

F-2.3
ASSESSMENT FOR LEARNING

Unit-I: Overview of Assessment and Evaluation

- a) Basic Concepts: Assessment for learning, assessment of learning, test, examination, measurement, evaluation.
- b) Objectives and Purpose of Assessment – Behaviorist and Constructivist Paradigm
- c) Types of evaluation- Teacher made and standardized tests, Norm referenced and criterion referenced testing;

Unit-II: Assessment Tools

- a) Test: Types (Essay, Objective, Objective based); characteristics of good tool (validity, reliability and usability).
- b) Construction of a Test: Planning (Blue Print), Preparation, Try Out and Evaluation
- c) Techniques of Assessment: Project work, Assignments, Practical work, Performance based activities.

Unit-III: Analysis and Implementation of Assessment

- a) Scoring procedure - manual and electronic, development of Rubrics
- b) Analysis and Interpretation: Calculation of percentages, Frequency distribution, Percentile Rank, Pie Chart, Bar Graph, Histogram, Frequency Polygon,
- c) Normal probability Curve- Meaning, characteristics and uses.

Unit-IV: Trends and Issues in Assessment

- a) Existing Practices: Continuous and Comprehensive Evaluation (CCE), Grading, Choice Based Credit System, Feedback in Improving learning and learners' development.
- b) Issues and Problems: Non-Detention Policy, the menace of coaching.
- c) Emerging Practices in Assessment: Standard Based Assessment, Online, Computer Based and Open Book Examinations.

Unit-I: Overview of Assessment and Evaluation

a) Basic Concepts: Assessment for learning, assessment of learning, test, examination, measurement, evaluation.

b) Objectives and Purpose of Assessment – Behaviorist and Constructivist Paradigm

INTRODUCTION

The term Assessment has been widely used by educators to evaluate, measure, and document the academic readiness, learning progress, and skill acquisition of students throughout their learning in life. Different terminologies are there for assessment and evaluation such as Measurement, Tests, Examination, Appraisal and Evaluation. There are certain Learning theories which are having conceptual frameworks describing how information is absorbed, processed and retained during learning. Behaviourism is a philosophy of learning that only focuses on objectively observable behaviors and discounts mental activities. Piaget proposed that a child's cognitive structure increases in sophistication with development, moving from a few innate reflexes such as crying and sucking to highly complex mental activities. Constructivist learning theory stated that the process of adjusting our mental models to accommodate new experiences. Assessments are classified based on the different purposes, scopes, attribute measured, nature of information gathered, nature of interpretation and context.

Meaning of Assessment

In education, the term assessment refers to the wide variety of methods that educators use to evaluate, measure, and document the academic readiness, learning progress, and skill acquisition of students from preschool through college and adulthood. It is the process of systematically gathering information as part of an evaluation. Assessment is carried out to see what children and young people know, understand and are able to do. Assessment is very important for tracking progress, planning next steps, reporting and involving parents, children and young people in learning.

PURPOSES OF ASSESSMENT

1. Teaching and Learning

The primary purpose of assessment is to improve students' learning and teachers' teaching as both respond to the information it provides. Assessment for learning is an ongoing process that arises out of the interaction between teaching and learning.

What makes assessment for learning effective is how well the information is used.

2 System improvement

Assessment can do more than simply diagnose and identify students' learning needs; it can be used to assist improvements across the education system in a cycle of continuous improvement:

- Students and teachers can use the information gained from assessment to determine their next teaching and learning steps.
- Parents and families can be kept informed of next plans for teaching and learning and the progress being made, so they can play an active role in their children's learning.
- School leaders can use the information for school-wide planning, to support their teachers

and determine professional development needs.

- Communities and Boards of Trustees can use assessment information to assist their governance role and their decisions about staffing and resourcing.
- The Education Review Office can use assessment information to inform their advice for school improvement.

The Ministry of Education can use assessment information to undertake policy review and development at a national level, so that government funding and policy intervention is targeted appropriately to support improved student outcomes.

PRINCIPLES OF ASSESSMENT

a. Reliability

A test can be reliable but not valid, whereas a test cannot be valid yet unreliable. Reliability, in simple terms, describes the repeatability and consistency of a test. Validity defines the strength of the final results and whether they can be regarded as accurately describing the real world.

b. Validity

The word "valid" is derived from the Latin *validus*, meaning strong. The validity of a measurement tool (for example, a test in education) is considered to be the degree to which the tool measures what it claims to measure; in this case, the validity is an equivalent to accuracy.

c. Relevance and transferability

In education, the term relevance typically refers to learning experiences that are either directly applicable to the personal aspirations, interests or cultural experiences of students (*personal relevance*) or that are connected in some way to real-world issues, problems and contexts (*life relevance*).

Relevance is the concept of one topic being connected to another topic in a way that makes it useful to consider the first topic when considering the second. The concept of relevance is studied in many different fields, including cognitive sciences, logic, and library and information science. Most fundamentally, however, it is studied in epistemology (the theory of knowledge). Different theories of knowledge have different implications for what is considered relevant and these fundamental views have implications for all other fields as well.

Transferability in research is the degree to which the results of a research can apply or transfer beyond the bounds of the project. Transferability implies that results of the research study can be applicable to similar situations or individuals. The knowledge which was obtained

in situation will be relevant in another and investigators who carry out research in another context will be able to utilize certain concepts which were initially developed. It is comparable to generalisability.

Transferability in research is utilized by the readers of study. Transferability can apply in varying degrees to many types of research. Transferability doesn't involve broad claims, but invites readers of research to make associations between elements of research and their own experience. For example, lecturers at a school may selectively apply to their own class's results from a research indicating that heuristic writing exercises aid students at the university level. It is important that adequate thick description of the phenomenon under study is given to allow audience to have a proper understanding of it, thus enabling them to compare the instances of the phenomenon explained in the research document with those that they have seen emerge in their situations.

LEARNING THEORY

Learning theories are conceptual frameworks describing how information is absorbed, processed and retained during learning. Cognitive, emotional, and environmental influences, as well as prior experience, all play a part in how understanding, or a world view, is acquired or changed and knowledge and skills retained.

Behaviorists look at learning as an aspect of conditioning and will advocate a system of rewards and targets in education. Educators who embrace cognitive theory believe that the definition of learning as a change in behavior is too narrow and prefer to study the learner rather than their environment and in particular the complexities of human memory. Those who advocate constructivism believe that a learner's ability to learn relies to a large extent on what he already knows and understands, and the acquisition of knowledge should be an individually tailored process of construction. Transformative learning theory focuses upon the often necessary change that is required in a learner's preconceptions and world view.

1 Behaviorism

Behaviorism is a philosophy of learning that only focuses on objectively observable behaviors and discounts mental activities. Behavior theorists define learning as nothing more than the acquisition of new behavior. Experiments by behaviorists identify conditioning as a universal learning process. There are two different types of conditioning, each yielding a different behavioral pattern:

- Classic conditioning occurs when a natural reflex responds to a stimulus.

The most popular example is Pavlov's observation that dogs salivate when they eat or even see food. Essentially, animals and people are biologically "wired" so that a certain stimulus will produce a specific response.

