SHRI MADHAV COLLEGE OF EDUCATION & TECHNOLOGY HAPUR

MONIKA POSWAL

Unit I: Measurement, Assessment and Evaluation

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.

Meaning of Measurement

Measurement is actually the process of estimating the values that is the physical quantities like; time, temperature, weight, length etc. each measurement value is represented in the form of some standard units. The estimated values by these measurements are actually compared against the standard quantities that are of same type. Measurement is the assignment of a number to a characteristic of an object or event, which can be compared with other objects or events. The scope and application of a measurement is dependent on the context and discipline.

Meaning of Evaluation

Evaluation is a broader term that refers to all of the methods used to find out what happens as a result of using a specific intervention or practice. Evaluation is the systematic assessment of the worth or merit of some object. It is the systematic acquisition and assessment of information to provide useful feedback about some object.

Interrelation among Assessment, Evaluation and Measurement Though the terms assessment and evaluation are often used interchangeably (Cooper, 1999), many writers differentiate between them. Assessment is defined as gathering information or evidence, and evaluation is the use of that information or evidence to make judgments (Snowman, McCown, and Biehler, 2012). Measurement involves assigning numbers or scores to an "attribute or characteristic of a person in such a way that the numbers describe the degree to which the person possesses the attribute". Assigning grade equivalents to scores on a standardized achievement test is an example of measurement.

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.

CHARACTERISTICS OF CLASSROOM ASSESSMENT The different characteristics of classroom assessment are given below.

- Learner-Centered- The primary attention of teachers is focused on observing and improving learning.

- Teacher-Directed- Individual teachers decide what to assess, how to assess, and how to respond to the information gained through the assessment. Teachers do not need to share results with anyone outside of the class.

- Mutually Beneficial Students are active participants. Students are motivated by the increased interest of faculty in their success as learners. Teachers improve their teaching skills and gain new insights.

- Formative Assessments are almost never "graded". Assessments are almost always anonymous in the classroom and often anonymous online. Assessments do not provide evidence for evaluating or grading students.

- Context-Specific Assessments respond to the particular needs and characteristics of the teachers, students and disciplines to which they are applied. Customize to meet the needs of students and course.

- Ongoing Classroom assessment is a continuous process. Part of the process is creating and maintaining a classroom "feedback loop". Each classroom assessment event is of short duration.

- Rooted in Good Teaching Practice Classroom assessment builds on good practices by making feedback on students' learning more systematic, more flexible and more effective.

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

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

Teacher-Made vs. Standardized Assessments In the broadest sense, assessments may be classified into two categories: teacher-made and standardized. Teacher-made assessments are constructed by an individual teacher or a group of teachers in order to measure the outcome of classroom instruction. Standardized assessments, on the other hand, are commercially prepared and have uniform procedures for administration and scoring. They are meant for gathering information on large groups of students in multiple settings.

MODE OF RESPONSE a. Oral Response and Written Assessments Student oral responses are longer and more complex, parallel to extended written response questions. Just as with extended written response, one evaluates the quality of oral responses using a rubric or scoring guide. Longer, more complicated responses would occur, for example, during oral examination or oral presentations. Written assessments are activities in which the student selects or composes a response to a prompt. In most cases, the prompt consists of printed materials (a brief question, a collection of historical documents, graphic or tabular material, or a combination of these). However, it may also be an object, an event, or an experience. Student responses are usually produced ―on demand,‖ i.e., the respondent does the writing at a specified time and within a fixed amount of time. These constraints contribute to standardization of testing conditions, which increases the comparability of results across students or groups.

Unit II: Assessment Tools

TESTS

Tests may be of different forms, such as psychological test used to measure mental and behavioural traits, achievement test to assess performances of students, etc. They can be used to assess both the scholastic and co-scholastic abilities of the students. Let us now understand the concept of the test. The tests are those instruments by which you, as a teacher, collect information as data through verbal and non-verbal responses of the students. A concise definition may be : a test is an instrument or systematic procedure for measuring a sample of behaviour (Gronlund, 1990).Further, a psychological test is defined as a standardized, repeatable procedures used to elicit and measure samples of human behaviour' (Kazdin, 2000). From the above two definitions, you can summarize the meaning of a test and the sample behaviour it measures as follows :

- Human abilities, including intelligence, aptitudes, skills, and achievement in various areas.
- Personality characteristics, which include traits, attitudes, interests, and values.
- Adjustment and mental health, which involves detecting signs and symptoms of psychological and neurological disorders and appraising the effectiveness of psychological functioning.

