

Assessment & Evaluation
ICT 6th Semester 2076 Notes
2076
by

Mukesh Singh
ICT 5th batch 2073

Sukuna Multiple Campus
Sundarbarancha -12, Morang

Unit-1

Assessment & Evaluation

Class _____

Page No. _____
Date _____

1.1 Concept of test, measurement, assessment and evaluation

A. Concept of Test

- Test is a systematic procedure for obtaining data or measurements on a particular behaviour.
- A test is defined as a group of questions or tasks to which learners are asked to respond orally, in writing or sometimes even in pantomime (एकांकिक).
- They may appear at any point in a course. Feedback from tests is usually to help learners by identifying weakness.
- Test is an instrument for measurement. It measures the level of skill or knowledge that has been reached.

Few definitions of Test:

- i) "A test is an organized succession of stimuli designed to measure quantitatively or to evaluate qualitatively some mental processes, traits and characteristics". - Bean
- ii) "Test is defined as a series of question on the basis of which some information is sought." - A.K. Singh
- iii) According to Oxford Advanced Learners' Dictionary :
"Test is examine and measure the qualities of sb or sth."
- iv) "Test is a tool to measure language proficiency of students".
- Hughes.

Although it may seem that all tests are the same, many different types of tests exist and each has a different

Purpose and style.

- Diagnosis tests
- Placement tests
- Progress or achievement tests
- Proficiency tests
- Internal tests
- External tests
- Objective tests
- Subjective tests

B. Concept of Measurement

The process of obtaining a numerical description of the degree to which an individual possesses a particular characteristic.

- Test is used to gather information.
- That information is presented in the form of measurement.
- That measurement is then used to make evaluation.

In education, the numerical value of scholar's ability, aptitude, achievement etc. can be measured and obtained using instruments such as paper and pencil test. It means that the values of the attribute are translated into numbers by measurement.

Some definitions of Measurement:

i) "Measurement is the assignment of numerals to objects and events according to rules." - S. Stevens

ii) "Measurement consists of rules for assigning numbers to

object in such a way as to represent qualities of attributes".
- J. C. Nunnally

Types of measurement:

a) Physical measurement → a measurement made by comparing a quantity with a standard unit.

b) Mental/Psychological measurement → the values of attributes are translated into numbers by measurement.

Physical Measurement

1. It is an absolute measurement.
2. The reference point is zero.
3. It has definite or certain order.
4. The trait is directly measured.
5. It is perfectly objective and valid.
6. It has great precision.

Mental/Psychological measurement

1. It is a relative measurement.
2. The reference point is group performance.
3. It has recognizable order.
4. The trait is indirectly measured with the help of behaviours.
5. It is a subjective measurement but tries to make it objective and valid.
6. It has less precision.

C. Concept of Assessment

Assessment is a process by which information is obtained relative to some known objective or goal.

It is the systematic process of documenting and using the empirical data on the knowledge, skills, attitudes and beliefs. By taking assessments, teachers try to improve student learning.

A test is a special form of assessment. All tests are assessments but not all assessments are tests.

Some Definitions of Assessment:

- i) "Assessment involves the use of empirical data on student learning to refine programs and improve students learning." - Allen (2004)
- ii) "Assessment is the systematic basis for making inferences about the learning and development of students". - Enwin (1991)

D. Concept of Evaluation:

Evaluation is the process of making judgements based on criteria and evidence.

In education, evaluation is the process of using the measurements gathered in the assessments. Teachers use this information to judge the relationship between what was intended by the instruction and what was learned. They evaluate the information gathered to determine what students know and understand, how far they have progressed and how fast, and how their scores and progress compare to those of other students.

Some definitions of evaluation:

i) "Evaluation is the process of determining to what extent the educational objectives are being realized."

- (Ralph Tyler, 1950)

ii) "Evaluation is a systematic process of collecting, analyzing and interpreting information to determine the extent to which pupils are achievement instructional objectives."

- Norman E. Grunlund and Robert L. Linn

Principles of Evaluation

iii) "Evaluation is a process of judging the value or something by certain appraisal." - Goods

Principles of Evaluation:

Evaluation is based on the following principles:

1. Principle of Continuity → Evaluation is a continuous process, which goes on continuously as long as the learner student is related to education. Whatever the child learns, it should be evaluated daily. Only then the learner could have better command on language.
2. Principle of Comprehensiveness → By comprehensiveness we mean to assess all aspects of the learner's personality. It is concerned with all round development of the child.
3. Principle of Objectives → Evaluation should be based on the objectives of education. It should be helpful in finding out where there is a need for redesigning and refining the learner's behaviour.
4. Principle of Learning Experience → Evaluation is also related to the learning experiences of the learner. In the process, we don't evaluate only the curricular activities of the learner but his co-curricular activities are also evaluated.
5. Principle of Broadness → Evaluation should be broad enough to cover all the aspects of life.
6. Principle of Child-Centredness → Children in the center, in the process of evaluation. The behaviour of child is the central point of assessment. It helps a teacher to know the grasping power of a child and usefulness of teaching.

material.

7. Principle of Application → Evaluation judges that student is better to apply her knowledge and understanding in different situations in order to succeed in life.

1.2 Types of evaluation : purpose, tools and uses

1.2.1 Diagnostic evaluation

This type of evaluation is concerned with finding out the reasons for students' persistent or recurring learning difficulties that cannot be resolved by standard corrective measures or formative evaluation.

The aim of diagnostic evaluation is to find out the causes of learning problems and plan to take remedial action.

1.2.2 Placement

In this type of evaluation, learners entry behaviour or capability is assessed to find out whether the student possess knowledge, skills and attitude needed to begin the course of instruction.

1.2.3 Formative

It is the evaluation used to monitor students learning progress during instruction with the purpose of providing on going feedback to students and teachers regarding success and failure of teaching / learning process.

1.2.4 Summative

This type of evaluation is given at the end of the course or unit of instructions to find out which student, to what extent has mastered the intended learning outcomes.

Types of Evaluation	Purpose	Tools	Uses
1. Placement Evaluation	<ul style="list-style-type: none"> To evaluate the pupil's entry behaviour in a sequence of instruction. To determine the level or position of the child in the instructional sequence. 	<ul style="list-style-type: none"> Entrance Exam Interviews Observation 	
2. Diagnostic Evaluation	<ul style="list-style-type: none"> To determine student's individual strengths, weaknesses, knowledge and skills. To collect data on what students already know about the topic 	<ul style="list-style-type: none"> Self assessment Interviews Test 	
3. Formative Evaluation	<ul style="list-style-type: none"> To improve students learning. To develop knowledge and understanding. 	<ul style="list-style-type: none"> Classwork Homework Unit test Oral question 	
4. Summative Evaluation	<ul style="list-style-type: none"> To evaluate students achievements. To assess knowledge and understanding at a given point in time. 		<ul style="list-style-type: none"> Final exam

Unit - 2

Characteristics of a Test

Page No. _____
Date. _____

2.1. Essential qualities of a test

2.1.1. Reliability

Reliability refers to the extent to which an assessment yields consistent information about the knowledge, skills, or abilities being assessed. This refers to the extent to which they obtained results are consistent or reliable. An assessment is considered reliable if the same results are yielded each time the test is administered. For example, if we took a test in History today to assess our understanding of World War I and then took another test on World War I next week, we would expect to see similar scores on both tests. This would indicate the assessment was reliable.

2.1.2. Validity

It refers to extent to which the test measures what it intends to measure. For example, when an intelligent test is developed to assess the level of intelligence, it should assess the intelligence of the person, not other factors. It means that it measures what it is supposed to measure. It tests what it ought to test. A good test which measures control of grammar should have no difficult lexical items. Validity explains us whether the test fulfills the objective of its development. There are many methods to assess validity of a test.

2.1.3. Objectivity

Objectivity is an important characteristic of a good test. It affects both validity and reliability of test scores. A test is objective when it makes for the elimination of the answerer's personal opinion bias judgement. The recognition of quality objectives in a test has been largely responsible for the development of an arched and objective type tests.

Objective based tests ~~are~~ measure or evaluate the entire human development in three domains that is cognitive, affective and psychomotor. As name suggest itself indicates they are based on particular objective of teaching and evaluating. They provide the proper direction and thus streamline the whole process of evaluation. These tests are all comprehensive.

2.1.4. Usability

Usability means the degree to which the tests are used without much expenditure of money time, money, and effort. It also means practicability. factors that determine usability are : administrability, Scorability, interpretability, economy and proper mechanical makeup of the test.

2.2 Methods of estimating reliability

2.2.1 Test-retest

This is the simplest method of determining the test reliability. To measure test-retest reliability, you conduct the same test on the same group of people at two different points in time. Then you calculate the correlation between the two sets of results.

You use it when you are measuring something that you expect to stay constant in your sample.

Example

Eg 1:- A test of colour blindness for trainee pilot applicants should have high test-retest reliability, because colour blindness is a trait that does not change over time.

Eg: In a questionnaire to measure the IQ of a group of participants, if you administer the test two months apart to the same group of people, but the results are significantly different, so test-retest reliability of the IQ questionnaire is low.

Note - The smaller difference between the two sets of results, the higher the test-retest reliability.

2.2.2 Parallel form

Reliability of test scores can be estimated by parallel form method. Parallel forms reliability measures the correlation between two equivalent versions of a test. You use it when you have two different assessment tools or set of questions designed to measure the same thing.

