GAROWE TEACHERS EDUCAITON COLLEGE (GTEC)

GAROWE, PUNTLAND, SOMALIA

LECTURE NOTES ON EDUCATIONAL MEASUREMENT AND EVALUATION
FOR
FOURTH SEMESTER FOR DIPLOMA IN PRIMARY TEACHER
AND
DIPLOMA IN SECONDARY TEACHER

Abdirahman Sheikhdom

Mail to: abdirahman.sh.ali@gmail.com

March to July 2023

"When you can measure what you are speaking about, and express it in numbers, you know something about it."Kelvin

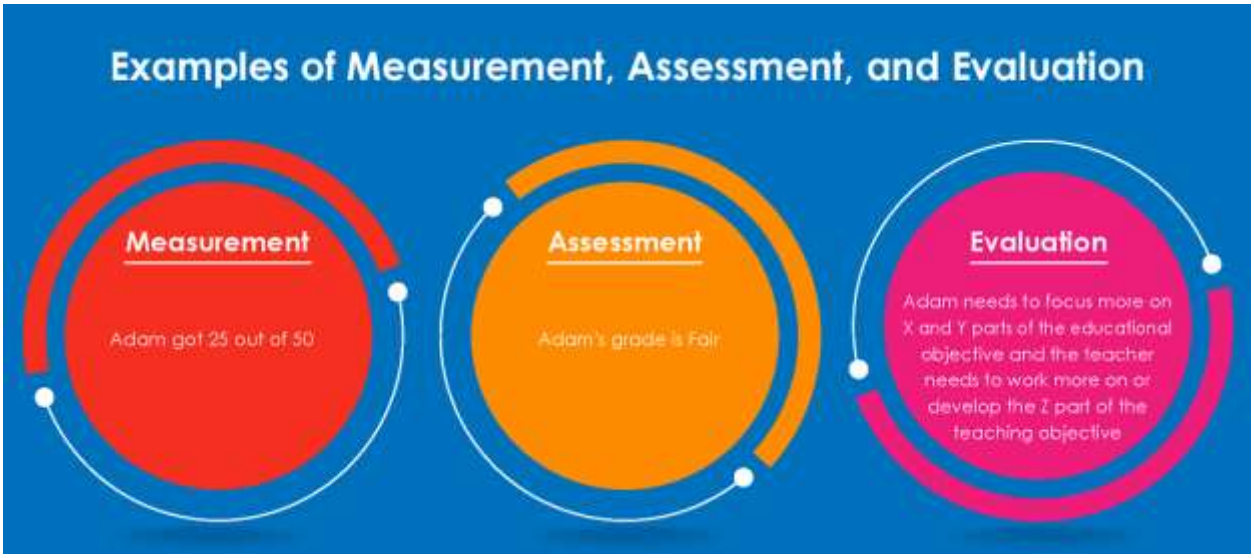
Assessment, Measurement and Evaluation in Education

Assessment refers to the process of collecting information that reflects the performance of a student, school, classroom, or an academic system based on a set of standards, learning criteria, or curricula. It contains both qualitative and quantitative information regarding students' performance.

Measurement refers to a process through which a learning phenomena, context, or experience is translated into a set of numerical variables that are representative. It contains purely quantitative data and information regarding student's academic performance.

Evaluation is the process of gathering and interpreting evidences on the changes in behaviour of the students as they progress through school.

Evaluation provides a teacher feedback to understand the academic performance and achievement. Through evaluation, a teacher can bring about certain required changes and modifications in the teaching methods and education system as a whole.



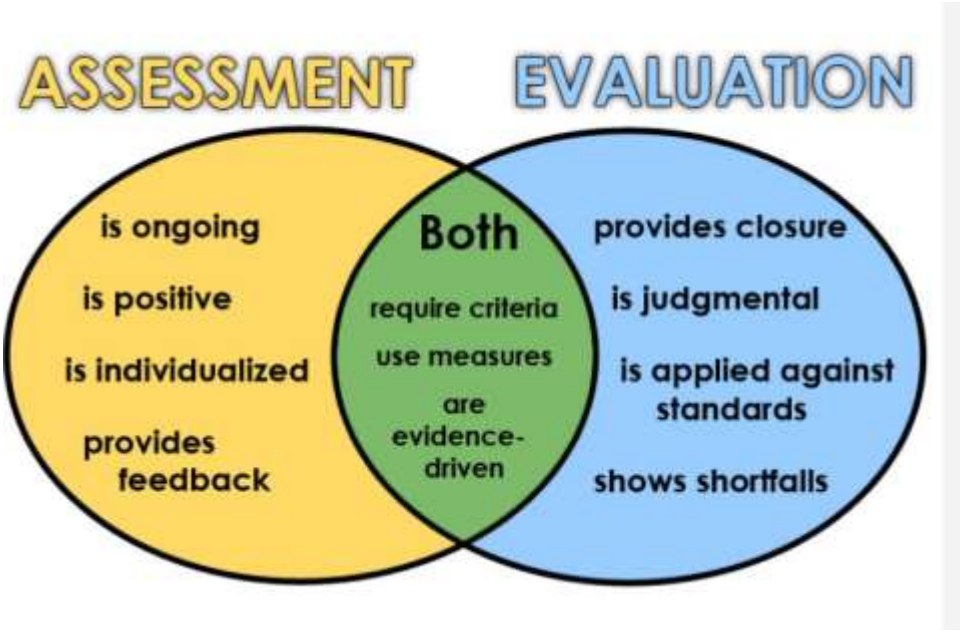
The differences between measurement assessment and evaluation in education are as follows:

Measurement	Assessment	Evaluation
1. Measurement is the process of the delegation of a numerical index to the object consistently and meaningfully. It has **quantitative information**.	Assessment is a process by which the information is obtained to achieve a particular goal. It has **qualitative and quantitative information.**	1. Evaluation refers to the comparison made between the score of one learner with that of other learners to judge the results. It has **quantitative information + qualitative informative + value judgment**
2. It involves observations which can be expressed in numerical terms.	It contains the information with marks, grades as well as the interest and goals of the student.	2. It includes observations which are quantitative as well as qualitative.
3. It is the assessment of numerals as per certain standards.	Assessment has the specific goal to find out the learning difficulties.	3. It is the assessment of symbols, grades, or levels as per the established standards.
4. It does not express the student's logical assumption.	It contains diagnostic tests, personal interactions and discussion with other teachers.	4. The student's logical assumption can be made.
5. It requires relatively less energy and time and is content oriented.	It requires some energy and time to find out the learning behaviour of the student.	5. It requires more energy and time and is objective oriented.

The relationship between measurement assessment and evaluation is as follows:

1. Evaluation and Assessment involve all the tasks of education. It goes beyond these things and although it depends on measurement.
2. Measurement explains a situation and evaluations and assessment describe the judgment of its worth and value.

3. Without evaluation, measurement is quite meaningless because evaluation adds a little significance to it.
4. Measurement makes a quantitative determination of an individual's performance and evaluation makes a qualitative determination of the individual's performance.



Aims and Objective of Assessment and Evaluation in Education

- It aims to enable the instructors to determine how much the learners have understood what the teacher has taught in the class and how much they can apply the knowledge of what has been taught in the class as well.
- Another aim is that the standardized tests involved in the process of measurement assessment and evaluation enable the students to make better use of the data available in the daily classroom.
- The basic purpose of measurement, assessment and evaluation is to determine the needs of all the learners.
- Measurement, assessment and evaluation also aims to enable educators to measure the skills, knowledge, beliefs, and attitude of the learners.

- Measurement, assessment and evaluation also aims to help the teachers to determine the learning progress of the students. Without measuring and evaluating their performance, teachers will not be able to determine how much the students have learned.

Conclusion:

Therefore, evaluation is more broader and comprehensive than assessment. Evaluation includes assessment. Measurement is a tool or method employed in assessment and evaluation

To sum up, we measure distance, we assess learning, and we evaluate results in terms of some set of criteria. These three terms are certainly share some common attributes, but it is useful to think of them as separate but connected ideas and processes.

Types of assessment/Evaluation in education

The main types of assessment in education are as follows:

Diagnostic Assessment/Evaluation: It is a type of pre-assessment which allows teachers to assess the strengths, weaknesses, knowledge, and skills of the learners before starting the teaching learning process

Formative Assessment/Evaluation: is the type of evaluation which is done during the teaching process. Its main aim is to offer continuous feedback to the student as well as the teacher. This enables one to make modifications in the instructional process if there are any requirements. It takes into account those units of the curriculum which are smaller and independent and through tests, the learners' performance is evaluated. Formative assessment is effective for making changes and timely corrections among the learners and teaching methodology.

Summative Assessment/Evaluation: it refers to the evaluation which is conducted towards the end of the academic session. It takes into account the achievements of the outcomes and the overall personality development of the learner at the academic session's end. It takes a broad aspect of learning into account.

Another types include:

Quantitative Assessment/Evaluation: it refers to the use of scientific measurements

and tools and includes the results which can be measured and counted. Few techniques and tools involved in quantitative evaluation are oral, written, and performance tests.

