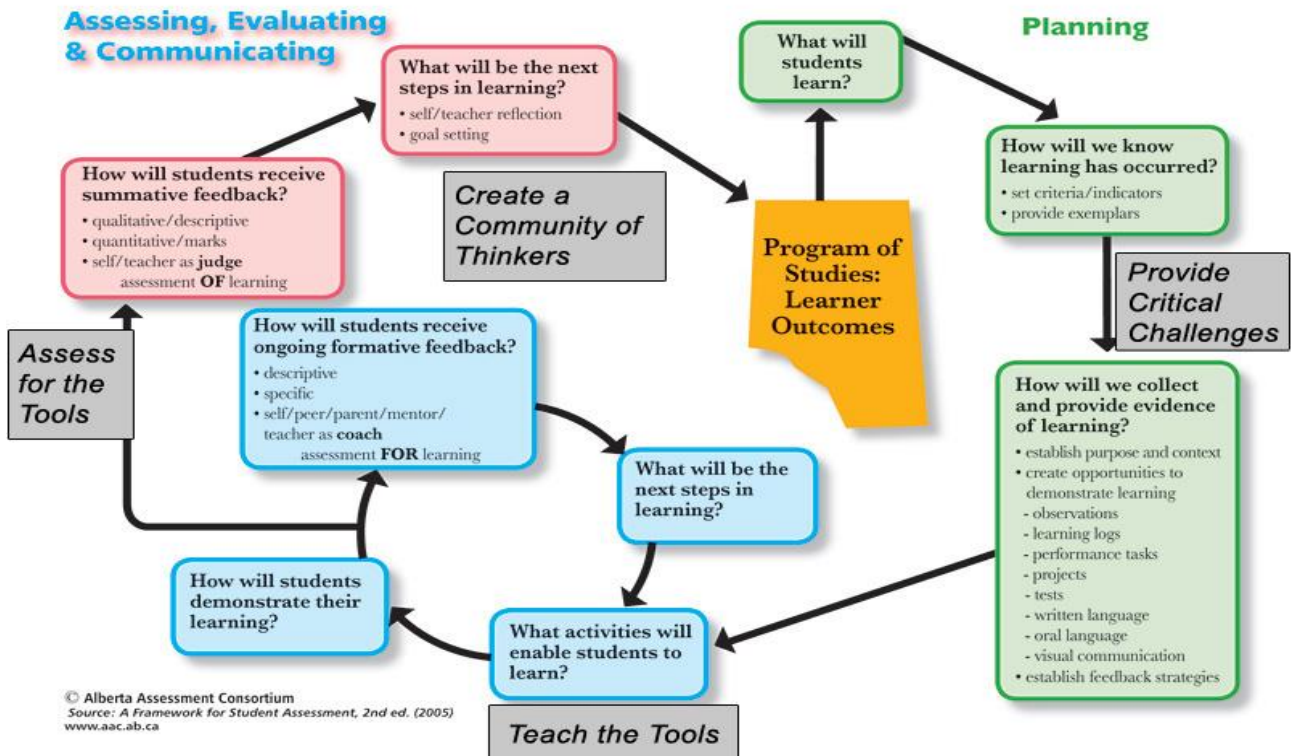


Figure 5: Qualities of Assessment Results

Assessing Student Learning in the Classroom



Key Ideas Units One and Two

- Assessment is the systematic basis for making inferences about the learning and development of learners.
- Principles and Qualities of Assessment results include:
 - Reliability: Consistency of test results
 - Validity: Accuracy of use and interpretation of test results
 - Fairness: The consideration of learner's needs and characteristics, and any reasonable adjustments that need to be applied to take account of them.

UNIT 3: PURPOSE OF ASSESSMENT

This session uses the understandings gained from Session 1 as building blocks to translate the principles into practice by exploring the why of assessment. It is important that teachers understand why we assess learners and who the key stakeholders are. This will inherently inform assessment practices by ensuring that the assessment principles are actual living documents.

Learning outcome

By the end of this unit the participant will be able to:

- Explain the purposes of assessment in terms of assessment of, as, and for learning
- Differentiate among the three purposes of assessment.
- Distinguish between formative and summative assessment
- Explain the essentials of these assessments for learners' development.

Importance of assessment in the teaching-learning process.

The assessment can benefit the students as well as benefit the teachers in many ways:
For students, assessment can mean:

- clarifying their instructors' expectations of them
- focusing more on learning as they...show more content...

It occurs as a part of everyday teaching and learning and uses information from the ongoing assessment to shape classroom teaching and learning. Assessment as learning puts the students in an active role as learners, using assessment information to self-assess, regulate and monitor their learning progress.

Assessment is important for several reasons:

1. Assessment results provide qualitative information that helps faculty determine how they might improve courses and/or programs through changes in curriculum, teaching methodologies, course materials or other areas. When integrated into the planning cycle for curriculum development and review, assessment results can provide a powerful rationale for securing support for curricular and other changes.
2. Assessment may provide comparative data that can give you valuable information on how well your students are meeting the learning outcomes for your course /program.
3. It brings about benefits for children gathering accurate information from young children is difficult and potentially stressful.

4. It should be tailored to a specific purpose and should be reliable, valid and fair. It is important that the teacher practice diagnostic assessment, formative assessment, summative assessment, and informal assessment in the classroom so that it can be used throughout the learning process so that the students are able to explore and use a wide range of assessment methods to monitor their learners' learning and progress.

Gibbs (2003) states that *assessment* has six main functions:

1. Capturing students' time and attention.
2. Generating appropriate student learning activities.
3. Providing timely feedback which students pay attention to.
4. Helping students to internalize the disciplines' standards and notion of equality.
5. Generating marks or grades which distinguish between students or enable pass/to fail decisions to be made.
6. Providing evidence for others outside the course to enable them to judge the appropriateness of standards on the course.

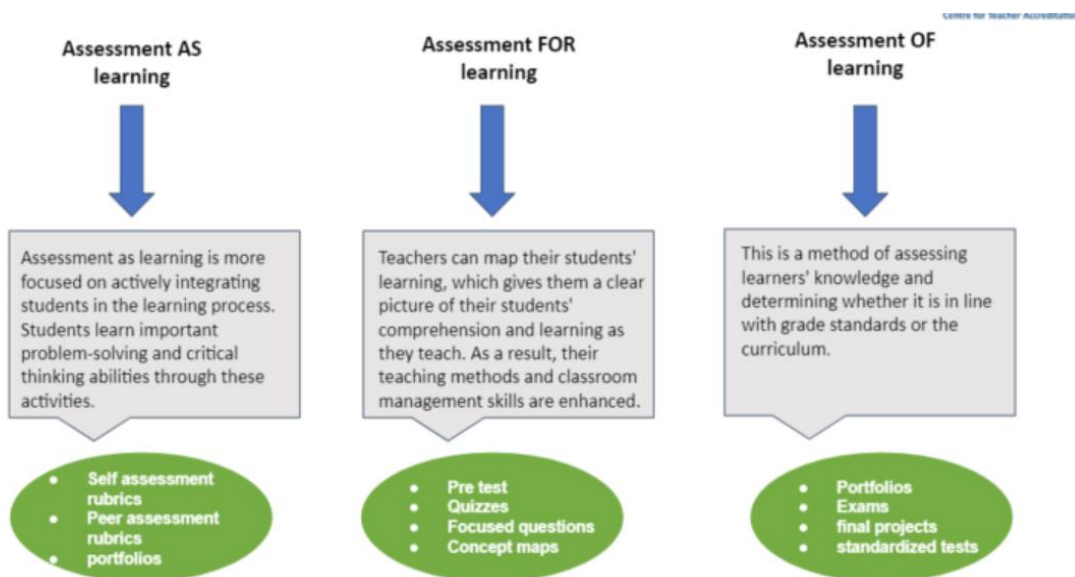


Figure 5: General Purposes of Assessment

Comparison of Assessment of Learning (AoL), Assessment for Learning (AfL), and Assessment as Learning (AaL).