The most common way to measure parallel forms reliability is to reproduce a large set of questions to evaluate the same thing, then divide these randomly into two question sets.

The same group of respondents answer both sets, and you calculate the correlation between the results. High correlation between the two indicates high parallel forms reliability.

Parallel forms of reliability means that, if the same students take two different versions of a reading comprehension test, they should get similar results in both tests.

Example

A set of questions is formulated to be given to a group of respondents. The questions are divided into two sets and the respondents are randomly divided into two groups. Both groups take both tests: group A takes test A first, and group B takes test B first. The results of two tests are compared, and the results are almost identical, indicating high parallel forms reliability.

2.2.3. Split-halves

The split-half method is a quick and easy way to establish reliability. However it can only be effective with large questionnaires in which all questions measure the same construct. This means it would not be appropriate for tests which measure different constructs.

This method is done by comparing the results of one half of a test with the results from other half. A test can be split in half in several ways e.g. first half and second half, or by odd and even numbers. If the two halves of the test provide similar results this would suggest that the test has internal reliability. The reliability of test could be measured using this method.

(1,3,5)
For example, taking all odd numbers in one half and all even-number (2,4,6,8 etc) in the other half. Then the scores of both the halves are correlated by using the Spearman-Brown formula.

$$\gamma_2 = \frac{2r_1}{1+r_1} - (5.1)$$

Where

r_1 = Reliability coefficient on full test.

r_1 = Correlation coefficient between halves.

For example by correlating both the halves, we found a coefficient of 70.

The reliability coefficient is 82 when the coefficient of correlation between half tests is 70. It indicates the to what extent the sample of the test items are dependable on sample of the content being measured.

2.2.4 Kuder-Richardson method

This method is also known as Rational Equivalence method or Inter-item method.

This is a method based on single administration.

Like split-half method, this method also provides a measure of internal consistency. It neither requires administration of two equivalent forms of tests nor it requires to split the tests into two equal halves.

Reliability coefficient is determined by using the Kuder-Richardson-21 formula (KR-21) which is given below:

KR-21 :

$$\gamma_{11} = \frac{n}{n-1} \times \left[1 - \frac{\sum pq}{\sigma_t^2} \right]$$

Where,

γ_{11} = Reliability Coefficient of the whole test

n = Number of items in the test.

σ_t^2 = Variance of the total test ($\sigma_t = S.D.$)

$p = \frac{\text{Number of persons answering item correctly}}{\text{Number of persons taking test}}$

$$q = (1-p) = \frac{\text{Number of persons answering items incorrectly}}{\text{Number of persons taking test}}$$

Σ = Summation sign indicating the pq is summed over all items

$$q = -p$$

$$p = 1 - q$$

Note

- Kuder-Richardson method and split-half method are appropriate for speed tests.
- Both Kuder-Richardson and split-half method do not measure the consistency of pupil response from day to day.
- Factors Affecting Reliability:

1. Factors related to test

(a) Length of test → Spearman Brown formula indicates the longer test is, the higher the reliability will be because the longer test provides adequate sample of the behaviour.

(b) Content of the test → According to Guilford homogeneity of test, content also increases the reliability of test.

(c) Characteristics of items → If test items are too easy or difficult for the group members, it will tend to produce scores of low reliability.

(d) Spread of scores → Larger/greater the spread of score the higher is the estimate of reliability will be.

2. Factors related to testee

(a) Heterogeneity of the group → The reliability coefficient for a heterogeneous group will be more than homogeneous group.

(b) Test wisdom of the students → Practice of students in taking the sophisticated tests increases the reliability. But when in a

group, all the students do not have same level of test burden, it leads to greater measurement errors.

(c) Motivation of the students → When the students are not motivated to take test, they will not represent their best achievement. This depresses the test scores.

(d) Factors related to testing procedures

(a) Time limit of test → When students get more time to take the test, they can make more guessing, which may increase the test scores. Therefore, speeding up a test increases the test reliability.

(b) Cheating opportunity given to the students
Cheating by the students leads to the measurement errors. Students copying correct answers from cheat sheets or listening to other students without knowing correct answers may get higher score than they actually deserve. They will make the observed score of cheater higher than their true score.

2-3 Types of Validity

Validity refers to the extent to which the test measures what it intends to measure.

There are following types of validity:

2.3.1. Content Validity

Content validity is a non-statistical type of validity that assesses whether a test is representative of all aspects of the construct.

To produce valid results the content of a test, survey or measurement method must cover all relevant parts of the subject it aims to measure. If some aspects are missing from the measurement, the validity is threatened.

Example Example

A mathematics teacher takes an algebra test at the end of the semester. The test should cover every form of algebra that was taught in the class. If some types of algebra are left out, then the result may not be accurate. Similarly, if he includes questions that are not related to algebra, the results are no longer a valid measure of algebra knowledge.

2.3.2. Criterion: concurrent and predictive

Criterion validity evaluates how closely the results of your test correspond to the results of a different test. (The criterion is an external measurement of the same thing. It is usually an established or widely-used test that is already considered or valid).

To evaluate the criterion validity, you calculate the correlation between the results of your measurement and the results of the criterion measurement. If there is a high correlation this gives a good indication that your test is

measuring what it intends to measure.

Example

for example, employee selection tests are often validated against measure of job performance (criterion), and if tests are often validated against measure of academic performance (the criterion).

Criterion validity is split into two types of validity:

i) ~~Reliability~~ Concurrent Validity

If the test data and criterion data are collected at the same time, this is referred to as concurrent validity, evidence.

Concurrent validity refers to the extent to which the results and conclusions concur with other studies and evidence.

ii) Predictive Validity

If the test data are collected first in order to predict criterion data collected at a later point in time, then this is referred to as predictive validity evidence.

Predictive validity refers to the extent to which the results and conclusions can be used to predict real life applications of the study.

2.2.3. Construct Validity

Construct validity refers to the extent to which a study or test measures the concept which it claims to. Construct validity evaluates whether a measurement tool really represents the thing we are interested in measuring.

[A construct refers to a concept (such as gender equality, freedom of speech etc.) or characteristic (such as intelligence, obesity, depression etc.) that can't be directly observed, but can be measured by observing other indicators that are associated with it]

Example of Construct Validity:

If you develop a questionnaire to diagnose depression, you need to know: does the questionnaire really measure the construct of depression? Or is it actually measuring the respondent's mood, self-esteem, or some other construct?

To achieve the construct validity, you have to ensure that your indicators and measurements are carefully developed based on relevant existing knowledge. The questionnaire must include only relevant questions that measure known indicators of depression.

Construction of Teacher Made Test

3.1. Concept of teacher made test

To measure the achievement level of students, tests are constructed by teachers for their classroom consumption, which is referred as teacher made test.

They are made by the teacher himself to measure the achievement of his pupil from time to time.

Teacher made tests are more than assessment devices. They are fundamental part of the educational process. They can define instructional purposes, influence what students study, and help instructors to gain perspective on their courses.

Teacher made tests are normally prepared and administered by for testing classroom achievement of students, evaluating the method of teaching adopted by teacher and other curricular programmes of the school.

Teacher-made test is one of the most valuable instrument in the hands of the teacher to solve his purpose. It is designed to solve the problem or requirements of the class for which it is prepared.

Features of teacher-made test

- i) The test items are arranged in order of difficulty.
- ii) These are prepared by teacher which can be used for prognosis and diagnosis purposes.
- iii) The test covers the whole content area and including larger number of items.
- iv) The preparation of the items conforms to the blueprint.

- PAGE NO. _____
Date _____
- v) A teacher-made test does not cover all the steps of a standardized test.
 - vi) Teacher-made tests may also be employed as a tool for formative evaluation.
 - vii) Preparation and administration of these tests are economical.
 - viii) Teacher-made tests are least used for research purposes.
 - ix) Teachers can understand the concept of re-teaching concepts and can decide remedial instruction.

Uses of teacher-made tests:

- i) Basically teacher-made tests are used to evaluate the progress of the students in school. However, the specific use of tests may vary from school to school and teacher to teacher.
- ii) To help the teacher whether the class is normal, average, ~~or~~ above average or below average.
- iii) To help in formulating new strategies for teaching and learning.
- iv) To measure ~~the~~ students' academic achievements in a given course.

- v) To assess how far the specified instructional objectives have been achieved.
- vi) To know the efficacy of learning experience.
- vii) To diagnose students learning difficulties and to suggest necessary remedial measures.
- viii) To certify, classify or grade the students on the basis of resulting scores.
- ix) These tests can be used as a tool for formative, diagnostic and summative evaluation.

3.3. Types of test items

Classroom test and assessments play a central role in the evaluation of student learning. The main goal of classroom testing and assessment is to obtain valid, reliable and useful information concerning student achievement.

There are different types of test items to check the learning process:

3.3.1. Subjective test : types, construction and uses

A subjective test is evaluated by giving an opinion. It can be compared with an objective test, which has right or wrong answers and so can be marked objectively. But subjective tests are more challenging and expensive to prepare, administer and evaluate correctly, but they can be more valid.

General characteristics of subjective test item:

- It include short-answer and essay items.
- Also known as "free-response", "constructed response" and "supply-type" items.

Types

There are two types of subjective tests:

1. Short-answer Items

Short answer type questions are the type, that can be answered by a word or a few sentences.

- It uses a direct question.
- It comes between objective and essay ~~items~~ type.