- Behavioral or operant conditioning occurs when a response to a stimulus is reinforced. Basically, operant conditioning is a simple feedback system: If a reward or reinforcement follows the response to a stimulus, then the response becomes more probable in the future. For example, leading behaviorist B.F. Skinner used reinforcement techniques to teach pigeons to dance and bowl a ball in a mini-alley.

Constructivism

Constructivism is a philosophy of learning founded on the premise that, by reflecting on our experiences we construct our own understanding of the world we live in. Each of us generates our own "rules" and "mental models," which we use to make sense of our experiences. Learning, therefore, is simply the process of adjusting our mental models to accommodate new experiences.

The guiding principles of Constructivism:

- Learning is a search for meaning. Therefore, learning must start with the issues around which students are actively trying to construct meaning.
- Meaning requires understanding wholes as well as parts and parts must be understood in the context of wholes. Therefore, the learning process focuses on primary concepts, not isolated facts.
- In order to teach well, we must understand the mental models that students use to perceive the world and the assumptions they make to support those models.

The purpose of learning is for an individual to construct his or her own meaning, not just

memorize the "right" answers and repeat someone else's meaning. Since education is

inherently interdisciplinary, the only valuable way to measure learning is to make assessment part of the learning process, ensuring it provides students with information on the quality of their learning.

How Constructivism impacts learning:

- Curriculum - Constructivism calls for the elimination of a standardized curriculum. Instead, it promotes using curricula customized to the students' prior knowledge. Also, it emphasizes hands-on problem solving.
- Instruction - Under the theory of constructivism, educators focus on making connections between facts and fostering new understanding in students. Instructors tailor their teaching strategies to student responses and encourage students to analyze, interpret and predict information. Teachers also rely heavily on open-ended questions and promote extensive dialogue among students.
- Assessment - Constructivism calls for the elimination of grades and standardized testing. Instead, assessment becomes part of the learning process so that students play a larger role in judging their own progress.

CLASSIFICATION OF ASSESSMENT

There are three types of assessment: diagnostic, formative and summative. Although there are three generally referred to simply as assessment, there are distinct differences between the three.

1.7.1 Prognostic Assessment

A prognostic assessment expands the findings of an assessment with analysis of abilities and potentials with a further dimension: the future development of the concerned person, as well as the necessary conditions, timeframe and limits.

Finding the right person for an executive position needs a reliable comprehension of the personality as well as the possibilities and limits concerning the personal development. Even an experienced and keen observer of human nature may get deluded, even recognized and proven test procedures may be incomplete or leading to wrong results – and misjudgments can become expensive in substantial and immaterial ways.

Six Goals of the Prognostic Personality and Abilities Assessment

Analysis of existing abilities and interests, including the not (yet) known ones and the development to be expected.

- If needed, a comparison with job description and profile of requirements.
- Basic conditions and needs for the development: how it can be enhanced and ensured.
- Period: how long the development will take until the defined goals can be reached.
- Limits of developmental possibilities, either referring to the defined goals (selection assessment), or generally, with a realistic time frame of 3 to 5 years.
- Quality assurance and sustainability: how the results can be monitored and ensured in the long term.

The prognostic assessment is suitable for all management levels including executive board and administrative council, but likewise for young people with the aim of a comprehensive potential analysis. Typically, the prognostic assessment is accomplished as an individual oneday-assessment. The objectives are defined individually.

1.7.2 Formative Assessment

Formative assessment provides feedback and information during the instructional process, while learning is taking place, and while learning is occurring. Formative assessment measures student progress but it can also assess your own progress as an instructor. A primary focus of formative assessment is to identify areas that may need improvement. These assessments

typically are not graded and act as a gauge to students' learning progress and to determine teaching effectiveness (implementing appropriate methods and activities).

❖ **Types of Formative Assessment:**

- Observations during in-class activities
- Homework exercises as review for exams and class discussions
- Reflections journals that are reviewed periodically during the semester
- Question and answer sessions, both formal—planned and informal—spontaneous
- Conferences between the instructor and student at various points in the semester
- In-class activities where students informally present their results
- Student feedback collected by periodically

❖ **Diagnostic Assessment:**

Diagnostic assessment can help you identify your students' current knowledge of a subject, their skill sets and capabilities, and to clarify misconceptions before teaching takes place. Knowing students' strengths and weaknesses can help you better plan what to teach and how to teach it.

➤ **Types of Diagnostic Assessments:**

- Pre-tests (on content and abilities)
- Self-assessments (identifying skills and competencies)
- Discussion board responses (on content-specific prompts)
- Interviews (brief, private, 10-minute interview of each student)

1.7.3 Summative Assessment

Summative assessment takes place after the learning has been completed and provides information and feedback that sums up the teaching and learning process. Typically, no more

formal learning is taking place at this stage, other than incidental learning which might take place through the completion of projects and assignments.

Types of Summative Assessment

- Examinations (major, high-stakes exams)
- Final examination (a truly summative assessment)
- Term papers (drafts submitted throughout the semester would be a formative assessment)
- Projects (project phases submitted at various completion points could be formatively assessed)
- Portfolios (could also be assessed during its development as a formative assessment)
- Performances
- Student evaluation of the course (teaching effectiveness)
- Instructor self-evaluation

c) Types of evaluation- Teacher made and standardized tests, Norm referenced and criterion referenced testing;

TESTS

1 Types of Tests

The test is to measure the ability, knowledge or performance developed during the course of learning. Here two types of test are given to assess the knowledge of students they are Teacher Made Tests and Standardized Tests.

a. Teacher Made Tests

These tests are designed by the teachers for the purpose of conducting classroom tests. These teacher made tests can be in the form of oral tests and written tests. These tests have a limited area of application and are prepared almost by all teachers according to their requirements.

A teacher is more concerned with the teacher – made tests as she is directly involved in its construction. Moreover, the teacher made tests have an advantage over standardized tests because they can be constructed to measure outcomes directly related to classroom specific objectives and particular class situations. These tests are within the means of every teacher and most economical. Teacher made oral tests are designed to measure the performance of students skills like listening and speaking in language learning. Written tests are designed to test the abilities of student's knowledge comprehension and written expression.

b. Standardized Tests

A Standardized test is one which norms have been established. The test has been given to a large number of students. A norm is an average score which measures achievement. So, every standardized test has norms. It is intended for general use and covers a wider scope of material than is covered in an ordinary teacher made test. A standardized test is one in which the procedure, apparatus and scoring have fixed so that precisely the same test can be given at different time and places. A standardized test is one which has been given to so many people that the test makers have been able to determine fairly accurately how well a typical person of a particular age or grade in school will succeed in it.

i. Role of Standardized Test

- ❖ Information becomes easier to convince the guardians of students
- ❖ Information in much less time than provided by other devices.
- ❖ Information for all guidance workers.
- ❖ Aspects of the behaviour which otherwise could not be obtained.
- ❖ Objectives and impartial informative about an individual.

ii. Steps Involved in Standardized Test

A standardized test is tried out and administered on a number of subjects for the expressed purpose of refining the items by subjecting the performances of the standard decision to pertinent statistical analysis. The steps for the standardized test is constructed by test specialists or experts they are