Assessment	Of Learning	For Learning	As Learning
Type	Summative	Formative	Formative
What	Teachers determine the progress or application of knowledge or skills against a standard.	Teachers and peers check progress and learning to help learners to determine how to improve.	Learner takes responsibility for their own learning and asks questions about their learning and the learning process and explores how to improve.
Who	Teacher	Teacher & Peers	Learner & Peers
How	Formal assessments used to collect evidence of learner progress and may be used for achievement grading on grades.	Involves formal and informal assessment activities as part of learning and to inform the planning of future learning.	Learners use formal and informal feedback and self-assessment to help understand the next steps in learning.
When	Periodic report	Ongoing feedback	Continual reflection
Why	Ranking and reporting	Improve learning	Deeper learning and learning how to learn
Emphasis	Scoring, grades, and competition	Feedback, support, and collaboration	Collaboration, reflection, and self-evaluation

Summary of Assessment of, for and as Learning



Key Ideas

- **Assessment as Learning:** Occurs when learners reflect on and monitor their progress to inform their future
- **Assessment for Learning:** Occurs when teachers use inferences about learners' progress to inform their teaching.
- **Assessment of Learning:** Occurs when teachers use evidence of learners learning to make judgments on learner achievement against goals and standards

Forms of Assessment

There are two ways of assessing learners namely formative assessment and summative assessment. In this session, we will explain the difference between these two types of assessment, outline some methods of evaluation, and assess why both are essential to learners' development.

Formative and Summative Assessment

Formative Assessment	Summative Assessment
<ul style="list-style-type: none"> i. Is intended to collect data on learner's learning and progress in a way that gives both the learner and the teacher feedback and relevant information while teaching and learning are ongoing. ii. Formative assessments can be <i>administered</i> before, during, or after a lesson. iii. Formative assessments can inform the teacher on the need to reteach or reinforce a post-learning experience/ outcome. 	<ul style="list-style-type: none"> i. It could be administered at the end of teaching a unit, topic, or level in learning progression such as midterm, end of term or end of the year. ii. The expectation is that at this point, new topics or content will be introduced. iii. In some rare cases, summative assessment could be formative in determining how new learning should and would be approached. iv. Assessment at this stage could be the norm-referenced or criterion reference
Examples:	
<ul style="list-style-type: none"> i. Pop quizzes ii. Reading checks iii. Learning games iv. Homework v. Discussion questions vi. Music rehearsals vii. Reflection paragraphs 	<ul style="list-style-type: none"> i. Graded tests ii. Research reports iii. Structured essays iv. Portfolio projects v. Book reports vi. Final exams vii. Recitals or concerts viii. Standardized tests ix. Science projects

IMPORTANCE OF FORMATIVE ASSESSMENT FOR LEARNING

Formative assessment is a flexible and informal way of assessing a learner's progress and their understanding of a certain subject matter. It may be recorded in a variety of ways, or may not be recorded at all, except perhaps in lesson planning to address the next steps.

- Formative assessment helps learners identify their strengths and weaknesses and target areas that need work.
- It also helps educators and governors recognise where learners are struggling and address problems immediately.
- At a school level, school leaders use this information to identify areas of strength and weakness across the institution, and to develop strategies for improvement.
- As the learning journey progresses, further formative assessments indicate whether teaching plans need to be revised to reinforce or extend learning.

IMPORTANCE OF SUMMATIVE ASSESSMENT FOR LEARNING

In the current education system, standard-driven instruction plays a significant role. Summative assessment, therefore,

- provides an essential benchmark to check the progress of learners, institutions, and the educational program of the country.
- Summative assessment contributes largely towards improving the curriculum and overall curriculum planning.
- When summative assessment data indicates gaps across the board between learner knowledge and learning targets, schools may turn to improved curriculum planning and new learning criteria to assess and improve their school attainment levels.

Differentiation in assessment

Differentiated assessment is the way by which teachers modify and match assessment with the varied characteristics/profiles of students to meet the student's individual needs, thereby enhancing their learning and boosting their ability to show what they have learned. Students differ in their previous learning experiences, readiness, learning styles, preferences, academic standing, abilities, strengths and weaknesses, culture, race, and backgrounds. Teachers use differentiated assessment to match and respond to the varying learning needs of diverse students in a classroom.

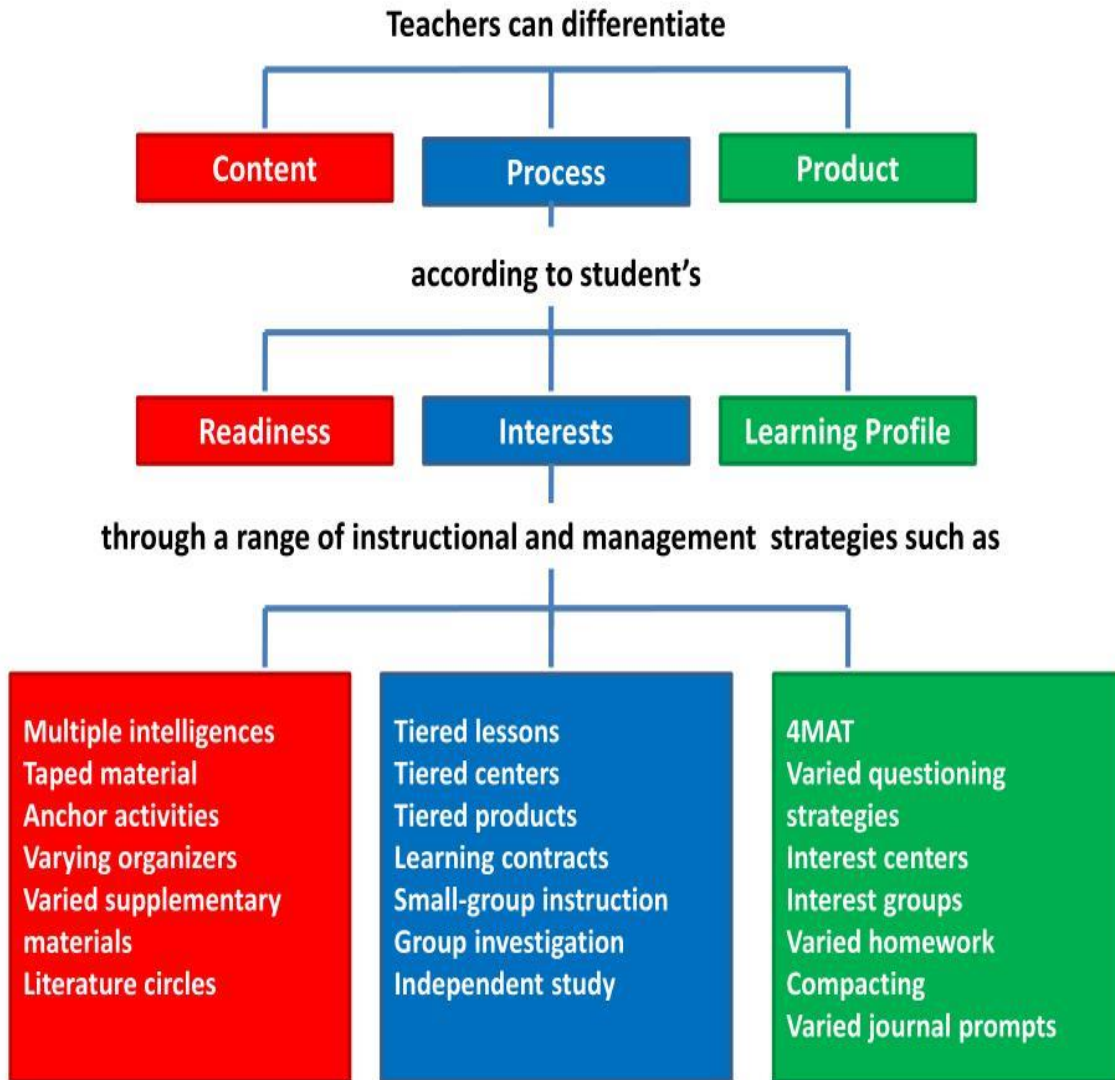
By differentiating assessments, teachers help diverse students to successfully demonstrate their competencies in particular ways that are fitting and effective for them. By providing various assessment methods/activities appropriate for particular types of students, the teachers can meet the students' individual needs, thereby helping them to be successful in their learning.

Designing various assessments apt for specific groups of learners provides more opportunities for students to effectively demonstrate what they have learned. Differentiated assessments also guide teachers on how they can differentiate, modify, and improve instruction.