- It can be answered in few sentences.
- It is highly thought provoking.
- It covers a wide range of content.
- It contains objectives individually like knowledge, understanding, synthesis, application, analysis and evaluation.

Construction / Suggestions for short answer items

- Avoid irrelevant clues.
- Avoid long/complex sentences.
- Use your own words.
- Use direct questions.

User of short answer items

- It is suitable for measuring wide variety of relatively simple learning outcomes.
- Used in interpreting diagrams, charts etc.

Advantages

- It covers a wide range of content.
- It can be answered in few sentences.
- It is easy to construct, because it measures simple learning outcomes.

Limitations/ Disadvantages

- Used only with lower level cognitive skills.
- Writing skill cannot be measured properly.
- It cannot test the expression ability of students.

2 Essay Items Test Items

The essay test is ~~the~~ probably the most popular of all types of teacher-made tests. It is a test item that requires the student to structure a rather long written response upto several paragraphs.

- It demands long answers.
- Allows freedom of response to a problem.
- Contains fewer questions than objectives and short answer questions.
- Student get much freedom to express his ideas.

Construction / Suggestions for Essay Test Items

- Prepare ~~essay~~ essay items that elicit the type of behaviour you want to measure.
- The question should give clarity.
- Specify time limits for thinking and writing.
- Use words like 'compare' or 'contrast' at the beginning of the question.

Advantages

- Easy to construct.
- ~~Less time consuming~~.
- Brings language mastery.
- Test pupil's ability to use knowledge.
- Reduce chance of on-the-spot copying.

Disadvantages

- It covers only few areas.
- It is of a time consuming type.
- Lack of reliability

3.3.2. Objective test items: types, construction and uses

An objective test item is a test that has right or wrong answers and so can be marked objectively.

These are questions that require a specific answer.

An objective question has only one potential correct answer and they leave no room for opinion.

An objective test is so named because system of scoring is objective rather than subjective.

Types of Objective Test items

1. Multiple Choice items

Multiple choice items is a form of an objective assessment in which respondents are asked to select only correct answers from the choices offered in a list.

Multiple choice items are composed of one question (the stem) and with multiple possible answers (choices), including the correct answer and several incorrect answers (distractors).

Construction / Suggestions for making multiple choice items

- The item should be clear, accurate and unambiguous.
- The item should give either "direct" or an incomplete statement.
- The item should have only one correct answer.
- Make alternatives approximately equal in length.

Advantages of Multiple choice items

- Can provide versatility in measuring all levels of cognitive ability.
- Provides highly reliable test scores.
- Provides scoring efficiency and accuracy.
- Provides a wide sampling of content or objectives.

Disadvantages / limitations

- Are difficult and time consuming to construct.
- Lead an instructor to favor simple recall of facts.

2. ~~Not~~ True/false Test Items :

True/false questions are only composed of a statement. Students respond to the questions by indicating whether the statement is true or false.

For example: True/false questions have only two possible answers.

Construction / Suggestions for making True/false test items

- The statement should not be used with out qualification.
- The statement should not be very long.
- The statement should not use words like always, never, no, only (false), sometimes & often (* ~~false~~ true).
- The statement should avoid negative statements.

Advantages of True/False test items

- Provides the widest sampling of content or objectives per unit of testing time.
- Provides scoring efficiency and accuracy.
- Provides highly reliable test scores.

Limitations/Disadvantages

- Can often include more irrelevant clues than do other item types.
- Can often lead an instructor to favour testing of trivial knowledge.

3. Matching Test Items

In general, matching items consist of a column of stimuli presented on the left side of the exam page and a column of responses placed on the right side of the page. Students are required to match the response associated with a given stimulus.

Construction Guidelines for Matching Items:

- The statement should give clear directions.
- The statement should contain small premises and responses.
- The statement should contain homogeneous premises and responses.

Advantages of Matching items,

- Requires short period of reading and response time.
- Provides highly reliable test scores.
- Provides scoring efficiency and accuracy.

Limitations / Disadvantages

- Have difficulty in measuring learning objectives ~~too~~ requiring more than simple recall of information.
- Are difficult to construct.

4. Completion Test Items

The completion test item requires the students to answer a question or to finish an incomplete statement by filling in the blanks with the correct words or phrases.

Advantages of Completion Test Items.

- Can provide a wide sampling of content.
- Can efficiently measure lower levels of cognitive ability.
- Can minimize guessing as compared to ~~the~~ multiple-choice or true-false items.

Limitations / Disadvantages

- Are difficult to construct.
- Have difficulty in measuring learning objectives.
- Can often include more irrelevant clues.
- Are more time consuming.

Construction Guidelines for Completion Test Items

- Omit only significant words from the statement.
- Avoid grammatical or other clues to the correct response.
- Make the blanks of equal length.

3.4 Taxonomy of educational Objectives : Cognitive domain

The word Taxonomy derived from the Greek word 'taxis' which means systematic classification. Prof. Benjamin Bloom and his associate, University of Chicago developed and classified the domains of educational objectives. Bloom (1956) presented her taxonomy related to cognitive domain giving emphasis to the hierarchy of cognitive process in attending knowledge and development of thinking. Later Krathwol (1964) introduced affective domain and Simpson (1966) developed psychomotor domain. They described hierarchical development of the three domains of the learner through instruction. The classification of educational objectives is known as Bloom's taxonomy of educational objectives.

Classification of Bloom's taxonomy

1. Cognitive domain 2 Affective 3 Psychomotor

1. Knowledge	Receiving	Perception
- Comprehension	Responding	- Set
- Application	Valuing	- Guided Response
- Analysis	Organization	- Mechanism
- Synthesis	Characterization	- Complex Overt Response
- Evaluation		- Adaptation
		- Organization
		- Origination

i. Cognitive domain

The cognitive domain is known as development of thinking process. It is concerned with sensation, perception and application of knowledge. It includes objectives of knowledge which aims to develop mental and intellectual capacity of learners.

The hierarchical development of cognitive domain is discussed below:

i) Knowledge

Acquisition of knowledge is the lowest level of cognitive domain. It includes the ability of the students to recall and remember the information learned in the classroom.

Specifications - (recall and memorization)

Example - At the end of class student will be able to define the vertebrate animal.

ii) Comprehension

It is meaningful recall and recognition of the learned content. Here the learner could understand and explain what he learned in the classroom in his own language.

Specifications - (classification, explanations, comprehension, translations etc.)

Example - At the end of the class, students will be able to explain the features of Computer.

iii) Application

The learners are able to apply or use the knowledge which is acquired and comprehended during the first two levels. It is the ability to apply knowledge through instruction in real life situation.

Specifications - (use, solve, judge, compute, develop, operate)
Example - At the end of the class, students will be able to operate the Computer.

iv) Analysis

Analysis is the breakdown of the materials into various components and to identify the interrelationship between the elements and find out how they are organized and related. To study the content by classifying the content into different parts is considered as analysis.

Specifications - (analyze, categorize, contrast, separate etc.)
Example -

v) Synthesis

To accumulate or combine various learned knowledge, skill and experience is known regarded as synthesis (i.e. Combination of components or elements to form a connected whole.)

Synthesis is the meaningful ability of the learners to integrate the acquired, comprehended, applied and analyzed knowledge into a comprehensive whole. It involves the ability to give a new shape or structure to statements or procedures.

Specifications - (Combine, summarize, formulate, produce)
Example - At the end of, students will be able to plan the

Curriculum in ICT.

vi) Evaluation.

To evaluate whether the teaching of content is on the basis of objectives or not is called evaluation. It is the ability to judge value of materials, aspects, methods, principles, theory, philosophy for a given purpose.

Specifications - (Judge, Compare, evaluate, measure.)

2 Affective Domain (Krathwohl, 1964)

Affective domain is related with the development of heart and mind of the child. It includes areas of emotions, feelings, interest, attitude and values.

Bloom and Krathwohl (1964) introduced the following hierarchy of affective domain:

i) Receiving

Learner's interest of receiving or attending any stimulus, activity and events is called regarded as receiving. For example, students are listening interestingly listening to Gandhian principles.

Specification - (receive, agree, accept).

ii) Responding

The active participation/attention of learners in teaching-learning activities is called responding. The learner tries to respond to the situation positively. For example, students show kindness towards elders and weaker people, hold honest behaviour in day-to-day life situations.

Specifications - (present, respond, recognize, discuss)

iii) Valuing

By responding in good way, students set guidelines for their behaviour. Accepting values, preference for values, commitment to values are the important behavioural changes in this level. P

for example, students develop positive attitude towards non-violent behaviour, truthfulness, honesty, etc.

Specifications (value, change, prepare, help).

iv) Organization

To accumulate different values into one is called organization. Through organizing different values students are able to develop their own code of conduct and standard of public life in the society.

For example, pupil identifies the inseparability of the values like non-violence, truthfulness and tolerance of Indian tradition.

Specifications - (compare, relate, organize)

v) Characterization

To decide about various knowledge, skills, behaviour and activities of learning is called characterization. Values are imbibed and forms part of the life style of the individual.

for example - the non-violence value becomes the part of the philosophy of the individual.

Specifications - (change, complete, review, accept, purpose).

3. Psychomotor Domain (Simpson, 1972)

Psychomotor domain deals with the action or performance level. Psychomotor domain includes knowledge, skills and capacity which are learned by physiological procedures.