- ❖ Proper planning
- ❖ Adequate preparations
- ❖ Try-out of the test
- ❖ Preparation of proper, norms
- ❖ Preparation of a manual containing instruction of administering a tool or test.
- ❖ Item analysis

iii. Teacher made Test vs. Standardized Tests

The standardized test is based on the general content and objectives common to many schools all over the country whereas the teacher made test can be adapted to content and objectives specific to his own situation. The standardized test deals with large segments of knowledge or skill whereas the teacher made test can be prepared in relation to any specific limited topic. The standardized test is developed with the help of professional writers, reviewers and editors of tests items whereas the teacher made test usually relies upon the skill of one or two teachers. The standardized test provides norms for various groups that are broadly representative of performance throughout the country whereas the teacher made test lack this external point of

reference.

iv. Characteristics of a Standardized Test

- Standardized tests are based on the content and objectives of teaching common to many schools.
- Not just one, but a team of experts are involved in the writing of test items.
- Items analysis is done on the basis of a pilot study, unlike in the case of a class room test.
- Norms are calculated for the purpose of comparison between grades, schools, age levels and sexes.
- They cover large segments of knowledge and skills.

74

- Test manuals are prepared.
 - Fairly a large same, not just one class is involved in the standardization of a test.
- Teacher need to test student performance. Test results are critical, not only because the affect careers, but because of the influence they exercise on motivation to learn. Teacher must be aware of different testing techniques, because they give useful information to both the teacher and the students. Testing techniques are often similar the teaching techniques, but with a different purpose.

Importance of Testing

1. Tests tell the teacher what the students can and cannot do and here are how successful the teaching has been.
2. They also tell the teacher what areas need to be taught in the future.
3. Test tell the students how well they are progressing, and where they need to focus their attention as learners.
4. Regular tests also encourage students to take their learning seriously, and give them a series of definite goals to aim towards.

Similarities

- They are both means of performance assessment.
- They both use the same type of test items.
- They both require validity, reliability, objectivity and efficiency.

Differences

- The classroom test may have more content validity than standardized tests.
- The quality of the test items on the standardized test is usually superior because they are prepared by test specialists and are revised on the basis of actual try out and item analysis.
- The procedures used in administering and scoring standardized tests are carefully described and they are standardized for each administration of the test.

The student's score on a standardized test is interpreted on basis of norms already developed on a large sample. But the student's score on a classroom test is interpreted in relation to the average score of the classroom students.

- Standardized tests are always accompanied with tests manuals which report data on test validity and reliability and provide directions for test administration. Such information is usually not available for classroom tests.

sadbhavna

	Norm-Referenced	Criterion Reference
Definition	<p>Norm-Referenced tests measure the performance of one group of test takers against another group of test takers.</p>	<p>Criterion Reference tests measure the performance of test takers against the criteria covered in the curriculum.</p>
	To measure how much a	To measure how much the test taker

Content

Referenced tests measure broad skill areas taken from a variety of textbooks and syllabi.

Criterion-Referenced tests measure the skills the test taker has acquired or finishing curriculum

Item characteristics

Each skill is tested by less than four items. The items vary in difficulty.

Each skill tested by at least four items to obtain an adequate sample of the student

<p>Score interpretation</p>	<p>In Norm-Referenced tests, if a test taker ranks 95%, it implies that he/she has performed better than 95% of the other test takers.</p>	<p>In Criterion Reference the score determines how much the curriculum understood by the test taker.</p>
-----------------------------	--	--

Administration	<p>Norm-referenced tests must be administered in a standardized format.</p>	<p>Criterion-referenced tests need not be administered in a standardized format.</p>
Score reporting	<p>Norm-Referenced test scores are reported in a percentile rank.</p>	<p>Criterion-Referenced test scores are reported in categories or percentages.</p>
	<p>In Norm-Referenced</p>	<p>In Criterion</p>

Unit-II: Assessment Tools

a) Test: Types (Essay, Objective, Objective based);

Essay Tests

Characteristics

The student is given a somewhat general directive to discuss one or more related ideas according to certain criteria. One example of an essay question is "Compare operant conditioning theory and information-processing theory in terms of basic assumptions, typical research findings, and classroom applications".

Advantages

Essay tests reveal how well students can recall, organize, and clearly communicate previously learned information. When well written, essays tests call on such higher-level abilities as analysis, synthesis, and evaluation. Because of these demands, students are more likely to try to meaningfully learn the material over which they are tested.

Disadvantages

Consistency of grading is likely to be a problem. Two students may have essentially similar responses, yet receive different letter or numerical grades. These test items are also very time consuming to grade. And because it takes time for students to formulate and write responses, only a few questions at most can be given.

It is important that instruments of measurement used for research purposes should be properly standardized. Individual score can be compared with the scores of others in a defined group. The evaluation tools serve a variety of uses. Selecting a proper tool is the base for accurate result of evaluation. An evaluation tool may be defined as a sophisticated means of intelligently and scientifically designed to evaluate what is required. Regardless of the type of tool used or how the results of evaluation are to be used, all types of evaluation should possess certain characteristics. The most important characteristics are validity, reliability, objectivity and usability.

5.7.1 Objective Type

An objective type of test item is one which the response will be objective. Objective type test item broadly classified into two:

- i. Supply type (Recall Type)

The respondent has to supply the responses.

- ii. Selection type (Recognition Type)

125

sadbhavna

The respondent has to select the responses from among the given responses.

a. Objective Type – 4 Types

- True – False Items (Alternate Response Type)
- Multiple Choice Items
- Matching Type Items
- Completion Type Test Items

b. Advantages of Objective Type Items

- A large amount of study material can be tested in a very short period time
- Economy of time.
- Objectivity of scoring.
- No bluffing
- It reduces the subjective element of the examiner to the minimum.
- If carefully planned, it can measure the higher mental process of understanding, application, analysis, prediction and interpretation.

c. Limitations of Objective type items

- Difficulty in preparing good items
- Problem of guessing.

sadbhavna

- Help in ensuring wide coverage of content.

i. Advantages of Short answer Type Items

- Large portion of the content can be covered in a test.
- No opportunity for guessing.
- Easy to construct, because it measures a relatively simple outcomes.
- It can be made quite objective by carefully fixing the value points.
- Useful in evaluating the ability to interpret diagrams, charts, graphs, etc.
- If carefully prepared, deep level objectives understanding, application and problem solving skill can be evaluated.

ii. Limitations of Short answer Type Items

- It is more subjective than the objective type of items.
- It may encourage student to memorize fact and develop poor study habits.
- Mechanical scoring is not possible.

5.7.3 Essay type

- It is free response test item.
- Help in ensuring a wide coverage of content and variety of objectives.
- Help in evaluating complex skills.

i. Advantages Essay Type Items

- Easy to prepare.
- Useful in measuring certain abilities and skills.
- Permit the examinee to write down comprehensively what he knows about something.
- Promote originality and creative thinking.
- Possibility of guess work can be eliminated.
- Reduce chance on the spot copying.
- Low printing cost.

characteristics of good tool
(validity, reliability and usability).