Differentiated assessments can be done by designing and providing various assessment methods and activities that are appropriate for each type of students such that they can effectively learn and demonstrate what they have learned. Differentiated assessments can be done by providing them with various options and opportunities to show their learning and proficiency. Some ways of differentiating assessments:

- Designing tiered activities
- Scaffolding struggling learners
- Challenging advanced learners with more mid-stimulating activities
- Adjusting questions
- Compacting
- Flexible grouping
- Flexible assignments and tasks based on students' learning styles
- Learning contracts
- Asking students to do:
 - Roleplaying
 - Unit collage
 - Individual projects
 - Visual presentations
 - Oral presentations
 - Written presentations
 - Summaries and reflections
 - Lists, charts, and graphic organizers
 - Group/collaborative activities
 - Comic books
 - Raps/songs/dances/other performances

TIPS



Allan, S.D. & Tomlinson, CA (2000). *Leadership for Differentiating Schools & Classrooms*. Alexandria, VA: ASCD

To Differentiate Instruction in the Classroom

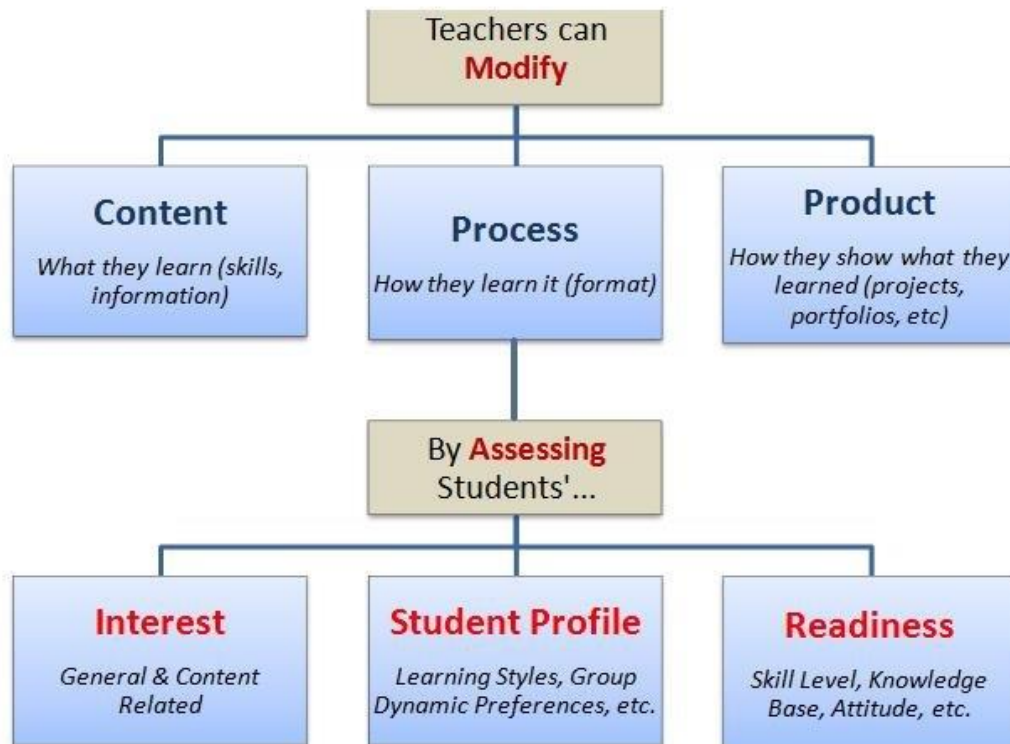


Diagram based on information from *The Differentiated Classroom: Responding to the Needs of All Learners* by Dr. Carol Ann Tomlinson (1999)

How is differentiated assessment different from other assessments?

Differentiated Instruction is a flexible student-centered approach to teaching that tailors instruction to meet students' individual learning needs. Differentiated Instruction can be carried out on several levels of content, process, product, and affect. It also takes into consideration and adapts the learning environment to students' needs. DI is characterized by the use of ongoing formative and summative assessment that guides instruction and allows for close monitoring of students' learning. DI implements flexible grouping. The following diagram illustrates the complex and multi-tiered nature of DI.

The self-assessment information can help you make decisions about adjusting your own approach to ensure that your teachers recognize the importance of DI in their classroom instruction and assessment and acquire additional skills needed for effective differentiated instruction. The indicators of effective differentiation outlined in the self-assessment are based on four general principles, or guidelines, of differentiated instruction found in current literature (C. A. Tomlinson, 1999):

In differentiated classrooms, teachers...

1. ***Create and sustain a responsive learning environment***, providing the foundation for long-term learning and positive connections to take place among students and adults.
2. ***Have clarity about the learning goals***, sharing, with students, what they should know, understand, and be able to do as a result of their learning.
3. ***Continuously assess student learning***, using information about what students already know, understand, and are able to do to make ongoing instructional decisions.
4. ***Establish flexible student groups and respectful tasks***, providing rigorous and engaging instruction matched to student needs.

3.6. Teachers who practice differentiation in the classroom may:

- Design assessment based on learners’ learning styles.
- Group learners by shared interest, topic, or ability for assignments.
- Assess learners’ learning using formative assessment.
- Manage the classroom to create a safe and supportive environment.
- Continually assess and adjust lesson content to meet learners’ needs.

Common Misconceptions about DI

#	Misunderstanding	Reality
1	Differentiation is a set of instructional strategies.	Differentiation is a philosophy—a way of thinking about teaching and learning. It is, in fact, a set of principles.
2	It’s adequate for a district or school leader (or professional developers) to tell, or even show, teachers how to differentiate instruction effectively.	Learning to differentiate instruction well requires rethinking one’s classroom practice and results from an ongoing process of trial, reflection, and adjustment in the classroom itself.
3	Differentiation is something a teacher does or doesn’t do (as in, “I already do that,” or “I tell our teachers that they already differentiate instruction.”).	Most teachers who remain in a classroom for longer than a day do pay attention to student variation and respond to it in some way— especially with students who can threaten order in the classroom. However, very few teachers proactively plan instruction to consistently address student differences in readiness, interest, and learning profile.
4	Differentiation is just about instruction.	Although differentiation is an instructional approach, effective differentiated instruction is inseparable from a positive learning environment, high-quality curriculum, assessment to inform teacher decision making, and flexible classroom management. To the degree that any one of those elements is weak, the others are also diminished.

Why use Differentiated Instruction?

- It's the law.
- Reduce the risk of underachievement.
- Alleviate discipline problems
- Increase motivation
- Address learner readiness
- Build self esteem
- Children already come to school differentiated.
- Because great minds DON'T think alike



Key Ideas

- Formative assessment forms part of the instructional process which helps teachers modify teaching methods and future lessons based on needs to improve learning.
- Summative assessment determines what learners know at a particular point in time/ end of a learning session to improve the overall school performance.
- Differentiated assessment is an ongoing process through which teachers gather data before, during, and after instruction from multiple sources to identify learners' needs and strengths

UNIT 4: TEST CONSTRUCTION, ADMINISTRATION, SCORING, AND GIVING FEEDBACK

Most classroom assessment involves tests that teachers have constructed themselves. It is estimated that mostly teacher-made tests are used in a typical classroom per year which results in perhaps billions of unique assessments yearly worldwide. This session seeks to equip participants with competencies for constructing, administering, and scoring teacher-made tests as well as providing effective feedback.

Learning outcomes

By the end of this unit, the participant will be able to:

- Describe how a test plan is used in test development to align the test to the content domain and learning objectives.
- Write and critique test items that match given learning objectives and depths of knowledge and that follow the item writing guidelines.
- Score and provide effective feedback

4.1. General Tips to Test Construction

1. **Start with your learning outcomes.** Choose objective and subjective assessments that match your learning outcomes and the level of complexity of the learning outcome.
2. **Use a test blueprint (Table of Specification-ToS).** A test blueprint is a rubric, document, or table that lists
 - a. the learning outcomes to be tested,
 - b. the level of complexity, and
 - c. the weight for the learning outcome (see sample).