Simpson presented the psychomotor domain as follows:

i) Perception

To perceive is necessary to learn anything. It consists the process of becoming aware of objects, qualities or relation through sense organs.
(distinguish, confirm, determine, explain)

ii) Set / Preparation

To prepare for certain stimulus is called set. It includes mental, physical and emotional preparation.
(start, act, response, show)

iii) Guided response

It is the overt behavioural act of a student under the guidance of the teacher. Student initially
(start, act, response, show)

iv) Mechanism

Mechanism in learning means to perform the action as machine. Students can perform the task in somewhat confident, proficient and habitual manner.
(apply, change, make, prepare)

v) Complex Overt Response

In this level, students attain a high degree of skill and the act can be carried out smoothly and efficiently.

For example, students are able to write many words easily and simply without any within a short period of time.

A ability to run, walk, jump and talk easily also comes under this category.

vi) Adaptation

Application of learned things into the behavioural activities in terms of situation or condition is called adaptation.

(reconstruct, adopt, change, adjust).

vii) Organization

vii) Origination

Student is able to originate a new pattern of action or style in doing the activity.

3.5 Teacher made test : Construction Process

A teacher made test does not require a well-planned preparation. Even then, to make it more efficient and effective tool of evaluation, careful considerations are needed to be given while constructing such tests.

The following steps may be followed for the preparation of teacher-made test:

3.5.1 Planning the test

Planning of the test includes:

i) Writing instructional objectives

The instructional objectives of given course are normally determined from students' needs and involve new knowledge, new interest, new skills, and new understanding expressed in term of behaviour.

The determining of the relative importance of these objects depends on the teachers' personal judgement.

ii) Preparing specification chart

The Table of Specification is a testing blueprint that is basic to the construction of teacher made test.

The Table of Specification is a chart that shows the topics that will be covered on a test.

It enables the teacher to prepare a test containing a representative sample of student knowledge in each of the areas tested.

Sample of ToS

S.N.	Content outline	No. of items
1.	Table of Specifications	10
2.	Test and item characteristics	20
3.	Test layout	5
4.	Test instructions	5
5.	Reproducing the test	5
6.	Test length	5
7.	Scoring the test	5
	Total	55

3.5.2. Preparing the test

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

The preparation stage includes the following three functions:

i) Preparing test items

- Test items must be appropriate for the learning outcomes to be measured.
- Test items should measure all types of instructional objectives and the whole content area.
- Test items should be free from ambiguity.
- Test items should be free from technical errors and irrelevant clues.
- Test items should be free from racial, ethnic and

sexual biasness.

ii) Preparing instructions

This is the most neglected aspect of the test construction. Generally, everybody gives attention to the construction of test items. So test makers do not attach directions with the test items.

It is the function of the test instructions to furnish the learning experiences needed in order to enable each examinee to understand clearly what he is being asked to do?

N. E. Gronlund has suggested that the test maker should provide clear-cut directions about:

- The purpose of testing.
- The time allowed for answering.
- The basis for answering.
- The procedure for recording answers.
- The methods to deal with guessing.

iii) Preparing scoring key and marking scheme

A scoring key increases the reliability of a test so that the test maker should provide the procedure for scoring the answer scripts. Directions must be given whether the scoring will be made by a scoring key (when the answer is recorded on the test paper) or by a scoring stencil (when answer is recorded on separate answer sheet). And how marks will be awarded to the test items.

In case of essay type items it should be indicated whether to score with 'point method' or with the

'Rating' method. In 'point method', each ~~item~~ answer is compared with a set of ideal answers in the scoring key. Then a given number of points are assigned.

In the 'scoring method', the answers are rated on the basis of degree of quality and determine the credit assigned to each answer. Thus, scoring key helps to obtain a consistent ~~set~~ data about the pupil's performance. So the test maker should prepare a Comprehensive Scoring key procedure along with the test items.

Standardized Test

1. Prepared by a team of experts or specialists.
2. Standard uniform procedures for administration.
3. Used to evaluate content and objectives common to schools in locality or state.
4. Content and objectives are determined by ministry of education, curricula and syllabi.
5. They ^{are} highly reliable.
6. Quality of test items is known.
7. Scores can be compared to norm-
groups.

Teacher-Made Test

1. Prepared by the same person as instructor, writer and evaluator.

2. No uniform procedures for administration.

3. Used to evaluate the content and objectives specific to school locality or state, the class taught by the teacher.

4. Content and objectives are determined by the teacher in the classroom.

5. It is unreliable.

6. Quality of test items is unknown.

7. Scores can be compared only in the concerned class or school.

Unit - 4

Administration, Scoring & Analysis of Test

Ajanta

Page No. _____
Date _____

To establish academic development achievement in any public school, teachers have to construct, administer and score test or examinations. For a valid and reliable assessment of academic achievement, the test should yield similar situations and consistent results / grades for comparable groups of students. The validity and reliability of assessment are dependent on many factors among which are administrations of the measurement.

4.1 Conditions and administration of test

Test administration is concerned with the physical and psychological setting in which student take the test, for the students do their best (Ainscian) 1994.

Test administration procedures are developed for an exam program in order to help reduce measurement error and to increase the likelihood of fair, valid and reliable assessment. Specially, appropriate standardized procedures improve measurement by increasing consistency and test security. Consistent, standardized administration of the exam allows you to make direct comparison between examinees scores, despite that fact that the examinees may have taken their tests on different dates, at different sites, and with different proctors. Furthermore, administration procedures

that protect the security of the test help to maintain the meaning and integrity of the score scale for all examinees.

Necessary Conditions for test administration:

i) Check the questions

Are the questions correct or not? Number of questions papers are adequate or not? Are there any errors/~~defects~~ in question? etc. should be checked before distributing the question papers.

ii) Give clear directions

After checking the question paper, directions about rules and regulations of the examination /test should be given to the students. These include basically, exam/test time, time to go out, result affecting / disturbing the test etc.

iii) Create appropriate environment

Peaceful and comfortable environment should be created inside and outside the exam hall so that students can write/present their answers in their own way they know. As a result, they can have accurate evaluation.

iv) Control the cheating

To make the test administration effective, cheating should be prohibited.

v) Collected the answer papers

After the exam time is over, students' answer papers should be collected. While collecting the answer papers, it should be collected from beginning (front side) to end. After collecting the answer papers, they should be counted and arranged serially and should be submitted to the concerned authority.

4.2. Scoring of Subjective and Objective answer sheets

After completion of test administration, the answer sheets are to be scored. Two types of questions are asked in the test: Objective and subjective (essay items). Objective items are short answers are short and fixed and hence they are easy to score and the scoring becomes appropriate. On the other hand, essay items can be answered in own way in which answers may vary from one another and hence they cannot be scored (marked) objectively.

* For scoring the objective items a scoring key is used (an answer key). Answers should be ticked or circled clearly so that the teacher can score it efficiently in a short time.

In case of essay items, it should be indicated whether to score with 'point method' or with the 'rating method'. In point method each answer is compared with a set of ideal answers in the scoring key. Then a given number of points are assigned.

In the rating method, answers are rated on the basis of degree of quality and determine the credit assigned to each answer. Thus, scoring key helps to obtain consistent data about the pupil's performance.

OR next answer

4.2 Scoring of Subjective and Objective Answer Sheets

There are various types of Exams and various types of Questions are asked on such exams. So the methods used to score the answer sheet are different.

Steps of Scoring of Answer sheet:

- Prepare the answer key.
- Check the answer key.
- Scoring Blindly.
- Check Machine-scored Answer sheets.
- Check Scoring
- Recording

Scoring of Subjective test:

There is lack of reliability in scoring the subjective test. Generally, content, language, writing style, organization of thought, expression, style etc. are considered while checking such answers. The following methods can be used to score the subjective test:

- Evaluate all answers to one question before going on to next question.
- Point Scoring Method
- By using grouping method
- Avoid personal bias in scoring
- Scoring by more than one Examiner

4.3 Statistical analyses of test scores

Test scores have to be organized before they can be subjected to statistical analysis. This chapter centers on the following ways of organizing test scores: Ordering, ranking and the use of frequency distribution method.

Ordering refers to the numerical arrangement of numerical arrangement of numerical observations of measurements (Gordon & Gordon, 1994). There are two ways by which test scores can be numerically arranged: (1) ascending order and (2) descending order.

1. Ascending order - arranged from lowest to highest.
2. Descending order - arranged from highest to lowest.

4.3.1. Frequency distribution

In statistics, a frequency distribution is a list, table or graph that displays the frequency of various outcomes in a sample. Each entry in the table contains the frequency or count of the occurrences of values within a particular group of intervals.

There are three levels/types of frequency distribution:

1. Individual Series

Individual Series is a statistical series in which all the observations are listed out and all the observations have a frequency of 1. The frequency of occurrence of all the values in such a series is only one. Such series are displayed without the frequency column. For example, marks

Obtained by 10 students in Evaluation and Assessment can be represented in individual series as follows:

24, 35, 40, 49, 60, 74, 81, 83, 85, 88

2. Discrete Series

A discrete series is one in which the different values of a variable are shown in a discontinuous manner along with their respective frequencies and at least one of the values has a frequency of more than 1. Such a series can also be arranged either in ascending or descending order.