1 Validity

Validity is the quality of data gathering instrument which enables to measure what it is supposed to measure. Validity refers to the degree to which the test actually measures what it claims to measure. Validity is also the extent to which inferences, conclusions and decisions made on the basis of test scores are appropriate and meaningful. Validity also refers to whether or not a test measures what it intends to measure. A test with high validity has items closely linked to the test's intended focus. A test with poor validity does not measure the content and competencies it ought to. Validity encompasses the entire experimental concept and establishes whether the results obtained meet all of the requirements of the scientific research method. A quality of a measurement indicating the degree to which the measure reflects the underlying construct, that is, whether it measures what it purports to measure.

a. Different Methods of Validity

Sometimes validity is also thought as utility. Basic to validity of a tool is to measure the right thing or asking right questions. The items of a questionnaire, inventory must appropriately sample a significant aspect of the purpose of the investigation. Validity is not absolute characteristic; it depends on purpose and method used. The six categories of validity are content validity, construct validity, criterion-related validity, concurrent validity, predictive validity and face validity.

i. Content Validity

Content validity refers to the connections between the test items and the subject-related tasks. It is judged by the degree of relationship between diagnostics techniques and achievements in curriculum. The content validity of academic achievement test in subjects is examined by checking the test items against the complete courses of study. The test should evaluate only the outline of the content related to the field of study in a manner sufficiently representative, relevant, and comprehensible. Based on the outline of the content indicating the kinds of knowledge and abilities which the students answer correctly. The overall judgment is based on the extent of agreement between the test and the instructional plan.

ii. Construct Validity

Construct validity is the relationship between the results of a technique of measurement and other indicators of the characteristics that are measured. It implies using the construct (concepts, ideas, and notions) in accordance to the state of the art in the field. Construct validity seeks agreement between updated subject-matter theories and the specific measuring components of the test. This type of validation is often used for measures of a psychological characteristic that is assumed to exist by empirical or theoretical deduction. The general mental ability comprises independent factors such as verbal ability, number ability, perceptual ability, special ability, reasoning ability and memory ability. In order to establish the construct validity of a test, it may be necessary to correlate the results of other tests.

iii. Criterion-Related Validity

It referred to as instrumental validity; it is used to demonstrate the accuracy of a measure or procedure by comparing it with another process or method which has been demonstrated to be valid. For example, imagine a hands-on driving test has been proved to be an accurate test of driving skills. A written test can be validated by using a criterion related strategy in which the

hands-on driving test is compared to it

iv. Concurrent Validity

Concurrent validity refers to the usefulness of a test in closely relating to measures or scores on another test of known validity. Tests are validated by comparing their results with a test of known validity. Concurrent validity indicates the relationship between a measure and more or less immediate behavior or performance of identifiable groups. Concurrent validity is considered when any test is used for the purpose of distinguishing between two or more groups of individuals whose status at the time of testing is different. Concurrent validity is used for statistical methods of correlation to other measures. Once the tests have been scored, the relationship between the examinees' status and their performance (i.e., pass or fail) is estimated based on the test.

v. Predictive Validity

Predictive validity refers to the usefulness of a test in predicting some future performance. Predictive validity is measured by the degree of relationship between a measured and subsequent criteria measure of judgments. This type of validity is used in tests of intelligence, test of aptitudes, vocational interest inventories and projective techniques. This type of validity is especially useful for test purposes such as selection or admissions.

vi. Face Validity

Face validity is the characteristics which appear to measure those which are actually sought to be measured. It is determined by a review of the items and not through the use of statistical analyses. Unlike content validity, face validity is not investigated through formal procedures. Instead, anyone who looks over the test, including examinees, may develop an informal opinion as to whether or not the test is measuring what it is supposed to measure. Face validity is not however a suitable measure of validity, sometimes it might be misleading.

Reliability

Reliability refers to the consistency of measurement, that is, how consistent are evaluation results from one measurement to another. Reliability is concerned with the extent to which an evaluation test is consistent in measuring what it is intended to measure. If the measurements are not consistent over different occasions or over different samples of the same performance domain, the evaluator can have little confidence in the results.

A test score would be reliable when there is good reason to believe that it is stable and trustworthy. These characteristics will depend on the extent to which the score is free from chance error. It is to be expected that the same test which is repeatedly administered on the same group of individuals, should yield the same pattern of scores.

a. Different Methods of Reliability

The various methods of estimating reliability are explained as follows

i. Test-Retest Method

In this test, the same tool or instrument is administered to the same sample on two different occasions. The resulting test scores are correlated and the correlation coefficient provides a measure of stability over a given period of time. If the results are highly stable, those respondents who are high on one administration of test will also be high on the other administration and the other respondents tend to remain in their same relative positions on both administrations. An important factor to be kept in mind is the time interval between tests when interpreting measures of stability. If the time interval is short (say 1-2 days), the consistency of results will be inflated because respondents will remember some of their answers from the first test. If the time interval is quite long (say 1 year), the results will be influenced by the actual changes in the respondent over that period of time. Therefore, the best time interval between test administrations will mainly depend on the use to be made of results.

ii. Equivalent-Forms Method

This method uses two versions of an instrument given to the same sample of respondents. The two forms of the instrument are administered to the same group of respondents in close succession, and the resulting scores are correlated. The correlated coefficient provides a measure of equivalence. It indicates the degree to which both forms of the test are measuring the same aspects of behaviour. The equivalent forms method reflects short term constancy of respondents' performance and the extent to which the test represents an adequate sample of the characteristics being measured.

iii. Split-Half Method

Reliability is also estimated from a single administration of a single form of a test. The test is administered to a group of respondents in the usual manner and then is divided in halves for scoring purposes. To split the test into halves that are most equivalent, the usual procedure is to score the even numbered and the odd numbered items separately. This produces two scores for each respondent, which, when correlated, provide a measure of internal consistency. A reliability coefficient is determined by correlating the scores of two half-tests. The split half method is similar to the equivalent forms method in that it indicates the extent to which the sample of test items is a dependable sample of the content being measured. A high correlation between the scores on the two-halves of a test denotes the equivalence of the two-halves and consequently the adequacy of the sampling. The advantage of this method is that all data for calculation of the reliability coefficient can be collected in one sitting thereby avoiding variations due to two sessions.

iv. Kuder-Richardson Method

Another method of estimating the reliability of test scores from a single administration of a single form of a test is by means of formulas developed by Kuder and Richardson. These formulas provide a measure of internal consistency as with the split-half method but do not require splitting the test in halves for scoring purposes. Kuder-Richardson estimates of reliability provide information about the degree to which the items in the test measure similar characteristics. For a test with relatively homogeneous content, the reliability estimate generally will be similar to that provided by the split half method. In fact, Kuder-Richardson estimate can be thought of as an average of all of the possible split half coefficients for the group tested. It is an advantage when considering tests with relatively homogenous content since the estimate does not depend on the way in which the items are confined to the two half test as in the split-half method. However, for tests designed to measure more heterogeneous learning outcomes, the Kuder-Richardson estimate will be smaller as compared to split half method and the later method is to be preferred.