Sample Test Blueprint

Pre and Post Assessment

Example of a Test Blueprint for Dental Hygiene 101

Program Objective/ Learning Outcome	Course Objective/ Learning Outcome	Bloom's Taxonomy Classification*	Number of Test Items	Point Value	% Weight of Test
An ability to apply knowledge of dental hygiene to personal care	Identifies active ingredient in toothpaste	Knowledge	1	2	2.4
	Explains why teeth should be cleaned at least twice per year	Comprehension	1	4	4.7
	Relates poor dental hygiene to poor overall health	Analysis	2	12	14.1
...
		TOTALS	23	85	100
*Classification of Learning.					
Note. Other learning classifications may be used in accordance with a program's accreditation standards.					
Number of Test Items and Point Values are determined by faculty.					

Screencast-O-Matic.com

A blueprint will make writing the test easier and contribute immensely to test validity. Note that Bloom's taxonomy can be very useful with this activity. For example:

EXAMPLE Blueprint FORMAT:

Growth Assessment Blueprint MHS

Essential Skill & Knowledge	Basic (Remember & Understand)	Standard (Apply & Analyze)	Expanded (Evaluate & Create)
#2: Developing and Using Models • (NGSS Practices #2)	1	2	1
#4 Analyzing and Using Data • (NGSS Practices #4) • Translate quantitative or technical information expressed in words in a text into visual form (e.g., a table or chart) and translate information expressed visually or mathematically (e.g., in an equation) into words. (CCSS RST.9-10.7) • Represent data with plots on the real number line (CCSS HSS.ID.A.1)	1	2	1
#7 Engaging in Arguments from Evidence • (NGSS Practices #7) • Assess the extent to which the reasoning and evidence in a text support the author's claim or a recommendation for solving a scientific or technical problem (CCSS RST.9-10.8)	1	2	1
#3 Planning & Carrying Out Investigations • (NGSS Practices #3)	1	2	1
#1 Reading Closely: • Cite strong and thorough textual evidence to support analysis of what a text says explicitly as well as inferences (CCSS RI.1) • Cite specific textual evidence to support analysis of science and technical texts, attending to the precise details of explanations or descriptions. (CCSS RST.9-10.1)	1	2	1

3. **Let your learners know what to expect on the test.** Be explicit; otherwise, learners may make incorrect assumptions about the test.
4. **Word questions clearly and simply.** Avoid complex questions, double negatives, and idiomatic language that may be difficult for learners, especially multilingual learners, to understand.
5. **Have a colleague or instructional assistant read through (or even take) your test.** This will help ensure your questions and exam are clear and unambiguous. This also contributes to the reliability and validity of the test
6. **Assess the length of the test.** Unless your goal is to assess learners' ability to work within time constraints, design your exam so that learners can comfortably complete it in the allocated time. A good guideline is to take the exam yourself and time it, then triple the amount of time it took you to complete the test or adjust accordingly.

7. **Write your test key (scoring guide/marking scheme) prior to learners taking the exam.** The point value you assign to each question should align with the level of difficulty and the importance of the skill being assessed. Writing the exam key enables you to see how the questions align with instructional activities. You should be able to easily answer all the questions. Decide if you will give partial credit to multi-step questions and determine the number of steps that will be assigned credit. Doing this in advance assures the test is reliable and valid.
8. **Design your test so that learners in your class have an equal opportunity to fully demonstrate their learning.** Use different types of questions, reduce or eliminate time pressure, allow memory aids when appropriate, and make your questions fair. An exam that is too easy or too demanding will not accurately measure your learners' understanding of the material.

4.2. Characteristics of test items/ questions, and how to choose which to use

Including a variety of test types in an exam enables the test constructor to better leverage the strengths and overcome the weaknesses of any individual item type. Multiple choice questions are popular for their versatility and efficiency, but many other question types can add value to a test. Some points to consider when deciding which, when, and how often to use a particular question type include:

- i. *Workload:* Some questions require more front-end workload (i.e., time-consuming to write), while others require more back-end workload (i.e., time-consuming to mark).
- ii. *Depth of knowledge:* Some question types are better at tapping higher-order thinking skills, such as analysing or synthesizing, while others are better for surface-level recall.
- iii. *Processing speed:* Some question types are more easily processed and can be more quickly answered. This can impact the timing of the test and the distribution of learners' efforts across different knowledge domains.

4.3. All test items should:

- i. Assess the achievement of learning outcomes for the unit and/or course
- ii. Measure important concepts and their relationship to that unit and/or course
- iii. Align with your teaching and learning activities and the emphasis placed on concepts and tasks
- iv. Measure the appropriate level of knowledge
- v. Vary in levels of difficulty (some factual recall and demonstration of knowledge, some application and analysis, and some evaluation and creation)

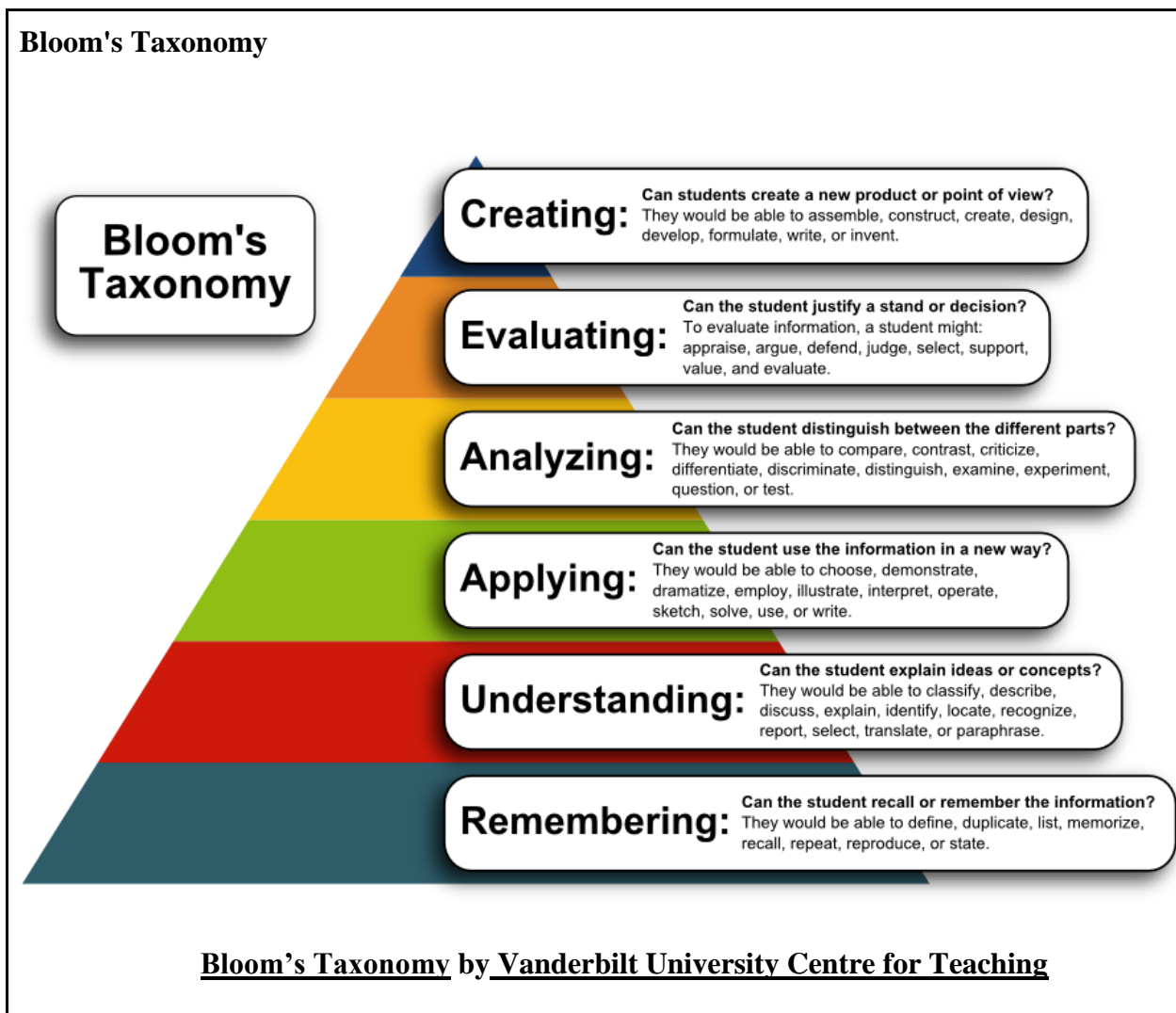


Figure 4.1: Bloom's Taxonomy with action verbs

4.4 Depth of knowledge (DOK) refers to the level of understanding required to answer a question or perform an activity. This concept is most often applied to the thinking that students do during assessments and other standards-driven evaluations. Depth of knowledge is largely believed to have been developed in the 1990s by Norman L. Webb, a researcher at the Wisconsin Center for Education Research. The depth of knowledge model has been highly popularized in the public education system.