A test can be used for two broad objectives. First, it attempts to compare the same student on two or more than two aspects of a trait, such as attitude and aptitude of the same student; and second, two or more than two students may be compared on the same trait like attitude of two students.

Some tests have been discussed in the subsequent sub-sections.

Paper Pencil Test

The paper pencil tests comprise a standard set of questions which are presented to the student in writing on paper or OpticalMark Recognition (OMR) sheets that requires completion of cognitive tasks in the formof response by the student on those papers by pencil/pen mark. These responses or answers are summarized/scored to obtain a numerical value that represents a characteristic of the student for which the test was administered. The paper pencil tests can focus on what the student knows (achievement), is able to learn (ability or aptitude), chooses or selects (interests, attitudes or values), or is able to do (skill).

Oral Test

Oral tests are those tests in which the response, solution or the task requires oral response to answer the question. Teachers' conversation with the students for the purpose of assessment, component of viva-voce examination for completion of a course, etc. can be called oral test. You might have noticed that in a written test, very little scope is left for the students to express themselves on any aspect whereas in oral test, students enjoy freedom to express themselves by citing many examples. In oral test, you can also use figures, diagrams, charts, maps, models, signs, etc. for asking students to explain the concepts covered in it. Oral tests have proved most valuable when used with students having language disabilities, the illiterates, shy feeling and the young children.

Aptitude Test

An aptitude test is also an instrument used to determine and measure an individual's ability and skills to acquire, through future training. Aptitude tests may be classified into two groups : multiple aptitude test, and special aptitude tests. Multiple aptitude tests are those which intend to measure various areas of aptitude (musical, mechanical, etc.) each by independent sub-tests, whereas the special aptitude tests measure only one specific aptitude like teaching, musical, etc. The multiple aptitude tests measure abilities of students in more than one area

simultaneously by using different sections of a test while the special aptitude tests measure ability of the students in one area. Aptitude tests are used to predict the future performance of the students. These predictions are for the performance based on specific criterion (for example you all appeared before an entrance test for admission to the B.Ed. programme. The test included teaching aptitude components required to get into the B.Ed. programme) which are prior to instruction, placement or training. The aptitude tests are also used for guidance, as well as prediction of success in some occupation. Training or academic courses are possible on the basis of scores on standardized aptitude tests. For example, for pursuing teacher education programmes, one needs to have teaching aptitude; for pursuing medical courses, one needs to have medical aptitude; for pursuing engineering courses, one needs to have engineering aptitude; for musical courses and sports related courses, one requires to have musical and the sports aptitude respectively.

Achievement Test

Achievement tests are administered on students to measure their learning outcomes. These tests are more prevalent in our schools. These tests show as to what has been learned by the students, rather than to predict future performance as in the case of aptitude tests. There are teacher made and standardized achievement tests. The choice of your achievement test will depend upon your purpose. The achievement tests can be used for following purposes:
to know the learning progress of the students.

- checking any weakness in the instruction.
- in formulation of learning objectives and provide an easy means of critical

examination of the content and the methods of teaching.

- adapting the instruction to the need of the individual learner.
- to know the effectiveness of any academic programme.

The achievement tests can be classified as :

- General achievement tests (batteries)
- Special achievement tests

General achivement tests attempt to measure the general educational achievement of the students at different stages which includes the common subject areas taught at the school in a particular class. Special achievement tests are meant for measuring the achievement of students in selected areas which may be grouped into two distinct groups – the diagnostic tests and the standardized summative achievement tests. Diagnostic tests are used to know the areas of difficulties of the students and accordingly to provide suitable remedial instructions to them. Standardized summative achievement tests are used to grade the students in a particular standard and also to certify them.

Intelligence Test

The term ‘intelligence’ is difficult to define in a single sentence acceptable to all. But intelligence can be understood as a general set of mental traits, which is often reported as the mental abilities. Test of mental ability or intelligence tests measure convergent thinking. Convergent thinking is a process of finding out the solution of a problem. Guilford defines convergent thinking is the ability to give correct answer to standard questions that do not require creativity. For instance, most of the school based tasks can be done through different tests. Many psychologists and educationists opined that intelligence is a product of heredity and environment, but it is a matter of debate to get the real contribution of heredity and environment for intelligence of an individual. Many researches you will find in this regard. In this section, our intension is to understand the measurement of intelligence and the tests used for measuring intelligence.
Intelligence tests are used to provide a very general measure, usually reporting

a global test score. As they are general, intelligence tests are useful in predicting a wide variety of tasks. Intelligence tests can be classified on various backgrounds.