For example, marks obtained by 20 students in Computer graphics subject can be represented in discrete series as follows:

Marks obt. (x)	Frequency (f)
46	2
59	7
65	6
72	4
81	1
Total Student (N)	N = 20

3. Continuous Series

A continuous series is one in which the different values of the variables are stated in a continuous manner along with their respective frequencies. Data can take on any value within a specified range. Such series can be arranged in ascending or descending order. Further, such series can be stated either in the form of exclusive or in the form of inclusive class intervals along with

their respective class frequencies.

For example, marks obtained by 50 students in English subject can be represented in continuous series as follows:

Marks obtained (X)	frequency (f)
0-20	1
20-40	6
40-60	26
60-80	15
80-100	2
Total Students (N)	= 50

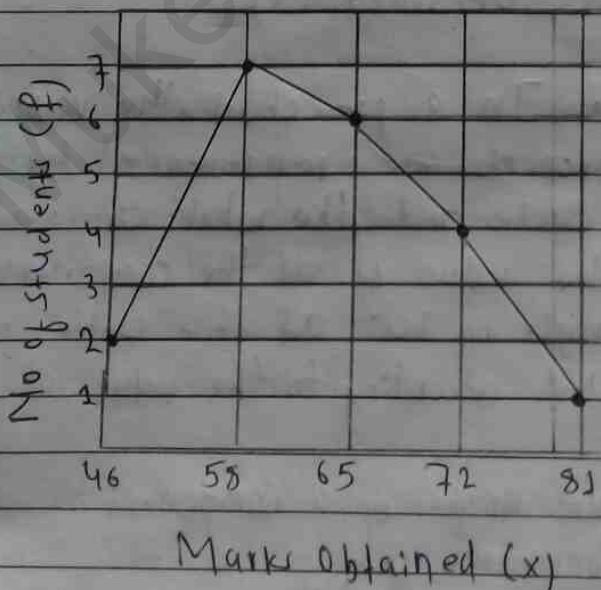
4.3.2. Graphical Representation : line-graph, bar-graph and pie-chart.

Graphical representation is another way of analysing numerical data. A graph is a set of chart through which statistical data are represented in the form of lines or curves drawn across the co-ordinated points plotted on its surface. Graphs are easy to understand and it is one of the most important learning strategies.

There are different types of graphical representation of test scores. Some of them are described below:

1. line-graph → A graph that uses line segments to connect data points and shows change in data over time.

Eg: Marks obtained by 20 students in Java subject can be represented in line graph as below:

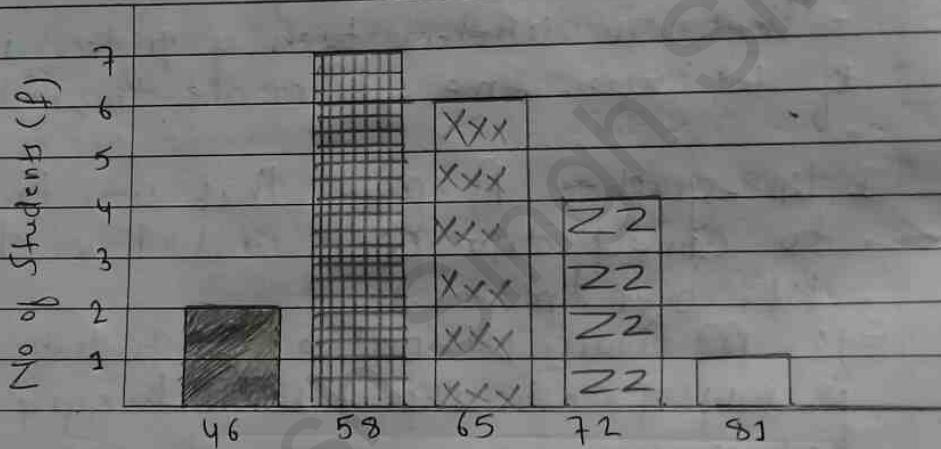


2 Bar-graph → A diagram in which the numerical values of variables are represented by the height or length of rectangles of equal width.

Example

Draw bar-graph of following test scores:

Mark Obt (x)	46	58	65	72	81
Frequency (f)	2	7	6	4	1



Marks Obtained (x)

3 Pie-Chart → In a pie-chart, the various categories or components are represented by the sectors of a circle and the whole circle represents the sum of the value of all the components. Clearly, the total angle of 360° at the centre of the circle is divided according to the value of the components.

The central angle of a component is

$$\frac{\text{Value of Component}}{\text{Total Value}} \times 360^\circ$$

Example

Draw a pie-chart from the following data:

Mark obtained (x)	Frequency (f)	Degree of f (central angle)
46	2	$\frac{2}{20} \times 360^\circ = 36^\circ$
58	7	$\frac{7}{20} \times 360^\circ = 126^\circ$
65	6	$\frac{6}{20} \times 360^\circ = 108^\circ$
72	4	$\frac{4}{20} \times 360^\circ = 72^\circ$
81	1	$\frac{1}{20} \times 360^\circ = 18^\circ$
Total	$N = 20$	360°

Showing the above data in pie chart:

4.3.3 Central Tendency : Mean, Median, Mode

The central tendency is stated as the statistical measure that represents the single value of the entire distribution or a dataset. It aims to provide an accurate description of the entire data in the distribution.

Measures of Central Tendency

Measures of Central Tendency are combined of two words i.e. 'measure' and 'central tendency'.

Measure means methods and central tendency means average value of any statistical series. Thus, we can say that central tendency means the methods of finding out the central value or average value of a statistical series of quantitative information.

Central Tendency

Mean	Median	Mode
------	--------	------

There are three measures of central tendency, such as:

1. The arithmetic Mean (Average) ✓
2. The Median ✓
3. The Mode ✓

1. Mean

The mean represents the average value of the dataset. It can be calculated as the sum of all the values in the dataset divided by the

number of values. It is popular method in measuring the central tendency of educational and psychological measures.

It can be calculated by the following ways:

- i) Individual Series → If the data in the distribution are in individual series (or un-grouped data) then mean can be calculated by following formula:

$$\bar{X} = \frac{\sum x}{N}$$

Where, \bar{X} = Mean

x = Individual Scores

$\sum x$ = Sum of all scores

N = No. of scores

Example

Calculate mean from the following test scores:

40, 48, 52, 55, 60

Soln:

$$\sum x = 40 + 48 + 52 + 55 + 60 = 255$$

$$N = 5$$

$$\text{Mean } (\bar{X}) = \frac{\sum x}{N} = \frac{255}{5} = 51$$

Hence, the mean score is 51.

- ii) Discrete Series → Discrete series means where frequencies ~~are given~~ of variables (scores) are given but the variable is without class intervals. When data is given along with their frequencies without any range.

It can be calculated by the following formula:

$$\bar{X} = \frac{\sum f x}{N}$$

Where, \bar{X} = mean, f = frequency of scores, x = scores
 $\sum f = N$ = sum of frequency of scores

Example

Calculate mean from the following test data.

Marks Obt. (x)	5	7	9	11	17
No. of Students (f)	4	5	3	7	6

Soln:

Showing the data in frequency table:

Marks Obt. (x)	No. of Stdu (f)	$f x$
5	4	20
7	5	35
9	3	27
11	7	77
17	6	102
	$N=25$	$\sum f x = 261$

By formula,

$$\bar{X} = \frac{\sum f x}{N} = \frac{261}{25} = 10.44$$

∴ Mean of the given test score is 10.44.

iii) Continuous Series → When data is given based on ranges (class intervals) along with their frequencies, it is continuous series.

If can be calculated by the following formula:

$$\bar{X} = \frac{\sum f m}{N}$$

Where, \bar{X} = mean, m = midpoint of class interval
 f = frequency, $N = \sum f$ = Number of Scores

Example: Calculate mean from following test score data.

Scores (x)	0-10	10-20	20-30	30-40	40-50	50-60
Frequency (f)	1	2	3	3	7	9

Soln:

Showing the data in frequency table:

Scores (x)	Frequency (f)	Mid-value (m)	fm
0-10	1	5	5
10-20	2	15	30
20-30	3	25	75
30-40	3	35	105
40-50	7	45	315
50-60	9	55	495
	N = 25		$\sum fm = 1025$

By formula,

$$\bar{x} = \frac{\sum fm}{N} = \frac{1025}{25} = 41.$$

∴ Mean of the given test data is 41.

Alternatively,

Continuous data mean can also be calculated from assumed mean by following formula:

$$\bar{x} = A + \frac{\sum fx'}{N} \times i$$

Where \bar{x} = mean, A = Assumed mean, f = frequency
 $N = \sum f = \text{No. of scores}$, i = class difference

$$x' = \frac{m - A}{i}$$

Q2 Median

Uses of Mean

- When the scores are distributed symmetrically.
- To measure central tendency with greater ability.
- To compute other statistics.

Advantages

- Easy to understand and calculate.
- It includes all scores of distribution.
- The result is reliable.

Q. Median

Median is a point in a distribution below which 50% (percent) cases and above which 50% (percent) cases lie.

OR, "The median is that value of the variable which divides the group into two equal parts, one part comprising all values ~~above~~ greater and other all values less than median".

Calculation of Median:

i) Individual Series → To find the value of Median, in this case, the terms are arranged in ascending or descending order first; and then middle term is taken called Median.

Formula for median in individual series is:

$$Md = \frac{N+1}{2}^{\text{th}} \text{ item}$$

Where N = Number of Scores.

Two cases arise in individual series :

(a) When number of terms is odd.

The terms are arranged in ascending or descending order and then are taken as Median.

Example: Find median of following test scores.