OBJECTIVITY

Objectivity is a central philosophical concept, related to reality and truth, which has been variously defined by sources. Generally, objectivity means the state or quality of being true even outside of a subject's individual biases, interpretations, feelings and imaginings. A proposition is generally considered objectively true (to have objective truth) when its truth conditions are met and are "bias-free"; that is, existing without biases caused by, feelings, ideas, etc. of a sentient subject. A second, broader meaning of the term refers to the ability in any context to judge fairly, without bias or external influence; this second meaning of objectivity is sometimes used synonymously with neutrality.

a. High objectivity tests

Standardized group tests have high objectivity because they are provided with scoring keys. For example intelligence, achievement, attitude and aptitude tests. These tests have certain standard answers.

b. Moderate objectivity

Binet and Wechsler Bellevue intelligence test administered individually have moderate

objectivity. The evaluation techniques Rorschach and Thematic Appreciation test also have moderate objectivity. The results obtained from test or other evaluative techniques require interpretation. The raw scores on a standardized test becomes much more meaningful when compared with average score obtained by reference groups arranged according to age, grade, years of study and type of person.

USABILITY

While selecting evaluation tool / instrument, practical considerations need to be kept in mind. Generally, the tests are administered by teachers having very limited training. The time available for testing is also limited. The cost of testing is also taken into consideration. All these factors must be taken into account when selecting evaluation tools.

Ease of Administration

Ease of administration is an important aspect if the evaluation instrument or test is to be administered by persons with limited training. For this, it is to be kept in mind that the questions asked are few in numbers. The time provided should be sufficient and the directions should be simple and clear, otherwise, persons who are not properly trained in administering tests may err in giving directions as to how to fill the test. This can have an adverse effect on the validity and reliability of test score.

b. Appropriate Time for Administration

The time provided for administering a test should be appropriate. If an attempt is made to cut down too much on the time allotted to testing, the reliability of the test score may reduce drastically. A safe procedure is to allot as much time as is necessary to obtain valid and reliable results.

c. Ease of Interpretation of Results

The success of an evaluation is determined by the use made of the evaluation results. If they are interpreted correctly, they will contribute effectively in decision-making process. If the results are misinterpreted, they will be of little value. Therefore, ease of interpretation of evaluation results is important, when the results are to be presented.

d. Cost of Administering Tests

The cost of administering test should not be a major consideration as it is comparatively inexpensive. However, in large-scale testing programs, use of separate answer sheets, machine scoring and reusable booklets can reduce the cost appreciably.

Cost

Tests that can be given in a short period of time are likely to gain the cooperation of the subjects and to conserve the time of all those involved in test administration. Teachers should examine the quality of highest quality and that may be obtained within the budget.

b. Administration

The use of any test should be judged in terms of the related competencies of the personnel. Tests that are interesting and enjoyable help to gain the cooperation of the subject. Those are dull or seem silly may discourage the subject. Under these unfavorable conditions, the test is not likely to yield useful results.

c. Scoring Method

Intelligence and personality tests require expertly trained personnel. Test should be reviewed carefully to determine practicability.

d. Mode of Interpretation

The tests such as personality, attitudes, aptitudes and interest should be interpreted with validity. The manual of tests should be studied carefully to determine the feasibility of interpretation of results. Test scores and result must be amenable for interpretation by the

researchers objectivity and uniformly.

e. Time

Suitable short test rather than a longer and more comprehensive one which gives more valid and reliable results. Several short tests will give more complete description of the individual. The length of a test has an important effect upon cooperation interest and effort of the examiner.

f. Availability of Tests

It is essential that parallel forms of the test be available for administration before and after a particular period of time. Therefore, the quality of each item should be important.

b) Construction of a Test: Planning (Blue Print), Preparation, Try Out and Evaluation

ACHIEVEMENT TEST

Any test designed to assess the achievement in any subject with regard to a set of predetermined objectives.

1 Major steps involved in the construction of achievement test

- Planning of test
- Preparation of a design for the test
- Preparation of the blue print
- Writing of items
- Preparation of the scoring key and marking scheme
- Preparation of question-wise analysis

i. Planning of test

- Objective of the Test
- Determine the maximum time and maximum marks

ii. Preparation of a design for the test

- Important factors to be considered in design for the test are:
 - Weightage to objectives
 - Weightage to content
 - Weightage to form of questions
 - Weightage to difficulty level.

iii. Weightage to objectives

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

iv. Weightage to content

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

v. 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.

democratic administration. The administration will be said to be democratic only if it is based on the aforesaid principles.

5.9 ITEM ANALYSIS

Item analysis is a process which examines student responses to individual test items (questions) in order to assess the quality of those items and of the test as a whole. Item analysis is the process of examining the student's responses to each test item to judge the quality of the item. Specifically, what one looks for is the difficulty and discriminating power of the item as well as the effectiveness of each alternative. How hard the items for the group tested, and how well does it distinguish between the more knowledgeable and the less knowledgeable students? These characteristics can be nearly independent of each other, except that a very easy to very hard item cannot discriminate well. If all students mark the item correctly, it has not distinguished between those who know more and those who know less about the concept. If all students mark an item incorrectly, then the item is not discriminating for the group. This information may be important to the teacher for quality-control or diagnostic purpose, but it does not help identify individual differences. Note that an item analysis is no substitute for meticulous care in planning, constructing, criticizing and editing items. Several different item-analysis procedures have been proposed. Some of the more elaborate are appropriate for research projects. The procedures proposed in the section are simple, but adequate for most classroom purposes.

➤ Steps in Item Analysis

In conducting an item analysis of a test consisting of objective type questions one should follow the steps listed (suggested by Ebel, 1972). It requires the six steps outlined in the following.

1. Arrange the scored tests of answer sheets in the order of score, from high to low.
2. Separate two sub groups of test sheets, an upper group, consisting of approximately 27% of the total group, who received highest score on the test, and lower group consisting of an equal number of papers from those who received lowest scores.
3. Count the number of times each possible response to each item was chosen on the papers of the upper group. Do the same separately for the papers of the lower group.
4. Record these responses on the copy of a test

c) Techniques of Assessment: Project work, Assignments, Practical work, Performance based activities.

PERFORMANCE BASED ASSESSMENT

Direct, systematic observation and rating of students' performance of an educational objective, often an ongoing observation over a period of time and typically involving the creation of products. The assessment may be a continuing interaction between teacher and student and should ideally be part of the learning process. The assessment should be a real-world performance with relevance to the students and learning Community.

Assessment of the Performance is done using a rubric or analytic scoring guide to aid in objectivity. Performance based assessment is a test of the ability to apply knowledge in a real life setting. Evaluation of the product of learning experiences can also be used to evaluate the effectiveness of teaching methods.

Stiggins:

Defines this Assessment as used of performance criteria to determine the degree to which a student has met an achievement target. Important elements of performance based assessment include clear goals or performance criteria clearly articulated and communicated to the Learner; the establishment of a sound sampling that clearly envisions the scope of an achievement target and the type of learning that is involved. Attention to extraneous interference and establishment of a clear purpose for the data collected during the assessment before the assessment is undertaken, keeping in mind the needs of the groups involved (teachers, Students, Parents, etc.).