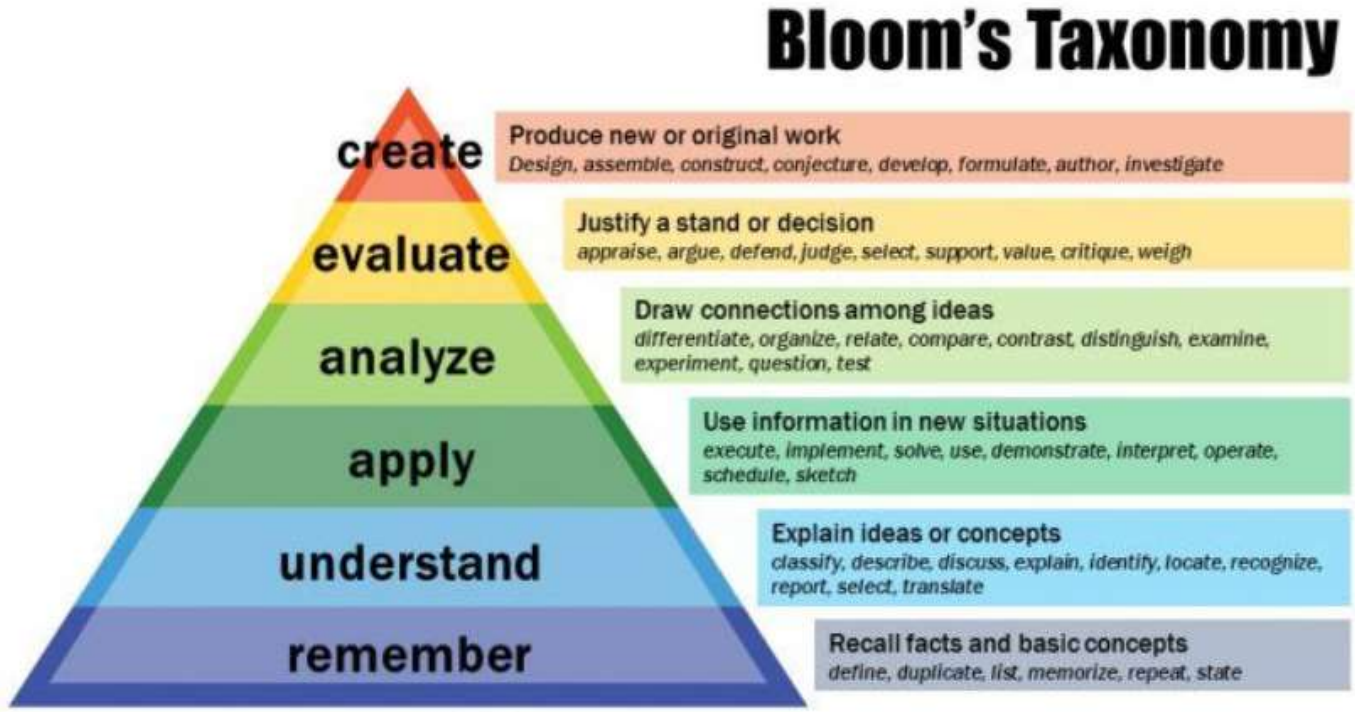
Qualitative assessment/Evaluation: it refers to an observation that is made using all the five senses and includes value judgment. It makes a quantitative determination of an individual's performance.

Difference Between Quantitative and Qualitative Evaluation

Quantitative Education	Qualitative Education
1. It involves specific measurements obtained from surveys.	It includes information gathered from interviews.
2. It is a form of short-term standardized exams.	It is a form of continuous long-term assessment.
3. It focuses on the right and correct answers.	It focuses on creative and open-minded answers.
4. It is timed and in a multiple choice format.	It is untimed and not in a multiple choice format.
5. It is oriented to product.	It is oriented to process.

Bloom Taxonomy

Benjamin Bloom developed this theory in 1956. The concept is based on the idea that learning is a sequential process. It consisted of six level: Remember, Understand, Apply, Analyse, Evaluate, and Create. The categories after Knowledge were presented as "skills and abilities," with the understanding that knowledge was the necessary precondition for putting these skills and abilities into practice.



How to Use Bloom's Taxonomy in Test Writing

Benjamin Bloom found that when questions are written in a complicated or inaccurate manner, the learner might answer them incorrectly even if they had studied the material well. To prevent this, the scientist suggested using specific action verbs in questions for each level of assessment. They help a trainer or teacher word a question correctly, provide an objective knowledge check, and allow the learner to understand what is expected of them in this task.

Analysis + Synthesis	Compare, contrast, separate, change, find, collect, combine, summarize, group, match, collect, set up	Questions on these levels make the learner go beyond the instructions and figure out the situation themselves.
Evaluation	Justify, judge, recommend, rate, evaluate, relate, predict, appraise, argue, support	This level checks if the learner can come up with a new solution based on new information, evaluate the situation, and act independently.
Comprehension	generalize, find, paraphrase, give an example, describe, estimate, infer, rewrite	simply learned the material by heart or if they truly understood the topic. It also shows if your course lacks explanations and details.
Application	Apply, decide, calculate, use, modify, transform, classify, arrange, discover, demonstrate, prepare, produce, write	This level checks if the learner can apply new knowledge in practice. It also shows if the course lacks practical value.

To check learners' knowledge effectively, Bloom suggests creating six types of questions. Each type corresponds to a specific taxonomy level. For example, you can first check how well the person learned the material, then find out what they understood, learn what knowledge they can apply, and so on.