~~55, 43~~ 54, 49, 59, 60, 43, 56, 64, 61, 25

Soln.:

Arranging the scores in ascending order, we get:

25, 43, 49, 54, 56, 59, 60, 61, 64

$$N = \text{No. of scores} = 9$$

Median,

$$M_d = \frac{N+1}{2}^{\text{th}} \text{ item}$$

$$= \frac{9+1}{2}^{\text{th}} \text{ item}$$

$$= 5^{\text{th}} \text{ item}$$

In the above data, 5th item is 56.

$$\therefore \text{median} = 56.$$

(b) When number of terms is even

The terms are arranged in order and the mean of two middle terms is taken as median.

Example - Calculate median from following test scores.

54, 60, 64, 61, 66, 68, 76, 70, 79, 78

Soln.:

Arranging the terms in ascending order, we get:

50, 54, 61, 64, 66, 68, 70, 76, 78, 79

Here

$$N = 10$$

$$\text{Median, } M_d = \frac{N+1}{2}^{\text{th}} \text{ term}$$

- $\frac{1+1}{2}$ th term

- 5.5th term

Here

Here 5.5th term lies between 5th and 6th term,

so,

$$\text{Median}(\text{MD}) = \frac{5^{\text{th}} \text{ term} + 6^{\text{th}} \text{ term}}{2}$$

$$= \frac{66 + 68}{2} = \frac{134}{2} = 67$$

∴ Median of the test scores is 67.

Q ii) Discrete Series → Here also the data is arranged in ascending or descending order, And the $(N+1)/2$ th term is taken after finding cumulative frequencies. Value of variable corresponding to that term is the value of median.

Example - Compute median from following test data:

Mark obt. (x)	50	55	57	63	74
Frequency (f)	2	3	9	4	2

Soln:

Showing the data in Cf table:

Mark obt (x)	frequency (f)	Cf	
50	2	2	
55	3	5	1
57	9	14	
63	4	18	
74	2	20	
N = 20			

By formula,

$$\text{Median} = \left(\frac{N+1}{2} \right)^{\text{th}} \text{ term}$$

$$= \left(\frac{20+1}{2} \right)^{\text{th}} \text{ term}$$

$$= 10.5^{\text{th}} \text{ term}$$

Cf greater than 10.5 is 14 and its corresponding score is 57.

∴ Median of the data is 57.

iii) Continuous Series → In this case cumulative frequency is taken and then value from the class-interval in which $(N/2)^{\text{th}}$ term lies is taken.

The formula for continuous data is:

$$M_d = L + \frac{\frac{N}{2} - cf}{f} \times i$$

Where

L = Lower limit of class interval upon which median lies

N = Number of scores

cf = Cumulative frequency of the class preceding the median class

f = frequency of ^{the} median class.

i = size of class interval

Example

Calculate median from the following test data:

Mark obt (x)	0-10	10-20	20-30	30-40	40-50	
frequency (f)	5	8	9	3	5	

Soln:

The above data can be represented in cf-table as:

Mark Obt. (x)	Frequency (f)	Cf
0-10	5	5
10-20	8	13
20-30	9	22
30-40	3	25
40-50	5	30
	$N = 30$	

Here,

$$\frac{N}{2} = \frac{30}{2} = 15^{\text{th}} \text{ term}$$

$$Md \text{ class} = \left(\frac{N}{2} \right)^{\text{th}} \text{ class} = 15^{\text{th}} \text{ class}$$

Cf greater than 15 is 22 and its corresponding class is 20-30.

$$\therefore Md \text{ Class} = 20-30$$

$$L = 20$$

$$f = 10$$

$$cf = 13$$

$$f_i = 9$$

$$\therefore Md = L + \frac{\frac{N}{2} - cf}{f} \times i$$

$$= 20 + \frac{15 - 13}{9} \times 10$$

$$= 20 + \frac{2}{9} \times 10 = 20 + \frac{20}{9} = \frac{200}{9}$$

$$= 22.22$$

\therefore Median of the test data is 22.22

2. ~~Median~~ →

Uses of Median:

- To compute central tendency within short period of time.
- When the distribution is asymmetrical.
- To identify midpoint of distribution.
- When the data are given in ordinal scale.
- When the extreme score affect the mean.

Advantages

- Easy to compute and understand.
- All the observations are not required for its computation.
- Extreme scores does not affect the median.
- Can be determined from unequal class intervals.

3. Mode

Mode is the most frequently occurring score in ~~this~~ a distribution. In other words, mode has the highest frequency associated with it. It is denoted by the symbol M_o or mode.

Methods to Calculate Median:

- i) Individual Series → The terms are arranged in any order ascending or descending. If each term of the series is occurring once then there is no mode, otherwise the value that occurs maximum is known as mode.

Mode is often denoted by Z

Example: find mode from the following test scores.

40, 49, 53, 57, 53, 60, 57, 68, 55, 59

Soln.

Arranging the scores in ~~arranging~~ ascending order,

40, 49, 53, 53, 55, 55, 57, 57, 57, 60, 68
 In the above scores 57 has occurred 3 times,
 which is with highest frequency.
 $\therefore \text{Mode}(M_o) = 57$.

ii) Discrete Series → Here, the mode is known by inspection method only. Here the variable is the mode where the frequency is highest. But this method is applicable only if:

- (a) There is a gradual rise or fall in the sequence of frequencies.
- (b) The highest frequency and the next highest frequency are not too close.
- (c) Maximum frequency is not repeated.

Example: Find mode from the following test data.

Mark obtained (x)	16	22	29	35	42
Frequency (f)	2	2	20	14	2

soln:

In the above data, the highest frequency is 20 and the score corresponding to this frequency is 29.
 $\therefore \text{Mode}(M_o) \text{ is } 29$.

iii) Continuous Series → In the case of continuous series, we only go one step ahead of the method for discrete series. We get the value of mode by interpolation as ~~it is~~ is the case with median. The following formula is used to calculate Mode (M_o)

$$\text{Mode}(M_o) = L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times i$$

Where,

L = Lower limit of modal class

f_m = frequency of modal class

f_1 = frequency preceding the modal class

f_2 = frequency succeeding the modal class

i = size of class interval

Example: Compute Mode from the following data.

Mark obt. (x)	0-10	10-20	20-30	30-40	40-50	50-60
frequency (f)	4	12	13	16	8	3

Soln:

In the above data, ~~the~~ 30-40 class has highest frequency.

∴ Modal class = 30-40

$$L = 30$$

$$f_m = 16$$

$$f_1 = 13$$

$$f_2 = 8$$

$$i = 10$$

$$\therefore \text{Mode} (M_o) = L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times i$$

$$= 30 + \frac{16 - 13}{2 \times 16 - 13 - 8} \times 10$$

$$= 30 + \frac{3}{11} \times 10$$

$$= 30 + \frac{30}{11} = \frac{360}{11}$$

$$= 32.73$$

$$\therefore \text{Mode} (M_o) = 32.73$$

Uses of Mode:

- To measure central tendency quickly.
- When the most frequently occurring value of the distribution is wanted.
- When the data is given in nominal scale.
- To identify typical case.

Advantages:

- Mode gives the most representative value of a series.
- It is not affected by any extreme scores like mean.
- Easy to understand and calculate.
- Helps in analyzing qualitative data.

4.3.4 Measure of Dispersion: Standard deviation

In statistics, dispersion is the extent to which a distribution is stretched or squeezed. It is also called variability, scatter or spread.

A measure of statistical dispersion is a non-negative real number that is zero if all the data are the same and increases as the data become more diverse. Most measures of dispersion have the same units as the quantity being measured.

There are four commonly used measures to indicate the dispersion (variability) within a set of measures. They are:

- 1) Range ✓
- 2) Quartile deviation ✓
- 3) Mean deviation ✓
- 4) Standard deviation ✓

We discuss only Standard deviation in this unit.

Standard deviation:

Standard deviation is the square root of the variance.

By definition, "Standard Deviation (S.D.) is the square root of the mean of the squared deviations of the individual scores from the mean of the distribution."

To compute SD, we square all the deviations separately. Find their sum, divide the sum by total number of scores and then find the square root of the mean of the squared deviations.

Computing SD in 3 different series:

1. Individual Series →

Deviation can be taken from Actual mean and following formula is used:

for actual mean (Direct method)

$$S.D = \sqrt{\frac{\sum x^2}{N}} \quad \text{Where,}$$

$\sum x^2 = \text{Sum of variance}$

$N = \text{No. of scores}$

For Assumed Mean (Indirect method)

$$S.D (\sigma) = \sqrt{\frac{\sum x^2 - (\sum x)^2}{N}} \quad \text{Where,}$$

$\sum x^2 = \text{sum of square of each score.}$

$\sum x = \text{sum of each score}$

$N = \text{Number of scores}$

Example 1. Calculate S.D. of following scores by direct and indirect method.

2, 3, 4, 5, 6, 7

Method 1 : Using direct method.

Mark Obtained (X)	X^2
2	4
3	9
4	16
5	25
6	36
7	49
$\sum x = 27$	$\sum x^2 = 139$

By formula,

$$S.D. = \sqrt{\frac{\sum x^2 - (\sum x)^2}{N}}$$

$$= \sqrt{\frac{139}{6} - \left(\frac{27}{6}\right)^2}$$

$$= \sqrt{2.92} = 1.71$$

$$\therefore S.D. (\sigma) = 1.71$$

Method 2: By indirect method.