On the basis of administration it can be classified into two categories :

a) Individual tests

b) Group tests

TOOLS

To measure the scholastic abilities of the students, we usually use the tests such as achievement, intelligence, aptitude, etc., but for measuring most of the co scholastic abilities, we use the tools such as rating scale, inventories, checklists, schedules, questionnaires, etc. In simple language, you can understand a tool as an instrument to measure something. In education, the word tool is used to measure various traits of students. Now, let us discuss each tool with details.

Rating Scales

Rating scale is one of the important tools widely used in psychology and education. It is used for assessing attitudes of students on any situation, idea, object, character, person or an attribute. In a rating scale, the opinions are given in various degrees such as strongly agree, agree, and disagree; highly satisfied, satisfied, and dissatisfied, etc. Arating scale is prepared always in odd number points like 3-point scale, 5-point scale, 7-point scale, or 9-point scale. It is in odd number points because, a definite middle measuring point will be possible only when the scale is odd points. You might have observed that many persons have the attitude to opine their opinions at the middle rating. Rating scales can be presented in different categories. The most commonly used categories of rating scales used in the schools are as follows :

(i) Numerical scale

(ii) Graphic scale

Questionnaire

A questionnaire is a device comprising a series of questions dealing with some psychological, social, and educational topic(s) given to an individual or a group of individuals, with the object of obtaining data with regard to some problems under investigation. Questionnaire is a common tool that we, the teachers usually apply to collect data from a situation, condition or prevailing practice.

The questionnaire can be classified in terms of the nature of the questions constructed to gather the data from the stakeholders. It may be closed or open ended. *Closed questionnaires* are those where the respondents answer in limited way, like responding in ‘yes’ or ‘no’; underlining the replied among the predefined responses, putting the sign ‘correct’ or ‘incorrect’. Whereas in *open questionnaires* the respondents are free to share, clarify and put their view.

Let us discuss certain examples of closed and open ended questions.

Examples of open-ended questions :

· Describe any one of the wars fought for independence in India.

· Write an essay on Quit India Movement.

Examples of closed-ended questions :

· In which year Quit India Movement took place?

a) 1941 b) 1942 c) 1943 d) 1944

Advantages of questionnaires : Some of the advantages of questionnaires are enumerated as follows :

- The responses are gathered in a standardized way.
- Questionnaires are more objective, certainly more so than interviews.
- Relatively quick to collect information.
- Potential information can be collected from a large portion of a group.

Limitations of questionnaires : Although the questionnaire is one of the widely used tools but it has also certain limitations which can be enumerated as follows :

- Not easy to be used with children or illiterates.
- Respondents may not agree to respond in writing.
- Sometimes it is difficult to construct questions on complex and crucial topics.
- Respondent interprets the questions from his/her angle of understanding.
- External factors may affect the response

Inventories

An inventory is constructed in the form of a questionnaire. But the inventories are more exhaustive than questionnaire. Inventories have been mostly used for measuring personality traits, interests, values and adjustments i.e. for assessing self-reporting affective domain of behaviour. It consists of a series of questions or statements to which the subject responds to by answering 'Yes' or 'No', or 'Agree' or 'Disagree'. This can also be answered in some similar ways to indicate preferences or to make those items that describes the subject's typical behaviour.

In the inventory, the statements are put in first person, For example, "I think I am comparatively more tense than others". In the questionnaire, there is a question in a second person, for example "Do you think you are more tense than other persons around you?"

Checklist

According to Koul (1997), a checklist is a simple device consisting of a prepared list of items which are prepared by the teacher to be relevant to the problem being studied. After each item a space is provided for the observer to indicate the presence or absence of the item by checking 'yes' or 'no', or a type of number of items may be indicated by inserting the appropriate word or number. The checklist is a systematic and quick way to gather data of the relevant factors and take actions accordingly. A very simple example is being given below :

For example :

Is the school building fire proof? Yes/No

Does the school follow rain harvesting policy? Yes/No

Is the school building earthquake proof? Yes/No

It is noteworthy that the responses collected on the checklist are as a matter of fact not any judgment. It is a good tool in gathering facts for educational survey, checking your school library, laboratory, game facilities, school buildings, textbooks, etc. It may also be used to check the availability of other facilities in your school.

Interview Schedule

Interview is a communication or conversation bywhich a person asks interviewer and interviewee responds verbally in the face-to-face situation. An interview can also be conducted through skype electronic media. It can be conducted through telephonic conversation or through by using internet. It is a common technique for collecting required information about an individual.

The interview schedule is a tool with the help of which the interview is conducted.