Example. When answering the first group of questions, Ali got 4 correct answers out of 10. The main objective of these questions was to check how well he had learned the material from the course. Because Ali failed more than half of the questions, there's no use testing him further and checking his practical skills. It will be best if the test automatically sends him to the beginning of the course for retraining or provides additional information on problematic issues.

Types of Test

In terms of the nature of the test, it can be classified as below:

- Objective test
- Subjective test

Objective Test is which require students to select the correct response from several alternatives or to supply a word or short phrase to answer a question or complete a statement. Objective tests are considered to be more reliable than subjective tests on the basis of its scoring procedure. Objective test, no judgment is required on the part of the scorer and the scoring is considered to be objective. Objective tests can be of various types such as-

- Multiple Choice
- Constant Alternative
- True and false
- Rearrangement
- Matching Type

Strengths of Objective Tests-

- These tests can be marked with complete reliability
- Large no. of questions can be answered in a relatively short period of time.
- These tests are easy to administer and mark.

Limitations of Objective Tests-

- These tests are difficult and initially expensive to construct.
- It is not possible to see the reasoning behind a wrong answer.
- These tests are more restricted.

Subjective Test permits the student to organize and present an original answer. There are different degrees of subjectivity in testing. A subjective test is evaluated by giving an opinion. Subjective tests can be of various types such as-

- Short Answer Type
- Long Answer Type
- Essay Type
- Conversation or Problem Solving

Strengths of Subjective Tests-

- Subjective tests are used to test ideas, culture, coherence and creativity.
- They aim to give an overall evaluation of the student's mastery of a language.
- These tests give students more freedom to express ideas, opinion and use synthesizing skills to change knowledge into creative ideas.

Limitation of Subjective Test:

- Subjective tests take much time to evaluate.
- These tests require more individual consideration of teacher.

Either essay or objective tests can be used to:

- Measure almost any important educational achievement a written test can measure
- Test understanding and ability to apply principles.
- Test ability to think critically.
- Test ability to solve problems.

Multiple Choice Item

Multiple – choice item (MCI) is one of the most widely used type of items in objective test.(multiple) can provide a useful means of learning and testing in various learning situations provided that it is always recognized that such items , test knowledge of grammar, vocabulary, etc. rather than the ability to use language.

Multiple – choice items (MCI): is usually divided into three groups or parts:

1- The Stem
2- The correct choice or correct answer
3- The distracters.

1. The Stem, the initial part of each multiple – choice items is known as the stem. It can be completed statement, an incomplete statement and question.
2. The correct choice or correct answer, or correct option or key. The answer can be a word or a group of words.
3. The distracters, incorrect options or incorrect answer. The distracters can be two or three or four options.

Example

- *I've been living here …… 1986.'*
    a. *For*
    b. *On*
    c. *In*
    d. *At*
    e. *Since*

Suggestions For Writing Multiple-Choice Test Items

The Stem

1. When possible, state the stem as a direct question rather than as an incomplete statement.

Undesirable: *Alloys are ordinarily produced by...*

Desirable: *How are alloys ordinarily produced?*

2. Present a definite, explicit and singular question or problem in the stem.

Undesirable: *Psychology...*

Desirable: *The science of mind and behavior is called...*

3. Eliminate unnecessary verbiage or irrelevant information from the stem.

Undesirable: *While ironing her formal, Jane burned her hand accidently on the hot iron. This was due to transfer of heat by...*

Desirable: *Which of the following ways of heat transfer explains why Jane's hand was burned after she touched a hot iron?*

4. Include in the stem any word(s) that might otherwise be repeated in each alternative.

Undesirable: *In national elections in the United States the President is officially*

a. chosen by the people.

b. chosen by members of Congress.

c. chosen by the House of Representatives.

**d. chosen by the Electoral College*

Desirable: *In national elections in the United States the President is officially chosen by*

a. the people.

b. members of Congress.

c. the House of Representatives.

**d. the Electoral College*

5. Use negatively stated stems sparingly. When used, underline and/or capitalize the negative word.

Undesirable: *Which of the following is not cited as an accomplishment of the Kennedy administration?*

Desirable: *Which of the following is NOT cited as an accomplishment of the Kennedy administration?*

Areas of evaluation include:

1- Teacher evaluation

1. Personality
2. Familiarity with the subject.
3. Teaching methods used by the teacher.
4. The teacher's ability to manage and control the class.
5. Employment of educational aids.
6. His ability to deal with students and take into account their tendencies, desires and trends.
7. Measuring the educational outcomes achieved by the teacher.
8. The teacher's appearance, behavior and intellectual attitudes.

How to evaluate a teacher?

- **Students**: where students evaluate their teachers
- **Self**-evaluation
- **Peer evaluation**
- **Principal**
- **Supervisor**

2- Learner Evaluation

✓ Assessing the extent to which the educational goals have been achieved in this learner.
✓ The ability and the preparation of the student mentally, physically, etc.
✓ Student's performance

3- Curriculum evaluation

✓ Its **goals** in terms of their relevance to the philosophy of society and the general objectives of education.
✓ Evaluating its **content** in terms of its relevance to the goals and organizing its knowledge in an organized manner consistent with the knowledge structure and its development to keep pace with scientific, cognitive and social changes
✓ Evaluating the **methods** of teaching the curriculum and the extent of their compatibility with the objectives, content, nature of students, capabilities and contemporary changes in

teaching and learning.