Mark Obt (x)	$x = (x - \bar{x})$	x^2
2	-2.5	6.25
3	-1.5	2.25
4	-0.5	0.25
5	0.5	0.25
6	1.5	2.25
7	2.5	6.25
$\sum x = 27$		$\sum x^2 = 17.50$

Here,

$$\bar{x} = \frac{\sum x}{N} = \frac{27}{6} = 4.5$$

By formula,

$$S.D. (\sigma) = \sqrt{\frac{\sum x^2}{N}} = \sqrt{\frac{17.50}{6}} = 1.71$$

$$\therefore S.D. = 1.71$$

2. Discrete Series \rightarrow Here also deviation can be taken from actual or assumed mean.

Formula for discrete series:

$$\sigma = \sqrt{\frac{\sum f d^2}{N} - \left(\frac{\sum f d}{N}\right)^2} \text{ where, } f = \text{frequency}$$

$d = x - A$ ($A = \underline{\text{assumed mean}}$)

$= \underline{\text{actual mean}}$

Example: Calculate S.D from following data:

Mark obt (X)	5	10	15	20	25	30
Frequency (f)	2	3	7	4	2	2
Soln:						

Showing the ~~first~~ data in frequency table:

Let Assumed mean, A = 20

Mark obt (X)	Frequency (f)	$d = X - A$	f_d	d^2	f_d^2
5	2	-15	-30	225	450
10	3	-10	-30	100	300
15	7	-5	-35	25	175
20	4	0	0	0	0
25	2	5	10	25	50
30	2	10	20	100	200
$N = 20$			$\sum f_d = -65$	$\sum f_d^2 = 1175$	

By formula,

$$\begin{aligned}\sigma &= \sqrt{\frac{\sum f_d^2}{N} - \left(\frac{\sum f_d}{N}\right)^2} \\ &= \sqrt{\frac{1175}{20} - \left(\frac{-65}{20}\right)^2} \\ &= \sqrt{58.75 - 10.56} = \sqrt{48.19} = 6.94\end{aligned}$$

$$\therefore S.D(r) = 6.94$$

3. Continuous Series → Here we take deviation from actual or assumed mean as desired from the mid point of class-intervals.

Formula for continuous series is:

$$\sigma = \sqrt{\left[\frac{\sum f_d'^2}{N} - \left(\frac{\sum f_d'}{N}\right)^2\right] \times i}$$

Where, $d' = \frac{m-A}{i}$

m = mid-point

A = Assumed mean

i = Size of class interval

Example: Calculate S.D. from following data.

Mark obt (X)	0-10	10-20	20-30	30-40	40-50
frequency (f)	7	2	4	9	8

Soln: Let Assumed mean = 25

Showing the data in table:

Mark X	Frequency f	mid-value m	$d' = \frac{m-A}{i}$	d'^2	fd'	fd'^2
0-10	7	5	-2	4	-14	28
10-20	2	15	-1	1	-2	2
20-30	4	25	0	0	0	0
30-40	9	35	1	1	9	9
40-50	8	45	2	4	16	32
$\Sigma f = 30$					$\Sigma fd' = 9$	$\Sigma fd'^2 = 71$

By formula,

$$\sigma = \sqrt{\left[\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N} \right)^2 \right] \times i}$$

$$= \sqrt{\left[\frac{71}{30} - \left(\frac{9}{30} \right)^2 \right] \times 10}$$

$$= \sqrt{[2.37 - 0.09]} \times 10$$

$$= \sqrt{2.28} \times 10$$

$$= 1.51 \times 10$$

$$= 15.1$$

$$\therefore S.D(\sigma) = 15.1.$$

User of Standard deviation

- To measure dispersion accurately in reliable way.
- To compute different statistics.
- To interpret normal distribution.
- To identify the nature of the group.

Advantages

- Its value is always definite.
- It is based on all values of deviation, distribution.
- Reliable and accurate.

Range)

On RP 11/06

Unit-5

Current Student Assessment System (in Nepal)

Ajanta

Page No. _____
Date _____

5.1. Existing student assessment system at school level

At the present time, it is found that from school to university level, various types of evaluation methods are used. Initially, placement evaluation is used to detect students' entry behaviour, while formative evaluation is used to improve them with the aim of improving the weakness of general nature.

Similarly, some schools have given some degree of priority to diagnostic evaluation, with the aim of diagnosing and improving the complex nature problems encountered during teaching. In addition, summative evaluation is used to determine whether the students have or have not achieved the expected achievements at the end of the academic session and then determine whether to make them pass or fail.

Among these various ~~forms~~ of evaluation, currently the policy and practice of constructive and summative (or decisive) evaluation methods are preferred in schools in Nepal.

The current status of the school level evaluation system in Nepal is discussed below:

Student Evaluation and Certification:

- For the students' evaluation and certification, school level examinations as well as external public examinations will be conducted.
- The school will be responsible for the classes 1 to 7 and class 9 and 11 examination under the school based examination. For this, a continuous assessment system (CAS) will be adopted up to

grade 7.

- For classes 8, 10 and 12 public examinations will be conducted at local, province and national level respectively. Class 8 final examination will be conducted by local level's education unit. Class 10 final examination will be conducted at Province level. Class 12 final examination will be conducted at national level by the National Education Board (NEB).
- To make the evaluation system objective, reliable and long lasting and to maintain uniformity, internal and external evaluation system should be implemented in all the subjects of local level, province level and national level public examinations. There should be 95% internal marks and 75% external marks evaluation in each subject of public examinations. Internal evaluation will be done by concerned school's subject teacher and the working procedure in this regard will be as per the approval of the Nation Curriculum Development and Evaluation Council (NCDEC).
- While conducting both School level as well as External public examinations, subjects having 3 or more credit hours will have 100 full marks and 3 hours for exam duration and the subjects having 2 or less credit hours will have 50 full marks and 2 hours 30 minute as exam duration.

- While preparing student grade sheets for verification of students' results, internal and external assessments will be carried out in accordance with the procedure adopted by the National Curriculum Development and Evaluation Council.
- Both theoretical and practical aspects will be included in the evaluation of the school teacher. The weight of the theoretical and experimental evaluation will be determined based on the nature of the subject. Generally, the weight of experimental evaluation will be less than 20% and not more than 60%. The weight of experimental work in the external public examination will be included in the internal evaluation.
- Students will be provided the opportunity to upgrade their grade in the subject they wish to.
- There will be coordination between CAs system and CGPA system.
- While conducting school internal examination and external public examinations, the examination committee should conduct examination according to the model and specificational grid provided by NCDC.

A. Student Assessment System at Basic Level:

In basic level, students' learning will be assured by formative evaluation more than the summative evaluation. The main aim of formative evaluation is to improve the learning level of students.

Basic level 1-3 Basic level 1-3

- From grade 1-3, there is a provision policy provision to keep a student's achievement archive file, which includes the student's classwork, homework, project work, students' learning achievement, practical changes in them and regularity.
- Grade 1-3 of basic level will be fully (100%) based on Continuous Assessment System.
- Students studying in grades 1-3 are categorized into classes A to C based on their classroom progress, activities.

Basic level 4-5

- Students studying in grades 4-5 are evaluated based on summative and formative evaluation.
- Students' learning achievement is evaluated by 50% CAS and 50% external summative

evaluation method.

- Along with them, students studying in grades 4 & 5 will have 40% from marks.

- How Classes 4 and 5 students are promoted using corrective and summative (deciive) assessment tools.

Basic Level 6-8

- Grade 6 and 7 students are evaluated by legislative leadership (school itself) using formative and ~~scor~~ dective (summative) assessment tools.

- Student achievement in grade 6 and 7 is assessed through Periodic assessment of 60% weight and continuous assessment of 40% weight. Continuous CAS includes practical work, classroom participation, improvement in behaviour, constructive works etc. By completing it a. a Continuous assessment by the teacher, there is a provision to classify them into A, B and C categories based on the progress made by them.

- Generally, in ~~class~~^{grades} 6 and 7 three examinations are held. First term consists of 20% weight, 2nd term consists 30% weight and final term ~~consists~~ consists 50% weight.

- Whereas at the end of basic level (in grade 8) there is a provision of conducting local level examination.

by the local governing body.

- Students of grade 8 are fully (100%) evaluated by ~~100~~ external local level examination.
- Pass marks for the examination is 40%.
Students ~~obtain~~ getting 80% or above are placed categorized in A category, 60 to 79% are placed in B category and 40 to 59% are placed in C category.
- Students are given 3 hours for 100 full marks in the written examination and 1 hour 5 minute for 50 full marks.
- While evaluating the students, appropriate evaluation process must be followed for the ~~B students~~ students with disability and students having special needs.

Ans
Basic

B. Secondary level 9-12

Continuous / Constructive and decisive (summative) evaluation procedures have been adopted to assess student achievement at the secondary level.

To conductive formative evaluation, as an integral part of teaching learning activities, assessment tools such as homework, classwork, project work, social work, extra activities, unit test, monthly test etc. are used. For students with special needs teacher should select appropriate process for evaluating such students.

To conduct summative evaluation, first terminal, second terminal and final examination ~~are~~ assessment tools are used to perform external evaluation.

Grade 9

Students of grade 9 are fully (100%) evaluated by the school itself. In grade 9, three examinations are held: 1st term consisting 10% weight, 2nd term consisting 30% weight and final exam consisting 60% weight.

