SEMESTER III

EDUCATIONAL MEASUREMENT AND EVALUATION

COURSE CODE : MS3TC3

UNIT - I: CONCEPT OF MEASUREMENT AND EVALUATION

Measurement and Evaluation – Concept, Meaning, nature and need. Relationship between measurement and evaluation, Functions of measurement and evaluation.

UNIT - II: TOOLS OF MEASUREMENT AND EVALUATION

Subjective and objective tools - Tests: Essay tests, objective test, scales, questioners, schedules, inventories, observation, interviews, performance tests, oral tests-diagnostic tests and remedial measures

UNIT - III: PSYCHOLOGICAL TESTING

Construction and Standardization of Psychological tests, Aptitude, Attitude, personality tests. Intelligence and its nature - Theories: Spearman, Thorndike, Thurston and Guilford - Types of intelligence test - their functions and uss.

UNIT - IV: STATISTICAL CONCEPTS

Test scores and their transformation: Z and T Scores, percentile-Interpretation of qualitative data Correlation analysis, Item analysis – Basic assumption, Methods

UNIT - V: NEW TRENDS IN EVALUATION

Grading System, Semester system, Continuous Comprehensive Evaluation, Question Bank, uses of computer in evaluation.

UNIT – I

CONCEPT OF MEASUREMENT AND EVALUATION

Concept of Measurement

As a student teacher at the secondary level, you are expected to understand the meaning of the term ‘measurement. “**Measurement is a process of assigning numbers to individuals or their characteristics according to specific rules**” (**Eble and Frisbie**, 1991). This is very common and simple definition of the term ‘measurement’. You can say that measurement is a quantitative description of one’s performance. Gay (1991) further simplified the term as a process of quantifying the degree to which someone or something possessed a given trait, i.e., quality, characteristics, or features.

You can generalize these definitions as measurement provides a quantified description of any trait, characteristics, or ability. For example, Mohan has scored 58 marks in a particular examination and Ahmad has scored 59, their individual scores are the measurement of their performance on a particular test. Similarly you can use the examples of weight and height of learners in your class. If, you measure their height in centimetres and weight in kilograms, you are assigning numerals (i.e. 125 cm. or 45 kg.) using some specific rules (i.e. height in centimetres and weight in kilograms). You can perform the following activity in your class to understand the concept of measurement in a better way:

- Measurement assigns a numeral to quantify certain aspects of human and non-human beings.
- It is numerical description of objects, traits, attributes, characteristics or behaviours.

- Measurement is not an end in itself but definitely a means to evaluate the abilities of a person in education and other fields as well.
- Nature of measurement and assessment vary for various traits or attributes of the students to deal with.

By now you might have understood the concept of measurement. Let us discuss the scales of measurement:

Scales of measurement: Whenever we measure anything, we assign a numerical value. This numerical value is known as scale of measurement. A scale is a system or scheme for assigning values or scores to the characteristics being measured (Sattler, 1992). Like for measuring any aspect of the human being we assign a numeral to quantify it, further we can provide an order to it if we know the similar type of measurement of other members of the group, we can also make groups considering equal interval scores within the group. These are called as measurement in different scales. There are mainly four scales of measurement, namely;

- Nominal
- Ordinal
- Interval and
- Ratio

Nominal scale: In nominal scale, a number is assigned for characterizing the attribute of the person or thing. That caters no order to define the attribute as high-low, more-less, big-small, superior inferior etc. In nominal scale, assigning a number is purely an individual matter. It is nothing to do with the group scores or group measurement. It is therefore, measurement in nominal scale has limited meaning, even if some experts do not consider it as a measurement. As discussed earlier, any student achievement related scores

(scores in subjects) or other measurement like height, weight, etc. are the examples of nominal measurement. Statistics such as frequencies, percentages, mode, and chi-square tests are used in nominal measurement.

Ordinal scale: Ordinal scale is synonymous to ranking or grading. It includes the characteristics of a nominal scale and provides an order to the measurement, like; when we know the achievement scores of students in a group, we can arrange them either in ascending (lowest to highest) or descending (highest to lowest) order. We can also interpret the result like; who stood first, second, 10th in the group, even the last one in the group. In ordinal scale, we can use the statistics such as median (measures of central tendency), quartile and percentile measures, correlation in rank difference method, and non-parametric tests.