4- Educational Administration Evaluation

Functions of Evaluation:

1. ***It measures student's achievement.***
2. ***It evaluates instruction.***
3. ***It motivates learning.***
4. ***It predicts success.***
5. ***It diagnoses the nature of difficulties.***

Importance of Evaluation:

i. Making of a number of decisions for example, selection, placement, promotion, decision making etc.

Evaluation Tools

- Observation
- Questionnaire
- Achievement test
- Interviews
- Referendum

Types of Achievement Test

- **Oral**
- **Written**
- **Objective**
- **Supply**
- **Short answer**
- **Completion type**
- **Selection**
- **Multiple choice**
- **Matching**
- **True or False**

- **Subjective**
- **Long essay**
- **Short essay**
- **Practical**

Examples

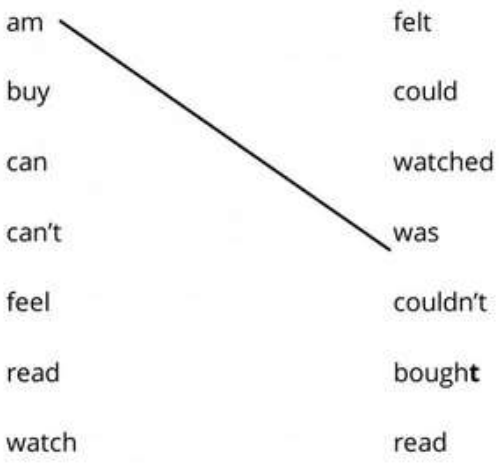
Short Answer: Who is the first president of Somalia?
Completion: A triangle with all sides and all angles equal is called……

Alternative/True or False: Somali took its independent in 1954 (…….)

Matching

1. Match the **present tense** on the left to the **past tense** on the right. Draw a line from present verb to past verb with online pen tool.

For example: am links to was.



Short Essay/Restrict Response:

- State three advantage of multiple choice questions and two disadvantage of multiple choice questions.

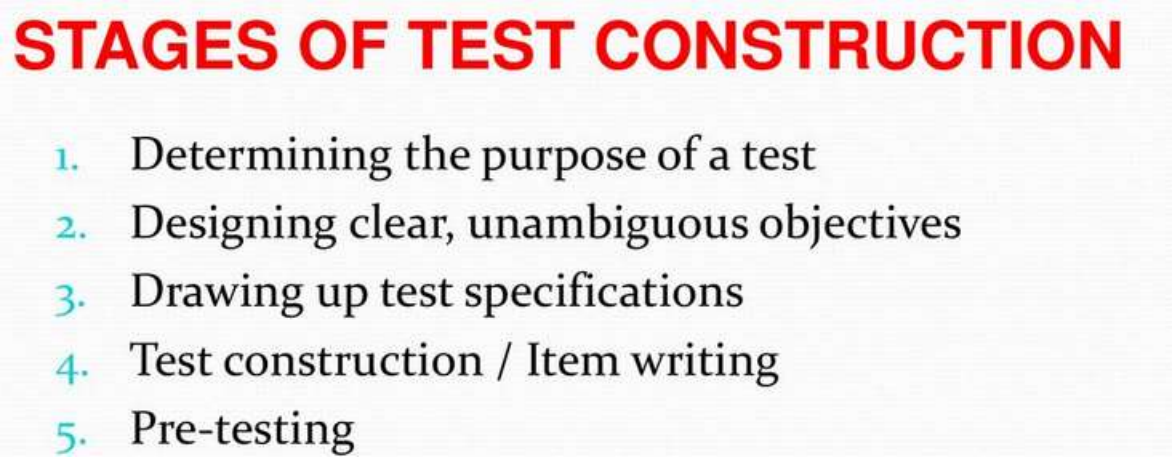
Long Essay/Extended Response:

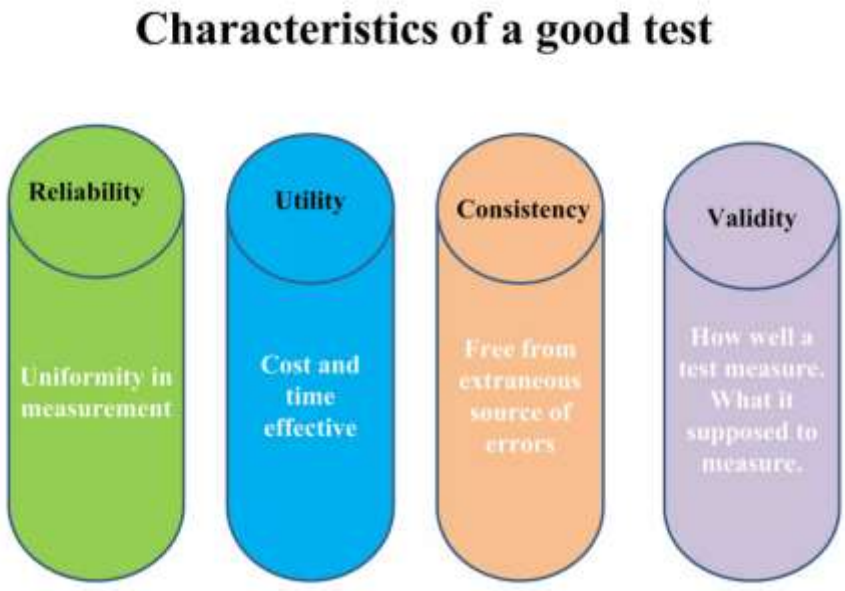
- Discuss the importance of the educational psychology in teaching and learning.

Stages of Test

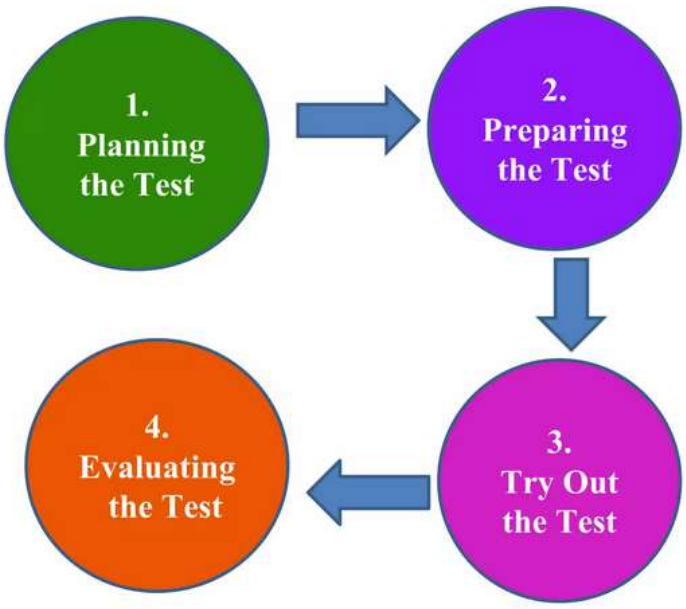
1. Test construction,
2. Test administration,
3. Scoring and
4. Analyzing the test.

Test Construction





4 main steps of Test Construction



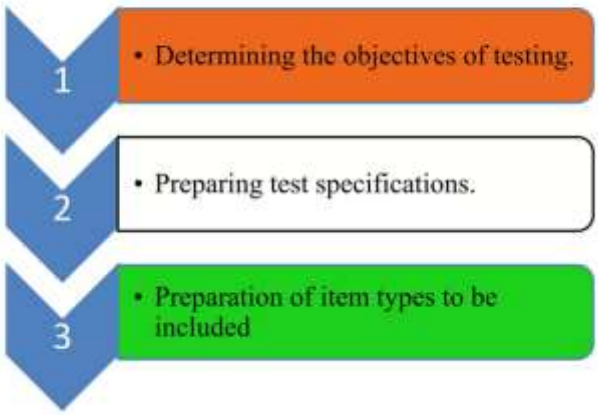
Planning

(1) What is to be measured?

(2) What content areas should be included and

(3) What types of test items are to be included.

Therefore the first step includes three major considerations.



1- Determining the objectives of testing

- ***a. Performance***
- ***b. Progress***
- ***c. Mastery level***

2- Preparing test specification

- Preparing Table of Specification or Blue Print
- It is three dimensional chart shows list of instructional objectives, content areas and type of items in its dimensions.

It includes four major steps:

- **Determining the weightage of instructional objective**

Objectives	Weightage in %	No. of Questions (marks)
Knowledge	15%	15
Understanding	45%	45
Application	30%	30
Skill	10%	10
Total	100%	100

- Determining the weightage to different content areas

If a book contains 150 pages and 100 test/items (marks) are to be constructed then the weightage will be given as, following:

Topic No.	Topic page no.	No. of items/Marks	% of items/Marks
1	1 to 25	17	16.7%
2	26 to 75	33	33.3%
3	76 to 110	23	23.3%
4	111 to 150	27	26.7%
	Total	100	100%

3- Preparing of item types to be included.