Grade 10

While in grade 10 SEE examination is conducted by ECB under the NEB, for evaluating the ~~student~~ achievement level of students. The SEE examination is conducted at the Province level.

The method of evaluation here is fully ~~external~~ based.

Subjects having 100 full marks is given 3 hours, & subjects having 75 full marks is given 2 hours 15 minutes

and the subjects having 50 or full marks,
is given 1 hour and 30 minutes.

Grade 11 & 12

- The method of evaluation for grade 11 and 12 is fully external based till now. ~~But~~
(But in the coming years, class 11 evaluation will be done by the concerned school and class 12 examination will be conducted externally and publicly by NEB.)
- At the end of sessions, NEB ~~exam~~ conducts external public examination for grade 11 and 12 at the National level.
(But in future only class 12 examination will be conducted by NEB).
- Subjects with full marks 100 and subjects with ~~75~~ full marks 75 both are given 3 hours time for the examination.
- Students of 11 and 12 will have to obtain 35% marks to pass the examination.

Grading System for Class 10 (SEL), 11 & 12 (MCB)
Examinations:

100 (75 + 25)

↓ ↓

S.N.	Obtained Marks (in Percentage)	Grade	Remarks	Grade Point
1.	90 and Above	A +	Outstanding	4.0
2.	Above 80 and below 90	A *	Excellent	3.6
3.	Above 70 and below 80	B +	Very Good	3.2
4.	Above 60 and below 70	B	Good	2.8
5.	Above 50 and below 60	C +	Above Average	2.4
6.	Above 40 and below 50	C	Average	2.0
7.	Above 20 and below 40	D	Below Average	1.6
8.	Above 1 and below 20	E	Insufficient	0.8
9.	0	N	Not Graded	0

✓

✓ B v. Good

5.2 Assessing Students with Special Needs

Evaluation is an essential beginning step in the special process for a child with a disability. Before a child can receive special education and related services for the first time, a full and individual initial evaluation of the child must be conducted to see if the child has a disability and is eligible for special education.

The government will adopt following measures to formulate inclusive assessment policy for the students with special needs:

- 1) For students with visual disability, there will be separate provision for students with low vision and for those who are blind.
 - a) For low vision students, provision of large size prints of the question, examination hall with extra light and reasonably extra time will be provided.
 - b) Since the blind students can read and write in Braille Script, there will be provision of question papers in Braille script and answer books to write in Braille Script. As an alternative to Braille Script, there could be arrangement of recording question papers and answers in audio tape, and examiner will be given slightly more time so that they can correct answer in the recorder. An alternative provision will be made in practical

Subject and substitution made in items for pictorial descriptions and drawing in subjects such as geography, geography and physics.

- 2) A flexible policy will be adopted for the students with poor hearing.
- 3) There will be comfortable seating arrangements for the students in the examination hall for physically weak students. Those who have disability in their arms and hands and thereby have problems in writing will be provided with extra time and other special provisions.
- 4) Apart from the examination, the other alternative arrangements will be made for the students with special needs.

5.3 Continuous Assessment System (CAS) : Concept, Process and Practice

Concept

Continuous Assessment System is a form of educational examination that evaluates a student's progress throughout a prescribed course. It is often used as an alternative to final examination system.

The trend was established that regular evaluation of the students should improve their weaknesses over time and the concept of Continuous Assessment System (CAS) was incorporated in Nepali schools.

The concept of CAS in Nepal started with the suggestion of the National Education Scheme 2049. It suggested that the system should be adopted to evaluate students at the primary level. Higher National Education Scheme 2055 emphasized its implementation. As a result, the Continuous Student Assessment Guide 2055 was released & was implemented in 5 districts (Jlam, Chitwan, Syangja, Surkhet and Kanchanpur) across the country as a test since the academic session of 2057. The School Sector Reform Plan (SSRP) 2066-2072 changed the structure of the school level and declared the class 1-8 as the basic level and 9-12 the secondary level. Later, the national curriculum format 2063 fixed the policy of expanding the ~~continuous~~ concept of continuous assessment from class 1-3 to 1-7.

Thus, the continuous evaluation assessment system has a long history, but its history is short in Nepal.

How much did the students learn on the basis of the frequency of the unit and quarterly examinations in this system, and the class work, homework, project work, behaviour, observation record? It is believed that non-learned behaviour will be facilitated by therapeutic teaching and improving learning.

Process

According to the current structure of Nepal's education system, (class-grade 1-8) is considered at the basic level. It is further divided into 3 stages:

- i) First stage: 1-3, ii) Second stage 4-5 and iii) Third stage: 6-8. (Grade 1-3 students are fully evaluated by CAs. Grade 4-5 students are evaluated by 50% CAs and 50% External examination and Grade 6-8 students are evaluated by 40% CAs + 60% External Examinations.

Following procedure is followed for CAs:

- i) To determine classroom and subjectwise learning achievement indicators.
- ii) Separate the ~~wor~~ workfile for each student by filling up the necessary information.
- iii) Record the ~~per~~ behaviour performance of the students.
- iv) In addition to the student's written performance

examinations during assessment, the following techniques will be the basis:

- Unit test
- Class work
- homework
- Project work
- Classroom participation
- Attendance
- ~~Constructive works~~
- Creative works, etc

v) To provide necessary support to achieve the achievements not achieved in regular classes. After the level of assessment has been increased, record the assessment form accordingly.

vi) To promote generous promotion to the students who have reached the age level but do not have the required minimum attendance or who do not ~~have~~ achieve the minimum learning achievement, which is achieved in the upper classes in consultation between teachers, parents and principals.

Recording

BY

Rewarding Process of Continuous Assessment System:

The main purpose of the continuous assessment system is to make the teaching process agile and to improve the timely detection of the weakness of the students observed during the teaching and learning activities.

Generally, the continuous assessment system should record the progress of the student as follows:

- i) Based on the criteria set out in Schedule -1, the students should take 1 to 3 tick marks (✓) on a textual basis.
- ii) The best learned behaviour should be given 3 tick marks (vvv), two tickmarks ^{up} for the learned in average (v) and one tickmark (✓) for the learned in general.
- iii) On the basis of the text, students should
- iv) After giving a tickmarks on a textual basis, students should be categorized into categories A, B and C every 3 months on the basis of the following regimen:

S.N.	Percentage	Category
1.	70% to 100%	A ✓
2.	40% to 70%	B ✓
3.	Below 40%	C ✓

- v) Percentage of the ~~text~~ textual markers should be calculated using the following formula:

Tickmarks % = Obtained no. of tickmarks $\times 100\%$
Total taught chapters $\times 3$

- v) In evaluating the lessons learned, the learning specified by the curriculum is completed or not should be viewed.
 - v) In each Quarter (3 months), the percentage determined from the textual assessment should be kept along with the written and oral examination score in the Student Progress Statement as per Schedule 2.
 - vi) As per the schedule 3 and 2, the form should be prepared by the school ~~is~~ itself.

Schedule - I

Continuous Student Assessment Form

School Name:

Academic Year:

class

- 54 -

Subject Teacher:

Signature

Verified by:

Schedule - 2Student Progress Description

Academic Year:

Name & School:

Class:

Roll. No.:

Subjects	Continuous Student Assessment Category (A, B, C)			1 st , 2 nd and Final term marks					
	1 st term	2 nd term	Final (Year) Exam ^(W)	First term FM	Obt.	Second term FM	Obt.	Yearly Exam FM	Obt.
			10 marks	30 marks	60 marks	60 marks			
Nepali									
English									
Maths									
Science, Health & Physical Education									
Social and Creative Skills									
Local Subject									
Total									

Subject Teacher's name:

Verified by:

Signature:

Characteristics of Continuous Assessment System,
Properties of

Characteristics of Continuous Assessment
System are mentioned below:

- i) It can evaluate all types of student's abilities.
- ii) Continuous student assessment is informal.
- iii) It emphasizes the corrective aspect rather than the validation of the learning.
- iv) Continuous student assessment is conducted continuously along with teaching learning without conducting at any given time.
- v) It uses examinations as well as other assessment tools.
- vi) There is also the provision of feedback to improve the learning.
- vii) It encourages weaker students to raise their achievement levels and motivates talented students to learn more.
- viii) It provides students to learn freely with the opportunity to learn freely without fear of examination.
- ix) It strengthens the triangular relationships

~~between~~ of teachers, students and parents.

x) ~~This~~ It makes the teacher active and responsive.

Applications of CAS in basic grade level (Grade 1 to 9)

In basic level, the CAS is used for the following purposes:

- i) To make a realistic assessment of all types of students' abilities,
- ii) To make student learning effective
- iii) To implement a generous promotion (class promotion) policy,
- iv) To provide the information required for the decision making,
- v) To save extra time and resources ~~for~~ needed for an evaluation,
- vi) To reduce class dropout (bulk) and repetition rates,
- vii) To inform parents or stakeholders about student achievement and progress,
- viii) To make the teacher responsible and active in the students' learning,

ix) To create the necessary learning environment in schools and at home by expanding teacher-student, teacher-parent relationships.

5.0

Challenges and issues of existing student assessment system at school level.

Problems of Implementing CAS in Schools, portwise :

- i) Finance
- ii) Adaptation of change
- iii) Training of teachers
- iv) Mode of Operation
- v) Record keeping
- vi) Favoritism
- vii) Overpopulation
- viii) Over concentration in written test
- ix) Uniformity