Interval scale: Interval scale carries all the characteristics of earlier scales like nominal and ordinal and added with an arbitrary zero point. That is, there is no absolute zero-point or true zero point. In this scale, we can group the scores into equal intervals like, scores within the intervals of: 0-5; 5-10; 10- 15; 15-20; 20-25; 25-30 etc. This is also called as equal interval scale as the size of the classes are equal, i.e. size of the class 0-5 is 5; 5-10 is 5; 10-15 is 5; 15-20 is also 5.As there is no absolute zero point in this measurement and the existing zero value is an arbitrary one, that's why all types of measurement done in education and social sciences are usually done by the interval scale. The statistics like mean, standard deviation, product moment correlation, t-test and f-test can be used in interval scale.

Ratio scale: Ratio scale is called as the highest scale in measurement. It carries all the characteristics of earlier discussed scales with a true or absolute zero point. As there is absolute zero point in this measurement, we can say that zero height means no height. But in the case of interval scale, we cannot say that zero intelligence means no intelligence. All types of measurements conducted in Physical Sciences such as Physics, Mathematics, etc. are done

by ratio scale. Ratio scales are almost non-existence in psychological and educational measurement except in the case of psycho-physical measurement. All mathematical applications – addition, subtraction, multiplication and division can be done in ratio scale. All statistical techniques are permissible with ratio scale.

Meaning of measurement

Suppose there is a basket full of fruits. Let us count the fruits. There are 22 apples, 12 oranges, 16 papayas, 4 pine apples and 8 mangoes. In all there are 62 fruits in the basket. This counting is called **enumeration**.

Now, if someone asks "what is the total weight of the basket?" When the basket is weighed, we find that that is 5.8 kg. This weighting answers the question "How much". So when we are trying to answer the question "How much" i.e. how much height, how much time, how much area, how much volume, how much pressure, how much water, how much electricity etc., we are dealing with measurement. Measurement answers the question "how much". In education, student's progress is measured in terms of marks or grades, teacher's effectiveness as improvement and modification ofthe behaviour of his students etc.

In every measurement, three things are involved, Firstly a set of objects i.e. the thing, object or person to be measured, secondly a set of numbers and thirdly a rule or rules for the assignment of a number to each object.

Measurement may be defined as the assignment of one of a set of numbers to each of a set of persons or objects according to certain established rules.

DEFINITION

- Measurement refers to the process of assigning numerals to events, objects etc.

according to certain rules (**Tyler,** 1963).

- Measurement consists of rules for assigning numbers to objects in such a way as to represent quantities of attributes (**Nunnally,** 1970).
- Measurement of any kind is matter of determining how much or how little, how great or how small, how much more than or how much less than (**Micheels& Karnes**, 1950).
- Measurement may be defined as the assignment of one of a set of numbers to each of a set of persons or objects according to established rules. (**Lindeman and Merenda**, 1979)
- Measurement is the assigning of numbers to things according to rules (**Glass and Stanley**, 1970).

Characteristics of good measurement tool

To measure the psychological traits with validity and reliability, the measuring instrument or tests should be far away from the aspects like personal errors, variable errors, constant errors and interpretative errors. The important characteristics of agood measuring tool are as follows:

(i) Should be valid: Validity of a test refers to its truthfulness. It refers to the extent to which a test measures what it actually wishes to measure. Suppose we want to know whether a Numerical Reasoning Test is valid. If it really measures the reasoning ability, the test can be said to be valid.

(ii) Should be reliable: Reliability means the consistency of a measuring instrument (how accurately it measures). It refers to the faithfulness of the test. To express in a general way, if a measuring instrument measures consistently, it is reliable. For example, a test is administered on English to the students of class VI. In this test, Ram scores 50. After a few days, the same test is administered and Ram scores 50. Here, the

test is reliable becausethere is consistency in the result.

(iii) Should be objective: Objectivity of a test refers to two aspects: (a) item objectivity (i.e., objectivity of the items), and (b) scoring objectivity (i.e., objectivity of scoring). By 'item objectivity' we mean that the items of the test must need a definite single answer. If the answer is scored by different examiners the marks would not vary. Ambiguous questions, lack of proper direction, double barrelled questions, questions with double negatives, essay- type questions must be avoided because they lack objectivity. By 'objectivity of scoring' we mean that by whomsoever scored, the test would fetch the same score. Thus, mostly the objective-type questions should be framed to maintain the objectivity of the test.