Interview schedule can be classified according to the purpose for which it is structured and used. If it is to resolve the research hypothesis, it is a research interview schedule and if it is for clinical purpose, it a clinical interview schedule. On the bases of the structure, the interview schedules are categorized as
structured or unstructured.

· **A structured** interview schedule is one in which the procedure to be followed is standardized and it is determined in advance of the interview. The same types of questions are presented to the interviewee and the wordings of the instructions to the interviewee are specified.

· In the **unstructured interview** schedule**,** the series of questions are decided in advance but the interviewer is largely free to reorganize the questions and timing to attain the objective of the interview.

Observation Schedule

Observation has been the first practice of assessment that we do in our classroom. Each observed incident, expression and reaction has useful data for teachers, hence the observation is an effective tool for us. Observation is the process in which one or more persons observe what is happening in a real life situation, and he/she classifies and records pertinent happening according
to pre-planned scheme. It is used to evaluate the overt behaviour use uniformly either American or British English i.e., spellings of words, etc. of individuals in both the controlled and uncontrolled situations. The observation schedules are the enumerations, listing of the facts or other data that are observed under observation process. Like questionnaires, the observation schedules are also classified as structured or unstructured. This can also be classified as participant and non-participant observation.

Merits of observation : The following are some of the merits of observation

· Through observation the data are gathered from a natural setup.

· The data observed are direct or first hand.

· As the data are first hand or direct, while doing observation we can correlate what is being said and what is being shown.

· Does not rely upon people's willingness or ability to provide information.

Demerits of observation : The observation has also the following demerits :

· The observation if not done with planning then it will not bring out authentic information.

· Observer's biases may affect the result.

· There may come '*Hawthorne effect'*, that is when the person gets to know
that he/she is being observed then the real problems may not be shared or shown.

· Does not develop the understanding why people behave in a particular manner.

Unit III: Standardization of Measuring Instrument

ESSENTIAL CRITERIA OF AN EFFECTIVE TOOL OF EVALUATION:

The essential qualities or criteria of an effective tool may be as follows:

- Reliability
- Validity
- Usability
- Objectivity
- Norm

Reliability

Reliability is the important criteria of a good test/tool. Reliability refers to consistency. A test which shows a consistent result in its frequent uses in different situations and places is called reliability of the test. The other synonyms that can be used for getting reliability of the test are: dependability, stability, consistency, predictability, accuracy, etc. It implies that the reliable test always provides a stable, dependable, accurate and consistent result in its subsequent uses.

Methods or techniques of reliability : There are three common methods of estimating the reliability coefficient of test scores. These methods are :

(i) Test-retest reliability.
(ii) Parallel-forms reliability.
(iii) Internal consistency reliability.

Let us discuss each method of reliability in detail.

(i) Test-retest reliability : Test-retest reliability means the same test is administered twice on the same group of sample within a given time interval and correlation is calculated between the two sets of scores (first and second administration). If the coefficient of correlation is positive and high, it is considered that the test is reliable. Let us discuss the procedures of using test-retest reliability.

Limitations of using the method : The following are the limitations of the test-retest method :

- As the same test is administered twice on the same group, there will be the threat of carry over effect, it means, during the second administration, the candidates may remember many items from the first administration.
- The scoring of second administration is usually high than the first one.
- Maintaining a gap of time between test and re-test is also again one of the important aspects to determining exact value of reliability. If time gap is very less, then carry over effect will be high and on the other side, if time gap is very high, maturity effects of the candidates may hamper the test results.
- This method is not free from errors. Memory, carry over, practice and maturity effects are high in this technique.

(ii) Parallel-form Reliability : Because of the error factors in test-retest method, parallel-form method is one of the alternate methods of the test retest method and it can minimize many of the errors occurred in the earlier method. In the parallel form method, two parallel tests are prepared keeping in consideration equivalence in all aspects such as similarities in content, objectives, types and number of items, time allowed in both the tests, level of difficulty, discrimination value, conditions of use, etc. The main effort by doing these is to make two equivalent forms of test.

Limitation of parallel form method : Parallel form method is also not completely free from errors. There are possibilities of making errors in this method also:

- Practice and carry over effect is not totally minimized, as both the tests are equivalent in nature in many respects except only the items are different and a time interval of 15 days to 6 months is given for testing the second form of the test. During this period, there is a chance that the students may practice the similar content and items and hence chances for getting better scores in second testing are generally more.
- Preparing two parallel forms of the tests is also a complex task.
- This method is comparatively time taking to get the reliability.