- **Objective**
- **Subjective**

4- Preparing of the Table of Specification

Class: VI
Subject: Science

BLUEPRINT

Time: 80 min
M.M: 40

Objectives / Contents	Knowledge (10%)			Understanding (52.5%)			Application (37.5%)			Total
	VSA	SA	LA	VSA	SA	LA	VSA	SA	LA	
Structure of plants					3(2)	5(2)	1(4)			20(8)
The Leaf		3(1)					1(2)	3(1)		8(4)
The Flower	1(1)					5(1)		3(2)		12(4)
TOTAL	4(2)			21 (5)			15(9)			40(16)

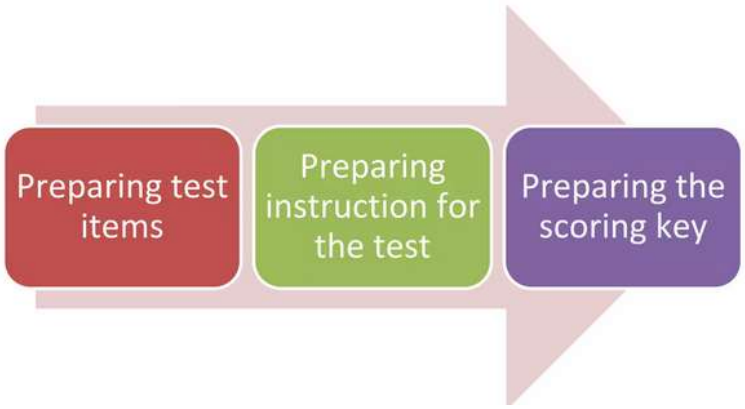
VSA: very short Answers
SA: Short Answers
LA: Long Answers

- Number in bracket shows number of questions
- Number outside the bracket shows marks given to each questions

- **Steps Three: Preparing the test**

After planning, preparation is the next important step in the test construction. In this step the test items are constructed in accordance with the table of specification. Each type of test item need special care for construction.

The preparation stage includes the following three functions:



Preparing the Test Items

- **The following principles help in preparing relevant test items:**
    1. Test item must be appropriate for the learning outcome to be measured
    2. Test item should measure all types of instructional objectives and the whole content areas.
    3. The test item should be free from ambiguity
    4. The test should be appropriate difficult level
    5. Arrange the questions in the order of their difficulty level.
    6. The test item must be free from technical errors and irrelevant clues
    7. The test item should be free from ethical bias

Preparing Instructional for test

Test maker should provide clear direction about

- Purpose
- Time
- Basis
- Procedure
- Methods

Marks Allocation to Test Items

- A **test score** is a piece of information, usually a number, that conveys the performance of an examinee on a **test**.

When allocating marks to the test items, consider the following:

- **Weightage to objectives**
- **Weightage to content**
- **Weightage to form of questions**
- **Weightage to difficulty level.**

This indicates what objectives are to be tested and what weightage has to be given to each objectives.

Sl.No	Objectives	Marks	Percentage
1	Knowledge	3	12
	Understanding		
2		2	8
3	Application	6	24
4	Analysis	8	32
5	Synthesis	4	16
6	Evaluation	2	8
Total		**25**	**100**

Weightage to content

This indicates the various aspects of the content to be tested and the weightage to be given to these different aspects.

Sl.No	Content	Marks	Percentage
1	Sub topic – 1	15	60
2	Sub topic – 2	10	40
Total		25	100

Weightage to form of questions

This indicates the form of the questions to be included in the test and the weightage to be given for each form of questions.

Sl.No	Form of questions	No. of Questions	Marks	Percentage
1	Objective type	14	7	28
2	Short answer type	7	14	56
3	Essay type	1	4	16
Total		**22**	**25**	**100**

Weight to Difficult Level

This indicates the total mark and weightage to given to different level of questions

Sl.No	Form of questions	Marks	Percentage
1	Easy	5	20
2	Average	15	60
3	Difficult	5	20
Total		**25**	**100**

Objectives / Form of Qtn / Content	Knowledge			Under-standing			Application			Analysis			Synthesis			Evaluation			Grant Total
	O	SA	E	O	SA	E	O	SE	E	O	SA	E	O	SA	E	O	SA	E	
Sub Topic - 1	2 (4)			1 (2)			2 (4)	2 (1)				4 (1)		2 (1)			2 (1)		15
Sub Topic – 2	1 (2)			1 (2)				2 (1)			4 (2)			2 (1)					10
Total Marks	3	0	0	2	0	0	2	4	0	0	4	4	0	4	0	0	2	0	25
Grand Total	3			2			6			8			4			2			

Test Administration

- Good testing practices rest in the hands of the examiner, who should ensure the testing exercise, runs smoothly.
- The period before the test, during the test and after the test should be effectively managed to realize a highly efficient testing period.

Period before Test

- Security of testing instruments: All test materials used in the assessment process, whether paper-and-pencil or computer-based must be kept secure.
- Test security is also a responsibility of test developers to ensure the test is not compromised over time.
- E.g: keep testing materials in locked rooms or cabinets and limit access to those materials to staff involved in the

- Examinees and parents have been notified regarding the test date and time.
- Candidates have been reminded to bring materials necessary for the test.
- All students with special needs (e.g. glasses and hearing aids) have been considered before the start of the test.
- All adequate invigilation has been planned.
- Examination administrators have read appropriate test administration procedures such as timing, examination regulations and test modifications.