The Purpose of the DOK Framework

Though originally developed for mathematics and science standards, DOK has been adapted for use in all subjects and is utilized most often in the creation of state assessment. This model ensures

that the complexity of assessments align with standards being assessed. When assessment follows the DOK framework, students are given a series of increasingly difficult tasks that gradually demonstrate that they are meeting expectations and allow assessors to evaluate their comprehensive depth of knowledge.

These assessment tasks are designed to capture the full scope of proficiency required to satisfy a standard, from the most basic to most complex and abstract units of knowledge and skill. That means that an assessment should include tasks from level 1 through 4—Webb identified four distinct depths of knowledge—and not too much of any one type of task. Assessment, just as the learning that precedes it, should be diversified and varied.

DOK in the Classroom

DOK is not reserved for state assessment small-scale, classroom assessment uses it too. Most classroom assessment consists of primarily level 1 and level 2 tasks because level 3 and 4 tasks are difficult to develop and score. However, teachers need to ensure that their students are exposed to a variety of tasks at differing levels of complexity to learn and grow to accurately assess whether expectations are met.

This means that teachers should design higher-level tasks though they require more time and effort because they offer benefits that simpler activities do not and show with more accuracy the full extent of a student's abilities. Teachers and students alike are best served by balanced assessment that calls on every depth of knowledge in some way.

Level 1

Level 1 is the first depth of knowledge. It includes recall of facts, concepts, information, and procedures this is the rote memorization and basic knowledge acquisition that makes higher-level tasks possible. Level 1 knowledge is an essential component of learning that does not require students to go beyond stating information. Mastering level 1 tasks builds a strong foundation on which to build.

Example of Level 1 Assessment Task

Question: Who was Dr. Kwame Nkrumah and what did he do?

Level 2

Level 2 depth of knowledge includes the limited application of skills and concepts. A common assessment of this is the use of information to solve multi-step problems. To demonstrate level 2 depth of knowledge, students must be able to make decisions about how to apply facts and details

provided to them as well as filling in any gaps using context clues. They must go beyond simple recall to answer questions about and make connections between pieces of information.

Example of Level 2 Assessment Task

Compare and contrast measurement, assessment and evaluation

Level 3

Level 3 DOK includes strategic thinking and reasoning that is abstract and complex. Students completing a level 3 assessment task must analyze and evaluate composite real-world problems with predictable outcomes. They need to apply logic, employ problem-solving strategies, and use skills from multiple subject areas to generate solutions. There is much multitasking expected of students at this level.

Example of Level 3 Assessment Task

Conduct and analyze the results of a survey about homework in your school. Decide what question you hope to answer. Represent this data in a graph and be able to present a conclusion about your findings.

Level 4

Level 4 includes extended thinking to solve complex and authentic problems with unpredictable outcomes. Students must be able to strategically analyze, investigate, and reflect while working to solve a problem, changing their approach to accommodate new information. This type of assessment requires highly sophisticated and creative thinking because it is open-ended by design there is no correct answer and a student must know how to evaluate their progress and determine whether they are on track to a feasible solution for themselves.

Example of Level 4 Assessment Task

Invent a new product or create a solution to a problem to make a fellow student's life easier.

Webb's Depth of Knowledge

<p>DOK Level 1 </p> <p style="text-align: center;">(Recall)</p> <p>Verbs: arrange, calculate, define, draw, identify, list, label, illustrate, match, memorize, recognize, tell, ...</p> <p>Focus: on specific facts, definitions, details, or procedures</p> <p>Note: there's one correct answer, and a combination of Level 1s does not make it a Level 2</p>	<p>DOK Level 2 </p> <p style="text-align: center;">(Skill / Concept)</p> <p>Verbs: categorize, cause/effect, classify, compare, distinguish, estimate, graph, interpret, modify, predict, relate, show, summarize, ...</p> <p>Focus: on applying skills and concepts • explaining how or why</p> <p>Note: there's one correct answer</p>
<p>DOK Level 3 </p> <p style="text-align: center;">(Strategic Thinking)</p> <p>Verbs: assess, cite evidence, compare, conclude, construct, critique, develop logical argument, differentiate, formulate, hypothesize, investigate, revise, ...</p> <p>Focus: on reasoning and planning in order to respond • complex and abstract thinking required • defending reasoning or conclusions</p> <p>Note: multiple answers or approaches</p>	<p>DOK Level 4 </p> <p style="text-align: center;">(Extended Thinking)</p> <p>Verbs: apply concepts, analyze, connect, create, critique, design, prove, ...</p> <p>Focus: on complex reasoning, planning, and thinking • make real-world applications in new situations</p> <p>Note: has multiple answers or approaches • often requires extended periods of time with multiple steps</p>

Webb's DOK Graphic by Tracy Watanabe is licensed under a Creative Commons Attribution 3.0 Unported License.

Assessment Formats

A **test** or examination is an assessment intended to measure a **test-takers** knowledge, skill, aptitude, physical fitness, or classification in many other topics. A **test** may be administered orally, on paper, on a computer, or in a confined area that requires a **test** taker to physically perform a set of skills.

Almost everybody has experienced testing during his or her life. Grammar tests, driving license test etc. Understanding the different types of testing, the kinds of results they provide, and how they complement one another help teachers determine what the best course of action is.

There are two general types of tests:

- Objective tests aim to assess a specific part of the learner's knowledge using questions which have a single correct answer.

- Subjective tests aim to assess areas of students’ performance that are complex and qualitative, using questioning which may have more than one correct answer or more ways to express it

Objective and Subjective Tests: Characteristics

These are some characteristics of objective and subjective tests:

Objective Tests characteristics:

- They are so definite and so clear that a single, definite answer is expected.
- They ensure perfect objectivity in scoring.
- It can be scored objectively and easily.
- It takes less time to answer than an essay test

Subjective Tests Characteristics

- Subjective items are generally easier and less time consuming to construct than are most objective test items
- Different readers can rate identical responses differently, the same reader can rate the same paper differently over time

Differences between Objective and Subjective Test Items

Objective type items	Subjective or essay items
<p>Learners select the correct response from several alternatives or supply a word or short phrase answer. These types of items are easier to create for lower order Bloom’s (recall and comprehension) while still possible to design for higher order thinking test items (apply and analyse).</p> <p>Objective test items include:</p> <ol style="list-style-type: none"> i. <u>Multiple choice</u> ii. <u>True-false</u> iii. <u>Matching</u> 	<p>Learners present an original answer. These types of items are easier to use for higher order Bloom’s (apply, analyse, synthesize, create, evaluate).</p> <p>Subjective test items include:</p> <ol style="list-style-type: none"> i. <u>Short answer essay</u> ii. <u>Extended response essay</u> iii. <u>Problem solving</u> iv. <u>Performance test items</u>

<p>iv. <u>Completion/Fill-in-the-blank</u></p> <p>Objective test items are best used when:</p> <ul style="list-style-type: none"> i. The group tested is large; objective tests are fast and easy to score. ii. The test will be reused (must be stored securely). iii. Highly reliable scores on a broad range of learning goals must be obtained as efficiently as possible. iv. Fairness and freedom from possible test scoring influences are essential. 	<p>(these can be graded as complete/incomplete, performed/not performed)</p> <p>Subjective test items are best used when:</p> <ul style="list-style-type: none"> i. The group to be tested is small or there is a method in place to minimize marking load. ii. The test is not going to be reused (but could be built upon). iii. The development of learners' writing skills is a learning outcome for the course. iv. Learner attitudes, critical thinking, and perceptions are as, or more, important than measuring achievement.
---	--

GENERAL STEPS OF TEST CONSTRUCTION”

The development of a good psychological test requires thoughtful and sound application of established principles of test construction. Before the real work of test construction, the test constructor takes some broad decisions about the major objectives of the test in general terms and population for whom the test is intended and also indicates the possible conditions under which the test can be used and its important uses.