(iii) Internal consistency reliability : Internal consistency reliability indicates the homogeneity of the test. If all the items of the test measure the same function or trait, the test is said to be a homogeneous one and its internal consistency reliability would be pretty high. The most common methods of estimating internal consistency reliability are the (a) Split-half method and (b) Rational equivalence method. Let us discuss split-halfmethod first.

(a) Split-half method : This method is also called as 'odd-even method'.

Let us discuss the procedures and conditions of using split-half method for determining reliability.

Validity

In the preceding sub-section you have studied about the reliability. Let us now discuss about validity. Validity tells us about the accuracy and truthfulness of a test. The accuracy of a test can be said if and when the test measures the the purpose it was constructed. Cronbach(1970) defines validity as 'validity is the extent to which a test measures what it purports tomeasure'. Further Freeman (1965) defines validity as 'an index of validity shows the degree to which a test measures what it purports to measure when compared with accepted criterion.'

From the above three definitions given we can say that validity talks not only about consistency but also accuracy and truthfulness of the test results.

Characteristics of validity : The following are the main characteristics of validity of a test :

i. Validity is an index of external correlation. The test scores are correlated with external criterion scores such as the test scores will be correlated with an earlier developed valid test prepared for measuring the same aspect.

ii. The criterion may be a set of operation, purpose or predictor for future course of performance.

iii. It deals with the psychological construct of a variable which is indirectly measured with the help of behaviours.

iv. No test in education and psychology is perfectly valid because measurement is indirect.

v. Validity endorses the reliability of a test. If a test is valid, it must be reliable, but a reliable test may or may not be valid.

vi. It refers to the truthfulness or purposiveness of test scores.

vii. It indicates the degree to which the test is capable to achieve the aims for which it is developed.

viii. Validity is best considered in terms of matter of degrees, such as high, moderate and low validity

Types of validity

Commonly, five types of validity are used in preparing tools. These are :

(i) Face Validity

(ii) Content Validity

(iii) Criterion-related Validity

· Predictive validity
· Concurrent validity
(iv) Construct Validity
(v) Factorial Validity
Let us understand the concept of the above types of validity.
(i) Face validity : Face validity is the first step to know validity of a test. This is also called validation by face. This method is not widely used because it never analyses the entire test and its items to determine the validity of the test. In the face validity, the appearance of the test, purpose of its construction, objectives it covers, dimensions it measures, language used, etc. are taken into consideration for determining the face validity of the test. This is the lowest level of determining validity of the test. This method can only be used in case of shortage of time for using other methods of validity. Further, before using other methods of determining validity, a judgment is taken, whether the test is validated by face or not. If the test lacks face validity then usually other methods of validity are not determined.
(ii) Content validity : Content validity is the second level of validity of the test. In this method, the format as well as the content of the test is examined and decision is taken for its validity.Anastasi (1968) defines content validity as, 'it involves essentially the systematic examination of the test content to determine whether it covers a representative sample of the behaviour domain to be measured.' Content validity basically matches the test items with the instructional objectives. Content validity is very important for achievement test. To know the content validity of a test, usually the items are examined as per the blue-print/table of specification on which the test is prepared.
(iii) Criterion-related validity : Unlike content validity, criterion-related validity can be objectively measured and declared in terms of numerical indices. The concept of criterion-related validity focuses on a set of 'external' criterion. The external criterion may be data of 'concurrent' information or of a future performance. The criterion related to concurrent information is called as 'Concurrent Validity' and criterion related to future performance is called as 'Predictive Validity'.
· **Predictive validity :** Predictive validity refers to the predictive capacity of a test. It indicates the effectiveness of the test in forecasting or predicting future outcomes in a specific area. In short, predictive validity determines how far the test is able to predict future result.
This can be better understood by an example. Suppose we have to prepare a medical entrance examination test to admit students in medical courses. The predictive validity of the test can be determined only when those qualified students who took admission in medical course performed well in the medical final examination. Predictive validity is a time consuming method. To get predictive validity of a test, you have to wait till the completion of the course. Sometimes, prediction can also take much time to correlate it with further criterion
such as those students who performed well in the entrance examination, whether they successfully completed the course or not, and further, those students who successfully completed the course did get a job/ placement or not. So, to determine the predictive validity of a test, there is a need to establish correlation of the scores between entrance examination result and the course completion result. If the correlation coefficient is positive and high, you can say that the test is valid. This type of validity is sometimes referred to as 'empirical validity' or 'statistical validity' as our evaluation is primarily empirical and statistical. You can test the validity empirically.
· **Concurrent validity :** Concurrent validity refers to the extent to which the test scores correspond to already accepted measures of performance. For example, suppose you have prepared a test of 'intelligence' and you want to know the concurrent validity of the test, you have to correlate the scores of the test administration with the scores of another established standardized test. Let us understand it with the help of

another example. The Intelligence test, which you have prepared and the intelligence test prepared by Stanford-Binet can be administered among the same group of students and correlation coefficient of two sets of scores can be determined. If the coefficient of correlation is high, we can say that the test has concurrent validity.