- The rooms where the test is to be conducted have adequate ventilation and lighting and have been properly arranged.
- Seats are arranged in such a way that candidates cannot look at each other's work.

During the Test

- Observe precision in giving instructions or clarifications.
- Avoid interruptions.
- Avoid giving hints to students who ask about individual items.
- Discourage cheating.
- Ensure that no eating takes place in the examination hall.
- Identify each examinee to prevent a situation where someone may attempt to take the examination on someone else's behalf.
- Handle emergencies appropriately.
- Inform students on progress of testing.

After the Test

- Orderliness is needed for a successful testing process until all the test materials are securely in the hands of the test administrators.
- After the completion of the examination, the following are expected;

a) All test materials and documents, both used and unused should be collected and accounted for. They may be kept in a secure and lockable facility.

b) Count through candidates' scripts to ensure their number corresponds with the names on the examination attendance register. Counting also eliminates scenarios where the attendance register shows a student attended an examination but his or her script is not available.

To realize a smooth test administration exercise, the period before, during and after the test should be carefully managed.

4. The test items should be of appropriate difficulty level: The items should not be so easy that everyone answers it correctly and also it should not be so difficult that everyone fails to answer it. The items should be of average difficulty level.

5. The test item must be free from technical errors and irrelevant clues:

For example: grammatical inconsistencies, verbal associations, extreme words (ever, seldom, always), and mechanical features (correct statement is longer than the incorrect). Therefore while constructing a test item careful step must be taken to avoid most of these clues.

6. Test items should be free from racial, ethnic and sexual biasness:

While portraying a role all the facilities of the society should be given equal importance. The terms used in the test item should have an universal meaning to all members of group.

VALIDITY AND RELIABILITY OF TEST

Validity

Validity of tests means that a test measures what it is supposed to measure or a test is suitable for the purposes for which it is intended. There are different kinds of validity include content validity, face validity, criterion-referenced validity, and predictive validity.

- **Content Validity** evaluates how well a test covers all relevant parts of the construct it aims to measure.
- **Face Validity** depends on the judgment of the external observer of the test. It is the degree to which a test appears to measure knowledge and ability based on the judgment of the external observer.
- **Criterion Validity** (or **criterion-related validity**) evaluates how accurately a test measures the outcome it was designed to measure. For example, the English test may be compared with the Centralized English test. If the correlation is high, i.e. r = 0.5 and above, we say the validity of the English test meets the criterion referenced, i.e., the Centralized English test.
- **Predictive Validity** suggests the degree to which a test accurately predicts future performance. For example, suppose we assume that a student who does well in a particular mathematics aptitude test should be able to undergo a physics course successfully. In that case, predictive validity is achieved if the student does well in the course.

Factors Affecting Validity

- Cultural beliefs
- Attitudes of testees
- Values
- Maturity
- Atmosphere
- Absenteeism

Reliability

The reliability of test scores is the extent to which they are consistent across different testing occasions.If candidates get similar scores on parallel tests, this suggests the test is reliable.

Methods of Estimating Reliability
Some of the methods used for estimating reliability include:

- **Test-retest method:** An identical test is administered to the same group of students on different occasions.
- **Alternate – Form method:** Two equivalent tests of different contents are given to the same group of students on different occasions. However, it is often difficult to construct two equivalent tests.

- **Split-half method** A test is split into two equivalent subtests using odd and even numbered items. However, the equivalence of this is often difficult to establish.

Reporting Test Result

Report

A document containing information organized in a narrative, graphic, or tabular form, prepared on periodic, recurring, regular, or as required basis.

Test Report

A test report is a document that contains

- A summary of test activities and final test result.
- An assessment of how well the testing is performed.

Based on the Test Report, the stakeholders can:

- Evaluate the quality of the tested report
- Make a decision on the software release. For example, if the test report performs that there're many defects remaining in the product, the stakeholder can delay the release until all the defects are fixed.

Report Test Results to Administration

At the school administration level, educators create tests to measure their students' understanding of specific content or the effective application of critical thinking skills. Such tests are used to evaluate student learning, skill level growth and academic achievements at the end of an instructional period, such as the end of a project, unit, course, semester, program or school year

School administration can take data from almost any dimension attendance, fundraising, standardized testing, you name it—and turn it into a report. Here's how reports will make it easier to run and track your school's achievements.

The test report aim to provide an account of the effectiveness of school work and student work to administration and the extent to which the school's development targets are achieved. The school should make use of the data and evaluation findings to inform future planning and puts continuous improvement in action. The report should be evidence-based and data-driven, and truly reflect the school's and students achievements and areas for improvement.

The test report also provides a channel for the school to report to administration on the major tasks in various areas, thus enhancing accountability.

The key elements of a test report include the following:

- Our School
- Achievements and Reflection on Major Concerns.
- Learning and Teaching
- Support for Student Development
- Student Performance.
- Financial Summary
- Feedback on Future Planning
- Appendix