Unit-III: Analysis and Implementation of Assessment

a) Scoring procedure - manual and electronic,

Development of Rubrics

RUBRICS AND ITS IMPORTANCE

The word 'rubric' comes from the Latin word for 'red'. The online Merriam-Webster dictionary lists the first meaning of *rubric* as "an authoritative rule" and the fourth meaning as "a guide listing specific criteria for grading or scoring academic papers, projects, or tests." A rubric is a coherent set of criteria for students' work that includes descriptions of levels of performance quality on the criteria. It should be clear from the definition that rubrics have two major aspects: coherent sets of criteria and descriptions of levels of performance for these criteria.

The genius of rubrics is that they are descriptive and not evaluative. Of course, rubrics can be used to evaluate, but the operating principle is you match the performance to the description rather than "judge" it. Thus rubrics are as good or bad as the criteria selected and the descriptions of the levels of performance under each. Effective rubrics have appropriate criteria and well-written descriptions of performance.

Purpose of Rubrics

Like any other evaluation tool, rubrics are useful for certain purposes and not for others. *The main purpose of rubrics is to assess performances.* For some performances, you observe the student in the process of doing something, like using an electric drill or discussing an issue. For other performances, you observe the product that is the result of the student's work, like a finished bookshelf or a written report. Some common kinds of school performances that can be assessed with rubrics. This by no means covers every possible school performance. It is just meant to help you think of the types of performances you might assess with rubrics.

b) Analysis and Interpretation: Calculation of percentages, Frequency distribution, Percentile Rank, Pie Chart, Bar Graph, Histogram, Frequency Polygon,

sadbhavna

We can compute the percentile rank in ordered data by using the formula,

$$PR\ 60 = 100 - \frac{(100 R - 50)}{N}$$

The student whose rank is the highest are 1 has a percentile rank,

$$\begin{aligned} PR\ (1) &= 100 - \frac{(100 \times 1 - 50)}{10} \\ &= 100 - \frac{50}{10} \\ &= 100 - 5 = 95 \end{aligned}$$

(When there are 10 ranks in the distribution)

7.9.3 Percentage Score

It is often feasible to find the percentage of a given sample which exhibits a certain behaviour or characteristic. When it is impossible to measure these attributes directly given the percentage of occurrence of behaviour, the question often arises of how much confidence we can place in the figure. How reliable an index is our percentage of the incidence of the behaviour in which we are interested. To answer this question we must compute the percentage by the

$$\text{equation } \delta \% = \sqrt{\frac{PQ}{N}}$$

➤ Where

P- the percentage occurrence of the behaviour.

Q – (1-P)

N – Number of cases

➤ Calculation of Percentage Score

In a study among elementary school children, 114 or 41.1% of the 348 children from homes of high socio-economic status were found to have cheated on various tests. Assuming our sample to be representative of children from good homes, how much confidence we place in this percentage? How well does it represent the population percentage?

example, an open ended survey question will yield raw data that cannot be used for statistical purposes as it is; however a multiple choice question will yield raw data that is either easy to convert to a standard score, or even can be used as it is.

7.4 FREQUENCY DISTRIBUTION

In statistics, a frequency distribution is a table 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 or interval, and in this way, the table summarizes the distribution of values in the sample.

A frequency distribution shows us a summarized grouping of data divided into mutually exclusive classes and the number of occurrences in a class. It is a way of showing unorganized data e.g. to show results of an election, income of people for a certain region, sales of a product within a certain period, student loan amounts of graduates, etc. Some of the graphs that can be used with frequency distributions are histograms, line charts, bar charts and pie charts. Frequency distributions are used for both qualitative and quantitative data.

7.5 MEANING OF MEASURES OF CENTRAL TENDENCY

You have an idea generally that an average which is very often used in daily practice. What is that average? You say an average mark of the pupil, an average height or weight of the pupils, an average income of the family and etc. The above concepts average indicates a single value which is the outcome of the total measure. The above typical measures indicate that the values in the data concentrate at the centre or somewhere in the middle of the distribution. Such measures are called measures of central tendency. Tendency of occurrence somewhere in the middle. Here, you are representing the performance of the group as a whole by the single measure and enable you to compare two or more groups in terms of their performance. It

7.9 MEASURES OF RELATIVE POSITION

Measures of relative position indicate where a score is in relation to all other scores in the distribution. In other words, measures of relative position permit you to express how well an individual has performed as compared to all other individuals in the sample who have been measured on the same variable. A major advantage is that they make it possible to compare the performance of an individual on two or more different tests.

➤ Types of Measures of Relative Position

A raw score on a test taken by itself has no value. It gets meaning only by comparing with some reference groups. The comparison is done with certain standard scores. Standard scores provide a method of expressing any score in a distribution in terms of its distance from the mean in standard deviation units. More frequently used relative positions are given below.



7.9.1 Percentiles

A **percentile** (or a centile) is a measure used in statistics indicating the value below which a given percentage of observations in a group of observations fall. For example, the 20th percentile is the value (or score) below which 20 percent of the observations may be found.

We have learn that the median is a point in a frequency distribution below which 50% of the scores and that Q_1 and Q_3 points below which 25% and 75% of the scores lie respectively. Using the same method by which, we may compute points below which 10%, 30%, 70%, 90% and etc., or any other percent of the scores. These points are called percentiles.

Example 1: Distribution of 100 scores on a grammar test.

Class Intervals	Frequency	Cum. Frequency
76 – 80	1	100
71 – 75	2	99
66 – 70	3	97
61 – 65	5	94
56 – 60	12	89
51 – 55	20	77
46 – 50	23	57
41 – 45	21	34
36 – 40	5	13
31 – 35	4	8
26 – 30	3	3
21 – 25	1	1
	N = 100	

$$PR = \frac{100}{N} \left[\text{Cumf} + \frac{(X-L)}{I} \times f \right]$$

Where,

X = the score whose percentile rank is required

N = total number of frequencies

Cumf. = Cumulative frequency up to the class interval containing X.

L = lower limit of the class interval containing X.

f = frequency of the class interval containing X.

I = size of the class interval.

Percentile rank of the score 60 using this formula,

$$\begin{aligned} PR &= \frac{100}{100} 77 + \frac{60-55.5}{5} \times 12 \\ &= 77 + \frac{4.5 \times 12}{5} \\ &= 77 + 10.8 \\ &= 87.8 \end{aligned}$$

7.8 GRAPHICAL REPRESENTATION OF DATA

Any data presented in the form of numbers and tables are not effective and do not create much interest in the reader. If the same data presented in graphs or charts will definitely attracts the attention of the reader and makes the process more effective and very easy to follow, to understand and to remember. Apart from this the devise attracts the attention of the reader.

Visual presentations impress more on the mind of a person and translate numerical values how they have been distributed in the data given.

Drawings are often called graphs. Graphs may be used effectively in understanding a group of test scores other measures in the research process. Thus graphs are more effective, appealing than numerical than numerical data and tables.