(iv) Construct validity : Construct validity is also called 'psychological or trait validity'. Construct validity means that the test scores are examined in terms of a construct. For example, the construct for achievement of a student may be his/her intelligence, practice, aptitude, interest, attitude, etc. Construct validity can be defined as the extent to which the test may be said to measure a theoretical construct or trait or psychological variable.

In construct validity the variables related to the test which contributes to that aspect are correlated and examined.

For example, this is a theoretical fact that intelligence and achievement are positively correlated with each other. Suppose you have to prepare an intelligence test and you want to know the construct validity of the test. For that, you have to correlate the intelligence test scores of the students with their achievement test scores. The assumption here is that those students who have done well in intelligence test will naturally do well in achievement test, because as per the theory, both are positively correlated with each other. In case the correlation is negative, it can be said that the intelligence test is lacking construct validity. This can also be correlated with other theoretical and psychological construct of an intelligence test with the assumptions as follows :

- Intelligence and achievement are positively correlated with each other.
- Intelligence and aptitude are positively correlated with each other.

Construct validity is to the extent test results are interpreted in terms of known psychological concepts and principles. Certain common examples of theoretical constructs of most psychological tests are intelligence, scientific attitude, critical thinking, reading, comprehension, study skills and mathematical aptitude, etc.

(v) Factorial validity : Factorial validity determines the correlation of the different factors/components with the whole test. Factorial validity is determined by a statistical technique known as factor analysis. It uses methods of expansion of inter-correlations to identify factors (which may be verbalized as abilities) constituting the test. The correlation of the test with each factor is calculated to determine the weight contributed by each such factor to the total performance of the test. This validity tells us about the factor loading. The factors responsible for achievement of students are called factor loading. This relationship of the different factors with the whole test is called the factorial validity

Factors affecting validity: A large number of factors influence the validity of the test. Gronlund (1981) has suggested the following factors:

i. Factors in the test itself:

The following factors that affect validity of a test are included in the test itself. These are also called as intrinsic factors.

- **Unclear direction :** If directions regarding how to respond to the items, whether it is permissible to guess and how to record the answers, are not clear to the pupil, then the validity will tend to reduce. Hence, clear direction should be given in the test.
- **Reading difficult vocabulary and sentence structures:** The complicated vocabulary and sentence structure meant for the student taking the test may fail in measuring the aspects of pupil performances; thus it results in lowering the validity.
- **Inappropriate level of difficulty of the test items:** When the test items have an inappropriate level of difficulty, it will affect the validity of the tool. For example, in criterion referenced test, failure to match the difficulty specified by the learning outcome will lower the validity.

· **Poorly constructed test items :** The test items which provide unintentional clues to the answer will tend to measure the pupils' alertness in detecting clues as well as the aspects of pupil performance which ultimately affect the validity.

· **Ambiguity :** Ambiguity in statements in the test items leads to misinterpretation, multi-interpretations and confusion. Sometimes, it may confuse the good students more than the poor ones resulting in the discrimination of items in a negative direction. As a consequence, the validity of the test is lowered.

· **Test items inappropriate for the outcomes being measured :**

Many a times we try to measure certain complex types of achievement, understanding, thinking, skills, etc. with test forms that are appropriate only for measuring factual knowledge. This affects the results and leads to a distortion of the validity.

· **Test too short :** Atest usually represents a sample of many questions. If the test is too short to become a representative one, then validity will be affected accordingly.

· **Improper arrangement of items :** Items in the test are usually arranged in terms of difficulty with the easiest items first. If the difficult

Usability

Usability refers how successfully you, as a teacher, use the test in a classroom situation. It has been observed that, many highly valid tests lack the quality of usability. The user fails to understand or feels it difficult to use the test. Therefore, a good test should have the quality of usability. While selecting an evaluation tool, you should look for certain practical considerations like easy for administration and scoring, easy for interpretation, availability of comparable forms and cost of testing. All these considerations induce a teacher to use tools of evaluation and such practical considerations are referred to as the 'usability'of a tool of evaluation. In other words, usability means the degree to which the tool of evaluation can be successfully used by the teacher and school administrators. So usability of a test includes comprehensibility, easy for administration and scoring, easy for interpretation, appearance of the test,economy and availability of the test for use.