These preliminary decisions have far-reaching consequences. For example, a test constructor may decide to construct an intelligence test meant for students of tenth grade broadly aiming at diagnosing the manipulative and organizational ability of the pupils. Having decided the above preliminary things, the test constructor goes ahead with the following steps:

1. Planning
2. Writing items for the test.
3. Preliminary administration of the test.
4. Reliability of the final test.
5. The validity of the final test.
6. Preparation of norms for the final test.
7. Preparation of manual and reproduction of the test.
8. PLANNING:

The first step in the test construction is the careful planning. At this stage, the test constructor address the following issues;

- **DEFINITION OF THE CONSTRUCT:** Definition of the construct to be measured by the proposed test.
- **OBJECTIVE OF THE TEST:** The author has to spell out the broad and specific objectives of the test in clear terms. That is the prospective users (For example Vocational counselors, Clinical psychologists, Educationalists) and the purpose or purposes for which they will use the test.
- **POPULATION:** What will be the appropriate age range, educational level and cultural background of the examinees, who would find it desirable to take the test?
- **CONTENT OF THE TEST:** What will be the content of the test? Is this content coverage different from that of the existing tests developed for the same or similar purposes? Is this cultural-specific?
- **TEST FORMAT:** The author has to decide what would be the nature of items, that is to decide if the test will be a multiple-choice, true-false, inventive response, or n some other form.
- **TYPE OF INSTRUCTIONS:** What would be the type of instructions i-e written or to be delivered orally?
- **TEST ADMINISTRATION:** Whether the test would be administered individually or in groups? Will the test be designed or modified for computer administration? A detailed agreement for preliminary and final administration should be considered.
- **USER QUALIFICATION AND PROFESSIONAL COMPETENCE:** What special training or qualifications will be necessary for administering or interpreting the test?
- **PROBABLE LENGTH, TIME STATISTICAL METHODS:** The test constructor must have to decide about the probable length and time for completion of test.
- **METHOD OF SAMPLING:** What would be the method of sampling i-e random or selective?
- **ETHICAL AND SOCIAL CONSIDERATIONS:** Is there any potential harm for the examinees resulting from the administration of this test? Are there any safeguards built into the recommended testing procedure to prevent any sort of harm to anyone involved in the use of this test.

- *INTERPRETATION OF SCORES*: How will the scores be interpreted? Will the scores of an examinee be compared to others in the criteria group or will they be used to assess mastery of a specific content area? To answer this question, the author must decide whether the proposed test will be criterion-referenced or norm-referenced.
- **MANUAL AND REPRODUCTION OF TEST**: Planning also include the total number of reproductions and a preparation of manual.

2. WRITING DOWN ITEMS:

ITEM:

A single question or task that is not often broken down into any smaller units. (Bean, 1953:15)

EXAMPLE: An arithmetical mean may be an item, a manipulative task may be an item, a mechanical puzzle may be an item and likewise sleeplessness may also be an item of a test. Items in a test are just like atoms in a matter that is they are indivisible.

The second step in item writing is the preparation of the items of the test. Item writing starts with the planning done earlier. If the test constructor decides to prepare an essay test, then the essay items are written down.

However, if he decides to construct an objective test, he writes down the objective items such as the alternative response item, matching item, multiple-choice item, completion item, short answer item, a pictorial form of item, etc. Depending upon the purpose, he decides to write any of these objective types of items.

PREREQUISITES FOR ITEM WRITING:

Item writing is essentially a creative art. There are no set rules to guide and guarantee the writing of good items. A lot depends upon the item writer's intuition, imagination, experience, practice, and ingenuity. However, there are some essential prerequisites that must be met if the item writer wants to write good and appropriate items. These requirements are briefly discussed as follows;

- **COMMAND ON SUBJECT MATTER**: The item writer must have a thorough knowledge and complete mastery of the subject matter. In other words, he must be fully acquainted with all facts, principles, misconceptions, Fallacies in a particular field so that he may be able to write good and appropriate items.
- **FULLY AWARE OF THE POPULATION**: The item writer must be fully aware of those persons for whom the test is meant. He must also be aware of the intelligence level of those

persons so that he may manipulate the difficulty level of the items for proper adjustment with their ability level. He must also be able to avoid irrelevant clues to correct responses.

- **FAMILIARITY WITH DIFFERENT TYPES OF ITEMS:** The item writer must be familiar with different types of items along with their advantages and disadvantages. He must also be aware of the characteristics of good items and the common probable errors in writing items.
- **COMMAND ON LANGUAGE:** The item writer must have a large vocabulary. He must know the different meanings of a word so that confusion in writing the items may be avoided. He must be able to convey the meaning of the items in the simplest possible language.
- **EXPERT OPINION:** After writing down the items, they must be submitted to a group of subject experts for their criticism or suggestions, which must then be duly modified.
- **CULTIVATE A RICH SOURCE OF IDEAS:** The item writer must also cultivate a rich source of ideas for items. This is because ideas are not produced in the mind automatically but rather require certain factors or stimuli. The common source of such factors are textbooks, Journals, discussions, questions for interviews, coarse outlines, and other instructional materials.

CHARACTERISTICS OF A GOOD ITEM:

An item must have the following characteristics.

- **CLARITY:** An item should be phrased in such a manner that there is no ambiguity regarding its meaning for both the item writer as well as the examinees who take the test.
- **MODERATELY DIFFICULT:** The item should not be too easy or too difficult.
- **DISCRIMINATING POWER:** It must have discriminating power, that is, it must clearly distinguish between those who possess the trait and those who do not.
- **TO THE POINT:** It should not be concerned with the trivial aspects of the subject matter, that is, it must only measure the significant aspects of knowledge or understanding.
- **NOT ENCOURAGE GUESSWORK:** As far as possible, it should not encourage guesswork by the subjects.

- **CLEAR IN READING:** It should not present any difficulty in reading.
- **INDEPENDENT FOR ITS MEANING:** It should not be such that its meaning is dependent upon another item and/or it can be answered by referring to another item.

GENERAL GUIDELINES FOR ITEM WRITING:

Writing items is a matter of precision. It is perhaps more like computer programming than writing a prose. The task of the item writer is to focus the attention of a large group of examinees, varying in background experience, environmental exposure and ability level on a single idea. Such a situation requires extreme care in the choice of words. The item writer must keep in view some general guidelines that are essential for writing good items. These are listed as under.

CLARITY OF THE ITEM:

The clarity in writing test items is one of the main requirements for an item to be considered good. Items must not be written as “verbal puzzles”. They must be able to discriminate between those who are competent and those who are not. This is possible only when the items have been written in simple and clear language. The items must not be a test of the examinee’s ability to understand the language.

The item writer should be very cautious particularly in writing the objective items because each such item provides more or less an isolated bit of knowledge and there the problem of clarity is more serious. If the objective item is a vague one, it will create difficulty in understanding and the validity of the item will be adversely affected. Vagueness in writing items may be because of several reasons such as poor thinking and incompetence of the item writer.

NON-FUNCTIONAL WORDS SHOULD BE AVOIDED:

Non-functional words must not be included in the items as they tend to lower the validity of the item. Non-functional words refer to those words which make no contribution towards the appropriate and correct choice of a response by the examinees. Such words are often included by the item writer in an attempt to make the correct answer less obvious or to provide a good distractor.

AVOID IRRELEVANT ACCURACIES:

The item writer must make sure that irrelevant accuracies unintentionally incorporated in the items are avoided. Such irrelevant accuracies reflect the poor critical ability to think on the part of the item writer. They may also lead the examinees to think that the statement is true.

DIFFICULTY LEVEL SHOULD BE ADAPTABLE:

The item must not be too easy or too difficult for the examinees. The level of difficulty of the item should be adaptable to the level of understanding of the examinees. Although it is a fact that exact decisions regarding the difficulty value of an item can be taken only after some statistical techniques have been employed, yet an experienced item writer is capable of controlling the difficulty value beforehand and making it adaptable to the examinees.

In certain forms of objective-type items such as multiple choice-items and matching items, it is very easy to increase or decrease the difficulty value of the item. In general, when the response alternatives are made homogenous, the difficulty value of the item is increased but when the response alternatives are made heterogeneous, except the correct alternative, the examinee is likely to choose the correct answer soon and thus, the level of difficulty is decreased.