7.8.1 Bar- Diagram

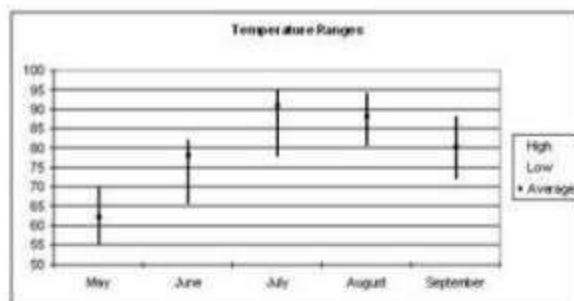
A **bar diagram** is a chart that uses bars to show comparisons between categories of data. The bars can be either horizontal or vertical. Bar graphs with vertical bars are sometimes called vertical bar diagrams. A bar diagram will have two axes. One axis will describe the types of categories being compared, and the other will have numerical values that represent the values of the data. It does not matter which axis is which, but it will determine what bar diagram is shown. If the descriptions are on the horizontal axis, the bars will be oriented vertically, and if the values are along the horizontal axis, the bars will be oriented horizontally.

a. Types of Bar- Diagram

There are many different types of bar diagrams. They are not always interchangeable. Each type will work best with a different type of comparison. The comparison you want to make will help determine which type of bar diagram to use. First we'll discuss some **simple bar diagrams**.

A simple **vertical bar diagram** is best when you have to compare between two or more independent variables. Each variable will relate to a fixed value. The values are positive and therefore can be fixed to the horizontal value.

sadbhavna



Range bar diagram

The difference between a **histogram** and a simple bar graph is that in a histogram, each bar represents a range of dependent variables instead of just one data point.

7.8.2 Pie- Diagram

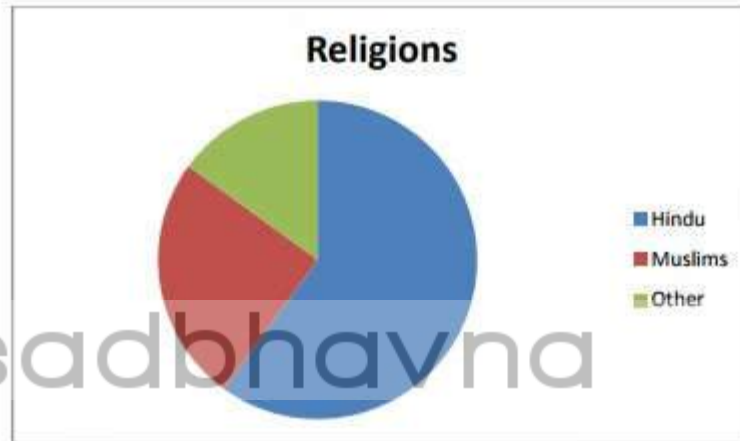
Pie- Diagram is otherwise called as sector graph or angular graph. It is called Pie-diagram because it resembles pie and with the help of a circle the data will be presented hence, also referred as circle graph or sector graph.

A circle may be sub-divided into sectors by subtending the angles at the centre of the circle. The given data is equated to 360 degrees. And each data is expressed in degrees. The sectors formed by the angle measured by the degrees of the data is proportional to the magnitude of the selected data.

❖ Construction of Pie- Diagram

For the following data the Pie- diagram is constructed and represented by means of the given data. In a country out of the total population 60% are Hindus, 25% are Muslims and 15% are other religions say. We can construct a pie-diagram as following.

Since the total degree in a circle are 360 degree, 60% of 360 degree is equal to 216 degree. So by using protractor 216 degree are measured to represent the proportion of Hindus. After presenting this sector, the sector representing the proportion of Muslims should be marked. Muslim constitute 25%. So allotted to the other religions.



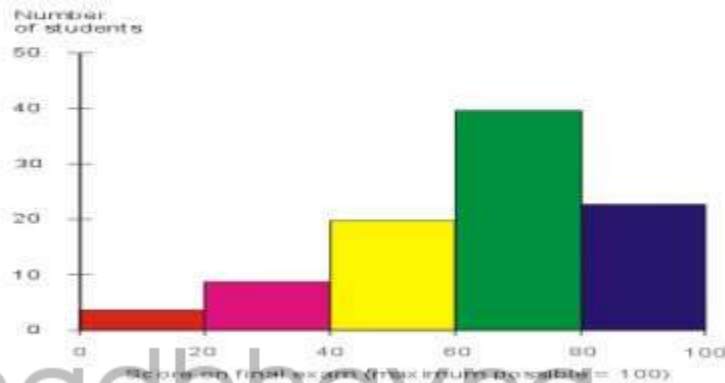
7.8.3 Histogram

Another way of presenting the data by means of a graph is Histogram. Histogram presents an accurate picture of the relative positions of the total frequency from one interval to the other interval. The frequencies within each interval of Histogram are presented by a rectangle, the base of which equals the length of the interval and height of which equals the numbers of the scores within a given interval are presented by the midpoint of the class interval. Whereas in as Histogram the scores are assumed to be spread uniformly over the entire interval, the area of each rectangle is directly proportional to number of measures in the interval. The other type of presenting the data is column diagram.

i. Construction of Histogram

The illustration, below, is a histogram showing the results of a final exam given to a hypothetical class of students. Each score range is denoted by a bar of a certain color. If this histogram were compared with those of classes from other years that received the same test from the same professor, conclusions might be drawn about intelligence changes among students over the years. Conclusions might also be drawn concerning the improvement or decline of the professor's teaching ability with the passage of time. If this histogram were compared with those of other classes in the same semester who had received the same final exam but who had taken

the course from different professors, one might draw conclusions about the relative competence of the professors.



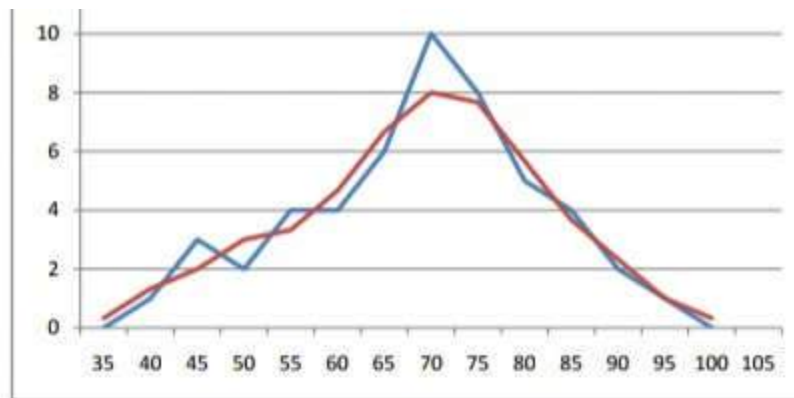
Some histograms are presented with the independent variable along the vertical axis and the dependent variable along the horizontal axis. That format is less common than the one shown here.

ii. Steps

- 1) Draw horizontal line at the bottom of a graph paper along which mark off units to represent the class intervals better to start with class interval of lowest value.
- 2) Draw a vertical line through the extreme end of the horizontal axis along which mark off units to represent the frequencies of the class intervals. Choose a scale which will make the largest frequency (the height of the y-axis) of the polygon approximately 75% of the width of the x-axis.
- 3) Draw rectangles with class units as base, such that the areas of rectangles are proportional to the frequencies of the corresponding class intervals.

iii. Uses

Histogram is the most popular graph used to represent continuous frequency distribution. The width of the height of the rectangle are proportional to the length of the class intervals, the graph thus formed by a series of such rectangles adjacent to one another is called histogram. Thus the area of the histogram is proportional to the total number of frequencies spread on all the class intervals.