Objectivity

Objectivity is another important feature for a good test. Objectivity of a test refers to two aspects of the test, viz;

· Item-objectivity, and

· Scoring-objectivity.

Item-objectivity means the item is having only one right answer. Many a times, you might have observed that a single item has two or more related answers. That affects the validity of the test. Apart from these, ambiguous questions, lack of proper direction, double-barreled questions, questions with double negatives, essay type questions, etc. do not have objectivity. So much care has to be exercised while framing the questions. Scoring-objectivity refers the test paper would fetch the same score. Scoring objectivity can be ensured by carefull maintaining the item-objectivity. In objective-type items, it is easy to ensure scoring objectivity whereas in subjective item, certain precautions needs to be taken to ensure scoring-objectivity such as carefully phrasing the essay items, making proper directions of scoring, making the items short-answer type instead of essay type item, etc.

Norm

Generally norm is considered as a standard, but technically, there is difference between the concepts of normand standard. Normcan be defined as the average or standard score on a particular test made by a specified population.' Thorndike and Hegen (1977) defines norm as 'average performance on a particular test made by a standardized sample.' Determining

normis one of the important criteria of a good test. Most standardized tests determine norm. Norm can be characterized as follows :

- It acts as a basis for interpreting test scores and minimize interpretive error of the test.
- It helps to transform the raw scores into standard scores or derived scores and put meaning to it.
- Norm suggests a level and therefore the individual departure from the level can be evaluated in quantitative term.
- Norms are necessary for the purpose of promotion, gradation, selection and classification of examinees.
- It refers to the average performance on a particular test made by standardized sample or specified population.

Unit IV : Data and Measures of Central Tendencies

MEASURES OF CENTRAL TENDENCY The following are the three measures of average or central tendency that are in common use :
(i) Arithmetic average or arithmetic mean or simple mean (ii) Median (iii) Mode

ARITHMETIC MEAN To find the arithmetic mean, add the values of all terms and them divide sum by the number of terms, the quotient is the arithmetic mean. There are three methods to find the mean : (i) Direct method: In individual series of observations x1, x2,… xn the arithmetic mean is obtained by following formula. 1 2 3 4 1..... n n x x x x x x A M n + + + + − = (ii) Short-cut method: This method is used to make the calculations simpler. Let A be any assumed mean (or any assumed number), d the deviation of the arithmetic mean, then we have . fd M A N = + ∑ (d=(x-A)) (iii)Step deviation method: If in a frequency table the class intervals have equal width, say i than it is convenient to use the following formula. fu M A i n = + × ∑ where u=(x-A)/ i ,and i is length of the interval, A is the assumed mean.

MEDIAN The median is defined as the measure of the central term, when the given terms (i.e., values of the variate) are arranged in the ascending or descending order of magnitudes. In other words the median is value of the variate for which total of the frequencies above this value is equal to the total of the frequencies below this value. Due to Corner, ―The median is the value of the variable which divides the group into two equal parts one part comprising all values greater, and the other all values less then the median‖. For example. The marks obtained, by seven students in a paper of Statistics are 15, 20, 23, 32, 34, 39, 48 the maximum marks being 50, then the median is 32 since it is the value of the 4 th term, which is situated such that the marks of 1st, 2nd and 3rd students are less than this value and those of 5th, 6th and 7th students are greater then this value. COMPUTATION OF MEDIAN (a)Median in individual series. Let n be the number of values of a variate (i.e. total of all frequencies). First of all we write the values of the variate (i.e., the terms) in ascending or descending order of magnitudes Here two cases arise:
Case 1. If n is odd then value of (n+1)/2th term gives the median. Case2. If n is even then there are two central terms i.e., n/2th and 1 2 th ⎛ ⎞ n + ⎜ ⎟ ⎝ ⎠ The mean of these two values gives the median.
(b) Median in continuous series (or grouped series). In this case, the median (Md) is computed by the following formula 2 d n cf M l i f − = + × Where Md = median l = lower limit of median class cf = total of all frequencies before median class f = frequency of median class i = class width of median class.