The item writer must keep in view the characteristics of both the ideal examinees as well as the typical examinees. If he keeps the typical examinees (who are fewer in number) in view and ignores the ideal examinees, the test items are likely to be unreasonably difficult ones.

STEREOTYPED WORDS SHOULD BE AVOIDED:

The use of stereotyped words either in the stem or in the alternative responses must be avoided because these facilitate rote learners in guessing the correct answer. Moreover, such stereotyped words failed to discriminate between those who really know and understand the subject and those who do not. Thus, stereotyped words do not provide an adequate and discriminatory measure of an index. The most obvious way of getting rid of such words is to paraphrase the words in a different manner so that those who really know the answer can pick up the meaning.

IRRELEVANT CLUES MUST BE AVOIDED:

Irrelevant clues must be avoided. These are sometimes provided in several forms such as clang association, verbal association, length of the answer, keeping a different foil among homogenous foils, giving the same order of the correct answer, etc. In general, such clues tend to decrease the difficulty level of the item because they provide an easy route to the correct answer.

The common observation is that the examinees who do not know the correct answer, choose any of these irrelevant clues and answer on that basis. The item writer must, therefore, take special care to avoid such irrelevant clues. Specific determiners like never, always, all, none must also be avoided because they are also irrelevant clues to the correct answer, especially in the two-alternative items.

INTERLOCKING ITEMS MUST BE AVOIDED:

Interlocking items must be avoided. Interlocking items, also known as interdependent items, are items that can be answered only by referring to other items. In other words, when responding

correctly to an item is dependent upon the correct response of any other item, the item constitutes an example of an interlocking or independent item. For example:

- Sociometry is a technique used to study the effect structure of groups. True/false
- It is a kind of projective technique. True/false
- It was developed by Morene et al. true/false

The above examples illustrate the interlocking items. Answer to items 2 and 3 only be given when the examinee knows the correct answer of item 1. Such items should be avoided because they do not provide an equal chance for examinees to answer the item.

NUMBER OF ITEMS:

The item writer is also frequently faced with the problem of determining the exact number of items. As a matter of fact, there is no hard and fast rule regarding this. Previous studies have shown that the number of items I usually linked with the desired level of reliability coefficient of the test. Studies have revealed that usually 25-30 dichotomous items are needed to have the reliability coefficient as high as 0.80 whereas 15-20 items needed to reach the same level of reliability when multipoint items are used.

These are the minimum number of items which should be retained after item analysis. An item writer should always write almost *TWICE* the number of items to be retained finally. Thus, if he wants 30 items in the final test, he should write 60 items.

In the speed test, the number of items to be written is entirely dependent upon the intuitive judgment of the test constructor. On the basis of his previous experiences, he decides that a certain number of items can be answered with the given time limit.

ARRANGEMENT OF ITEMS:

After the items have been written down, they are reviewed by some experts are by the item writer himself and then arranged in the order in which they are to appear in the final test. Generally, items are arranged in increasing order of difficulty those having the same form (say alternative form, matching, multiple-choice, etc.) and dealing with the same contents are placed together.

3. PRELIMINARY ADMINISTRATION:

Before proceeding toward administration of the test review by at least three experts. When the test have been written down and modified in the light of the suggestions and criticisms given by the experts, the test is said to be ready for experimental try-out.

THE EXPERIENTIAL TRYOUT/ PRE-TRY-OUT:

The first administration of the test is called EXPERIMENTAL TRY-OUT or PRE-TRY-OUT. The sample size for experimental try out should be 100. The purpose of the experimental try out is manifold. According to Conrad (1951), the main purpose of the experimental try out of any psychological and educational test is as follows:

- **DETERMINES VAGUENESS AND WEAKNESSES:** Finding out the major weaknesses, omissions, ambiguities, and inadequacies of the Items.
- **DETERMINING DIFFICULTY LEVEL OF EACH ITEM:** Experimental try out helps in determining the difficulty level of each item, which in turn helps in their proper distribution in the final form.
- **DETERMINES TIME LIMIT** Helps in determining a reasonable time limit for the test.
- **DETERMINES APPROPRIATE LENGTH OF A TEST.** Determining the appropriate length of the tests. In other words, it helps in determining the number of items to be included in the final form.
- **IDENTIFYING WEAKNESSES IN DIRECTIONS.** Identifying any weaknesses and vagueness in directions or instructions of the test.

PROPER TRYOUT:

The second preliminary administration is called PROPER TRYOUT. At this stage test is delivered to the sample of 400 and must be similar to those for whom the test is intended. The proper try-out is carried out for the item analysis. ITEM ANALYSIS is the technique of selecting discriminating items for the final composition of the test. It aims at obtaining three kinds of information regarding the items. That is;

- **ITEM DIFFICULTY:** Item difficulty is the proportion or percentage of the examinees or individuals who answer the item correctly.
- **DISCRIMINATORY POWER OF THE ITEMS:** The discriminatory power of the items refers to the extent to which any given item discriminates successfully between those who possess the trait in larger amounts and those who possess the same trait in the least amount.
- **EFFECTIVENESS OF DISTRACTORS:** Determines the non-functional distractors.

FINAL TRYOUT:

The third preliminary administration is called the Final tryout. The sample for final administration should be at least 100. At this stage, the items are selected after item analysis and constitute the test in the final form. It is carried out to determine the minor defects that may not have been detected by the first two preliminary administrations.

The final administration indicates how effective the test will be when it would be administered on the sample for which it is really intended. Thus, the preliminary administration would be a kind of “DRESS REHEARSAL” providing a sort of final check on the procedure of administration of the test and its time limit. After final tryout, expert opinion should be considered again.

4. Reliability OF The Final Test

The basis of the experimental or empirical tryout the test is finally composed of the selected items, the final test is again administered on a fresh. For this purpose, we check the reliability of the test and it indicates the consistency of scores. In simple words, it is defined as the degree to which a measurement is consistent. If findings from the research are replicated consistently then they are reliable. Reliability also refers to the self-correlation of a test. A correlation coefficient can be used to assess the degree of reliability; if a test is reliable it should show a high positive correlation.

Types of Reliability

- Internal reliability
- External reliability

Internal Reliability- Internal reliability assesses the consistency of results across items within a test.

External Reliability- External reliability refers to the extent to which a measure varies from one use to another.

Errors in Reliability: At a time scores are not consistent because some other factors also affect reliability e.g.

- Noise
- Health
- Time

There is always a chance of 5% error in reliability which is acceptable.

TYPES OF ERRORS

- Random error
- Systematic error

Random error

A random error exists in every measurement and is often a major source of uncertainty. These errors have no particular assignable cause. These errors can never be totally eliminated or corrected. These are caused by many uncontrollable variables that are an inevitable part of every analysis made by human beings. These variables are impossible to identify, even if we identify some they cannot be measured because most of them are so small.

Systematic error

Systematic error is caused due to instruments, machines, and measuring tools. It is not due to individuals. Systematic error is acceptable we can fix and handled it.

WAYS OF FINDING RELIABILITY:

Following are the methods to check reliability

- Test-retest
- Alternate form
- Split –half method

TEST-RETEST METHOD:

It is the oldest and commonly used method of testing reliability. The test-retest method assesses the external consistency of a test. Examples of appropriate tests include questionnaires and psychometric tests. It measures the stability of a test over time.

A typical assessment would involve giving participants the same test on two separate occasions. Each and everything from start to end will be the same in both tests. The results of the first test need to be correlated with the result of the second test. If the same or similar results are obtained then external reliability is established.

The timing of the test is important if the duration is too brief then participants may recall information from the first test which could bias the results. Alternatively, if the duration is too long it is feasible that the participants could have changed in some important way which could also bias the results.

The utility and worth of a psychological test decrease with time so the test should be revised and updated. When tests are not revised systematic error may arise.

ALTERNATE FORM:

In an alternate form, two equivalent forms of the test are administered to the same group of examinees. An individual has given one form of the test and after a period, the person is given a different version of the same test. The two forms of the test is then correlated to yield a coefficient of equivalence.

- **Positive point:** In alternate form no deal to wait for time.
- **Negative point:** It is very hectic and risky task to make two tests of equivalent level.