As smoothing gives a picture of what an investigator might have brought to his data, had been numerous, less subject to error than the original data, one has to present the original frequency polygon along with the smoothed frequency polygon. If N is large, smoothing may not greatly change the shape of the graph and hence it is often unnecessary.

iv. Uses

The frequency polygon is used in comparing two or more graphs plotted on the same axes. When the repeated scores of frequencies are there in the data, the frequency polygon gives a clear idea about the classified data. However, it is less precise because it does not represent accurately the scores of the frequencies of the class interval but is more useful when the lines of the graph are going to coincide each other.

7.8.5 Cumulative Frequency Curve

Cumulative Frequency Curve is called an Ogive Curve. We convert the cumulative frequencies into cumulative percentage frequencies and then plot the graph with the cumulative percentage frequencies corresponding to the class interval. This curve differs from the cumulative frequency graph. In that cumulative frequency graph, frequencies are not graphed to be expressed in the form of cumulative percents. Therefore, in this graph, the ogive cumulative percents can be calculated by dividing each cumulative frequency by the total frequency. The conversion of cumulative frequencies into cumulative

Unit-IV: Trends and Issues in Assessment

a) Existing Practices: Continuous and Comprehensive Evaluation (CCE),

Grading, Choice Based Credit System,

sadbhavna

$$\sigma \% = \sqrt{\frac{41.4\% \times 58.6\%}{348}} = 2.6\%$$

The sampling distribution of percentages can be taken as normal when N is large. When P is less than 95% and greater than 5%. The σ % is interpreted like the σ mean. In the present problem, 99 confidence interval further population percentage is $41.4\% \pm 2.58 \times 2.6\%$ or 34.7% to 48.1%. We may feel sure that the percentage of children is general who cheated on tests of the sort used in this study will be at least 34.7% and will not be larger than 48.1%.

7.9.4 Grade Point Averages

Your grade point average (GPA) is calculated by dividing the total amount of grade points earned by the total amount of credit hours attempted. Your grade point average may range from 0.0 to a 4.0.

For example:

- A = 4.00 grade points
- A- = 3.70 grade points
- B+ = 3.33 grade points
- B = 3.00 grade points
- B- = 2.70 grade points
- C+ = 2.30 grade points
- C = 2.00 grade points
- C- = 1.70 grade points
- D+ = 1.30 grade points
- D = 1.00 grade points
- D- = 0.70 grade points
- WF/F=0 grade points

P/NP (Pass/No Pass) courses are not factored in the student's GPA. I (Incompletes) and W (Withdrawals) do not receive grade points and do not have an effect on the GPA.

sadbhavna

iii. Weightage to objectives

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

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

iv. 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

v. 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

sadbhavna

vi. Weightage to difficulty level

This indicates the total mark and weightage to be 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

vii. Preparation of the blue print

Blue print is a three-dimensional chart giving the placement of the objectives, content and form of questions.

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	0	2	0	25
Grand Total	3			2			6			8			4			2				

Note: O – Objective Type, SA – Short Answer Type, E – Essay Type

The number outside the bracket indicates the marks and those inside indicates the number of questions.

sadbhavna

viii. Writing of items

- The paper setter writes items according to the blue print.
- The difficulty level has to be considered while writing the items.
- It should also check whether all the questions included can be answered within the time allotted.
- It is advisable to arrange the questions in the order of their difficulty level.
- In the case of short answer and essay type questions, the marking scheme is prepared.
- In preparing marking scheme the examiner has to list out the value points to be credited and fix up the mark to be given to each value point.

Marking Scheme

Q. No	Value points	Marks	Total Marks
1	Value Point – 1	½	2
	Value point – 2	½	
	Value point – 3	½	
	Value point – 4	½	
2	Value Point – 1	½	2
	Value point – 2	½	
	Value point – 3	½	
	Value point – 4	½	

sadbhavna

sadbhavana

ix. Preparation of Question-wise Analysis

Question-wise Analysis

Q. No	Content	Objectives	Form of Questions	Difficulty Level	Marks	Estimated Time (In Mts.)
1	Sub topic – 1	Knowledge	Objective Type	Easy	½	1
2	Sub Topic – 2	Understanding	Objective Type	Average	½	1
3	Sub Topic – 2	Application	Objective Type	Easy	½	1
4	Sub Topic – 1	Knowledge	Objective Type	Easy	½	1
5	Sub Topic – 2	Understanding	Objective type	Average	½	1
5	Sub Topic – 1	Analysis	Short answer	Average	2	3
6	Sub Topic – 1	Synthesis	Short Answer	Difficult	2	3
7	Sub topic – 2	Application	Short answer	Easy	2	3
8	Subtopic – 1	Analysis	Essay	Average	4	10

5.7 TYPE OF TEST ITEMS

- Objective type
- Short answer type
- Essay Type

5.7.1 Objective Type

An objective type of test item is one which the response will be objective. Objective type test item broadly classified into two:

Feedback in Improving learning and learners' development.

b) Issues and Problems: Non-Detention Policy, the menace of coaching.

c) Emerging Practices in Assessment: Standard Based Assessment, Online,

Computer Based and Open Book Examinations.

sadbhavna

6.10.2 Computer - Based Examination

Computer-based testing (CBT) has become widespread in recent years. Some states now use an online platform as the primary delivery mode for one or more computer-based tests used for accountability purposes. When CBT was emerging in state testing in the early 2000s, Thompson, Thurlow, Quenemoen, and Lehr (2002) examined the implications of CBT for students with disabilities. There was not much literature about the use of CBT for large-scale assessments at that time, and Thompson et al. worked with states to explore what needed to be considered during development for students with disabilities and how states might address the needs of these students for accommodations in a CBT environment.

Since the early 2000s, much has occurred in CBT. CBT seems to have advantages over paper and pencil testing, both for states that run the assessment programs and for the students who participate in them. These advantages are recognized by the U.S. Department of Education, which in one of its major initiatives (Race to the Top Assessment Program), encouraged the development of CBT. There currently is strong interest in CBT and advocates have identified

many positive merits of this approach to assessment including: efficient administration, student preference, self-selection options for students, improved writing performance, built-in accommodations, immediate results, efficient item development, increased authenticity, and the potential to shift focus from assessment to instruction (e.g., Becker, 2006; Salend, 2009; Thompson et al., 2002). CBT also allows new ways of assessing students that move beyond the traditional multiple choice and constructed response items. For example, innovative assessments are now being developed that enable students to manipulate data and role play. Yet, as states move forward with CBT they are discovering that it is important to consider not only the positive benefits, but also potential negative unintended consequences. These include, for example, the possibility that additional training will be needed for students with disabilities to interact successfully with computers and the challenges of determining the best way to present some accommodations such as screen readers.