MODE The word mode is formed from the French word La mode‘ which means in fashion‘. According to Dr. A. L. Bowle ‗the value of the graded quantity in a statistical group at which the numbers registered are most numerous, is called the mode or the position of greatest density or the predominant value.‘ Mode According to other statisticians, The value of the variable which occurs most frequently in the distribution is called the mode.‘ The mode of a distribution is the value around the items tends to be most heavily concentrated. It may be regarded at the most typical value of the series‖.
Definition. The mode is that value (or size) of the variate for which the frequency is maximum or the point of maximum frequency or the point of maximum density. In other words, the mode is the maximum ordinate of the ideal curve which gives the closest fit to the actual distribution.
Method to Compute the mode: (a) When the values (or measures) of all the terms (or items) are given. In this case the mode is the value (or size) of the term (or item) which occurs most frequently.

Unit V: Measurs of Variability & Correlation

Theoretical Base of the Normal Probability Curve The normal probability curve is based upon the law of Probability (the various games of chance) discovered by French Mathematician Abraham Demoiver (1667-1754). In the eighteenth century, he developed its mathematical equation and graphical representation also. The law of probability and the normal curve that illustrates it are based upon the law of chance or the probable occurrence of certain events. When any body of observations conforms to this mathematical form, it can be represented by a bell shaped curve with definite characteristics.

The characteristics of the normal probability curve are: 1) The Normal Curve is Symmetrical: The normal probability curve is symmetrical around it's vertical axis called ordinate. The symmetry about the ordinate at the central point of the curve implies that the size, shape and slope of the curve on one side of the curve is identical to that of the other. In other words the left and right halves to the middle central point are mirror imagesThe Normal Curve is Unimodel: Since there is only one maximum point in the curve, thus the normal probability curve is unimodel, i.e. it has only one mode.

3) The Maximum Ordinate occurs at the Center: The maximum height of the ordinate always occur at the central point of the curve, that is the mid-point. In the unit normal curve it is equal to 0.3989. 4) The Normal Curve is Asymptotic to the X Axis: The normal probability curve approaches the horizontal axis asymptotically; i.e. the curve continues to decrease in height on both ends away from the middle point (the maximum ordinate point); but it never touches the horizontal axis. Therefore its ends extend from minus infinity (- ∞) to plus infinity (+ ∞).

5) The Height of the Curve declines Symmetrically: In the normal probability curve the height declines symmetrically in either direction from the maximum point. 6) The Points of Influx occur at point ±1 Standard Deviation (± 1 σ): The normal curve changes its direction from convex to concave at a point recognised as point of influx. If we draw the perpendiculars from these two points of influx of the curve to the horizontal X axis; touch at a distance one standard deviation unit from above and below the mean (the central point). 7) The Total Percentage of Area of the Normal Curve within Two Points of Influxation is fixed: Approximately 68.26% area of the curve lies within the limits of ± 1 standard deviation (± 1 σ) unit from the mean.

8) The Total Area under Normal Curve may be also considered 100 Percent Probability: The total area under the normal curve may be considered to approach 100 percent probability; interpreted in terms of standard deviations.

Meaning

- A correlation coefficient of 1 means that for every positive increase in one variable, there is a positive increase of a fixed proportion in the other. For example, shoe sizes go up in (almost) perfect correlation with foot length.
- A correlation coefficient of -1 means that for every positive increase in one variable, there is a negative decrease of a fixed proportion in the other. For example, the amount of gas in a tank decreases in (almost) perfect correlation with speed.
- Zero means that for every increase, there isn't a positive or negative increase. The two just aren't related.

What is Spearman's Rank correlation coefficient?

Spearman's Rank correlation coefficient is used to identify and test the strength of a relationship between two sets of data. It is often used as a statistical method to aid with either proving or disproving a hypothesis e.g. the depth of a river does not progressively increase the further from the river bank. The formula used to calculate Spearman's Rank is shown below.

$r = 1-6*\Sigma d^2 /n(n^2-1)$

The Product Moment Correlation Coefficient

The product moment correlation coefficient is a measurement of the degree of scatter. It is usually denoted by r and r can be any value between -1 and 1. It is defined as follows:

$$r = \frac{S_{xy}}{\sqrt{(S_x S_y)}}$$

where s_{xy} is the covariance of x and y, $s_x = \frac{1}{n}\Sigma(x - \bar{x})$.

The product moment Correlation

The product moment correlation coefficient (pmcc) can be used to tell us how strong the correlation between two variables is.

A positive value indicates a positive correlation and the higher the value, the stronger the correlation. Similarly, a negative value indicates a negative correlation and the lower the value the stronger the correlation.

If there is a perfect positive correlation (in other words the points all lie on a straight line that goes up from left to right), then r = 1.

If there is a perfect negative correlation, then r = -1.

If there is no correlation, then r = 0. r would also be equal to zero if the variables were related in a non-linear way (they might lie on a quadratic curve rather than a straight line, for example).