SPLIT-HALF METHOD:

The split-half method assesses the internal consistency of a test. It measures the extent to which all parts of the test contribute equally to what is being measured. The test is technically spitted into odd and even form. The reason behind this is when we making test we always have the items in order of increasing difficulty if we put (1,2, —10) in one half and (11,12, —20) in another half then all easy question/items will go to one group and all difficult questions/items will go to the second group.

When we split the test we should split it with same format/theme e.g. Multiple questions – multiple questions or blanks – blanks.

5. VALIDITY OF THE TEST:

It refers to the extent to which test claim to measure what it claims to measure.

If a test is reliable then it is not necessary, that it is valid but if a test is valid then it must be reliable.

TYPES OF VALIDITY:

- External validity
- Internal validity

EXTERNAL VALIDITY: It is the extent to which the results of a research study can be generalized to different situations, different groups of people, different settings, different conditions etc.

INTERNAL VALIDITY: It is basically the extent to which a study is free from flaws and that any differences in a measurement are due to an independent variable.

TYPES OF VALIDITY

- Face validity
- Construct validity
- Criterion-related validity

- **FACE VALIDITY**

Face validity is determined by a review of the items and not through the use of statistical analysis. Face validity is not investigated through formal procedures. Instead, anyone who looks over the

test, including examinees, may develop an informed opinion as to whether or not the test is measuring what it is supposed to measure. While it is clearly of some value to having the test appear to be valid, face validity alone is insufficient for establishing that the test is measuring what it claims to measure.

- **CONSTRUCT VALIDITY:**

It implies using the construct correctly (concepts, ideas, notions). Construct validity seeks agreement between a theoretical concept and a specific measuring device or procedure.

For example, a test of intelligence now a days must include measures of multiple intelligences, rather than just logical-mathematical and linguistic ability measures.

- **CRITERION RELATED VALIDITY:**

It states that the criteria should be clearly defined by the teacher in advance. It must take into account other teachers' criteria to be standardized and it also needs to demonstrate the accuracy of a measure or procedure compared to another measure or procedure which has already been demonstrated to be valid.

6. NORMS:

When psychologists design a test to be used in a variety of settings, they usually set up a scale for comparison by establishing norms.

Norm is defined as the average performance or scores of a large sample representative of a specified population. Norms are prepared to meaningfully interpret the scores obtained on the test for as we know, the obtained scores on the test themselves convey no meaning regarding the ability or trait being measured. But when these are compared with the norms, a meaningful inference can immediately be drawn.

Types of norms:

- Age norms
- Grade norms
- Percentile norms
- Standard scores norms

All these types of norms are not suited to all types of tests. Keeping in view the purpose and type of test, the test constructor develops a suitable norm for the test.

- **AGE NORM**

Age norms indicate the average performance of different samples of test takers who were at various ages at the time the test was administered. If the measurement under consideration is height in

inches for example we know that scores (heights) for children will gradually increase at various rates as a function of age up to the middle to late teens.

The child of any chronological age whose performance on a valid test of intellectual ability indicated that he or she had intellectual ability similar to that of the average child of some other age was said to have the mental age of the norm group in which his or her test score fell.

The reasoning here was that irrespective of chronological age, children with the same mental age could be expected to read the same level of material, solve the same kinds of math problems, and reason with a similar level of judgment. But some have complained that the concept of mental age is too broad and that although a 6-year-old might, for example, perform intellectually like a 12-year-old, the 6-year-old might not be very similar at all to the average 12 years old socially, psychologically, and otherwise.

- **GRADE NORMS:**

Grade norm was designed to indicate the average test performance of test-takers in a given school grade, grade norms are developed by administering the test to representative samples of children over a range of consecutive grade levels.

Like age norms, grade norms have a widespread application with children of elementary school age, the thought here is that children learn and develop at varying rates but in ways that are in some aspects predictable.

One drawback of grade norms is that they are useful only with respect to years and months of schooling completed. They have little or no applicability to children who are not yet in school or who are out of school.

- **PERCENTILE NORMS:**

Percentile system is ranking of test scores that indicate the ratio of score lower from higher than a given score. A percentile is an expression of the percentage of people whose score on a test or measure falls below a particular raw score. A more familiar description of test performance, the concept of percentage correct, must be distinguished from the concept of a percentile. A percentile is a converted score that refers to a percentage of test takers.

Percentage correct refers to the distribution of raw scores-more specifically, to the number of items that were answered correctly multiplied by hundred and divided by the total number of items.

Because percentiles are easily calculated they are a popular way of organizing test data and are very adoptable to a wide range of tests.

- **STANDARD SCORE NORMS:**

When a raw score is converted into a formula it becomes standard scores. For example, marks obtained in the paper may be in 100% are applicable only in a specific area but when they are converted in GPA, they become standard score.

7. PREPARATION OF MANUAL AND REPRODUCTION OF THE TEST:

The last step in test construction is the preparation of a manual of the test. In the manual, the test constructor reports the psychometric properties of the test, norms, and references. This gives a clear indication regarding the procedures of the test administration, the scoring methods, and time limits if any of the test.

It also includes instructions as well as the details of the arrangement of materials that is whether items have been arranged in random order or in any other order. The test constructor finally orders for printing of the test and the manual.

Scoring and giving feedback

A scoring guide for assessment, usually in the form of a matrix or grid, is a tool used to interpret and grade learners' assessment against criteria and standards. Scoring guide are sometimes called "marking scheme", "rubric", or "scoring guides". Scoring guides can be designed for any content domain.

Benefits of Assessment rubrics:

1. provide a framework that clarifies assessment requirements and standards of performance for different grades.
2. enable very clear and consistent communication with learners about assessment requirements and about how different levels of performance earn different grades.
3. They allow assessors to give very specific feedback to learners on their performance.
4. when learners are involved in their construction, encourage them to take responsibility for their performance.
5. when used for self-assessment and peer assessment, make learners aware of assessment processes and procedures, enhance their meta-cognitive awareness, and improve their capacity to assess their own work.
6. can result in richer feedback to learners, giving them a clearer idea where they sit in terms of an ordered progression towards increased expertise in a learning domain.

7. by engaging colleague teams in rubric-based conversations about quality, help them develop a shared language for talking about learning and assessment.
8. help assessors efficiently and reliably interpret and grade learners' work.
9. systematically illuminate gaps and weaknesses in learners' understanding against criteria, helping teachers target areas to address

Feedback

Feedback is any response regarding a learner's performance or behavior. It can be verbal, written, or gestural. The purpose of feedback in the assessment and learning process is to improve a learner's performance. It is essential that the process of providing feedback is a positive, or at least a neutral, learning experience for the learner. Negative feedback can discourage learner effort and achievement. Teachers have the distinct responsibility to nurture a learner's learning and to provide feedback in such a manner that the learner does not leave the classroom feeling defeated.

Characteristics of Effective Feedback

1. Educative in Nature

Providing feedback means giving learners an explanation of what they are doing correctly and incorrectly, with the focus of the feedback on what the learners is doing right. It is most productive to a learner's learning when they are provided with an explanation as to what is accurate and inaccurate about their work. One technique is to use the concept of a "feedback sandwich" to guide your feedback: Compliment, Correct, Compliment.

2. Given In a Timely Manner

When learner feedback is given immediately after showing proof of learning, the learner responds and remembers the experience about what is being learned more positively. If we wait too long to give feedback, the learner might not connect the feedback with the learning moment.

1. Sensitive to the Individual Needs of the Learner

It is vital that we take into consideration each individual when giving learner feedback. Our classrooms are full of diverse learners. Some learners need to be nudged to achieve at a higher level and other needs to be handled gently so as not to discourage learning and damage self-esteem.

2. Answers the Questions

Effective teaching and learning have shown that learners want to know where they stand regarding their work. Providing answers to the following four questions on a regular basis will help provide quality learner feedback.

- i. What can the learner do?
- ii. What can't the learner do?
- iii. How does the learner's work compare with that of others?
- iv. How can the learner do better?

Feedback is one of the most powerful influences on learning and achievement, but this impact can be either positive or negative.” - Hattie & Timperley, 2007

Key ideas

- Designing tests involves the planning, preparing, administering, scoring, statistically analyzing, and reporting results of tests
- Score the essay test when you are physically sound, mentally alert, and in an environment with very little or no distraction.
- Feedback improves learner confidence, motivation to learn, and ultimately, a learner's attainment.