(iv) Should be usable and practicable: 'Usability' refers to the practicability of the test. In the teaching–learning situation, by usability we mean the degree to which the test (or the measuring tool) can be successfully used by teachersand school administrators.

(v) Should be comprehensive and precise: The test must be comprehensive and precise. It means that the items must be free from ambiguity. The directionsto test items must be clear and understandable. The directions for administration and for scoring must be clearly stated so that a classroomteacher can easily understand and follow them.

(vi) Should be easy in administering: If the directions for administration are complicated, or if they involve more time and labour, the users may lag behind. For example, Wechsler Adult Intelligence Scale (WAIS) is a good test, but itsadministration is very difficult.

(vii) Should be economical: A measurement tool should be less time consuming. The cost of the test must be reasonable so that the schools/educational institutions can afford to purchase and use it.

(viii) Should be easy in scoring: The scoring procedure of the test should be clear and

simple. The scoring directions and adequate scoring key should be provided to the scorer so that the test is easily scored.

(ix) Should be easily available: Some standardized tests are well-known all over India, but they are not easily available. Such tests have less usability. It is desirable that in order to be usable, the test must be readily and easilyavailable.

(x) Should have good and attractive get up/appearance: The quality of papers used, typography and printing, letter size, spacing, pictures and diagrams presented, its binding, space for pupil's responses etc., need to be of verygood quality and attractive.

Nature of Educational Measurement and Evaluation

Evaluation is an act or process that assigns 'value' to a measure. When we are evaluating, we are making a judgment as to the suitability, desirability or value of a thing. In the teaching–learning situation, evaluation is a continuous process and is concerned with more than just the formal academic achievement of students. Evaluation refers to the assessment of a student's progress towards stated objectives, the efficiency of the teaching and the effectiveness of the curriculum. Evaluation is a broad concept dealing not just with the classroom examination system; but also evaluating the cognitive, affective and psychomotor domain of students. The success and failure of teaching depends upon teaching strategies, tactics and aids. Thus, the evaluation approach improves the instructional procedure. Glaser's basic model of teaching refers to this step as a 'feedback function'.

J.M. Bradfield defines evaluation as 'the assignment of symbols to phenomenon in order to characterize the worth or value of the phenomenon usually with reference to some social, cultural and scientific standards'. Wright Stone stated, 'evaluation is a relatively new technical term introduced to designate a more comprehensive concept of measurement than is implied in conventional test and examination'. Hanna defined evaluation as 'the process of gathering and

interpretingevidence on change in the behaviour of all students as they progress through school'.

Evaluation takes place with the help of tests and measurements. In a classroom situation, teachers first use classroom tests to evaluate students according to their different traits. After getting the answer papers, teachers provide some numerals to the answer papers, this step is known as measurement. So measurement deals with only the quantitative description. After the measurement step, the teachers arrange the students as first, second, third etc., according to their achievements. This step is evaluation. So evaluation is a philosophical and subjective concept. It includes bothquantitative and qualitative descriptions, and value judgment.

Therefore, Evaluation = Quantitative Description (Measurement) and/or Qualitative Description (Non-measurement) + Value Judgments.

Functions of Measurement:

In psychology and education, the results of measurement serve various functions such as:

1. Classification:

Measurement helps in classifying the people in placing them in different categories. In school, army or industry, classification is sometimes very essential. In school, students have to be classified according to their achievement or ability.

In the army, officers and soldiers may be classified in accordance with their battalions or work-assignment or station. In industry, workers may be classified according to various levels or positions of work. Such a classification is called 'placement classification'. This implies gradation of workers. Some are placed higher and others lower.

2. Selection:

Selection is done in industrial establishments, in the army and in the Civil Services. Measurement of one kind or the other is an essential tool for the same.

Various measurement techniques are employed such as aptitude and ability tests, interviews, projective techniques, situational tests, achievement tests etc. Whenever there is selection, we have to select a few and reject many. Measurement tools are, therefore, applied cautiously.

3. Comparison:

Individual difference in certain traits among various persons is universal phenomenon. By the use of statistics, measurement helps to compare one trait of an individual or a group with that of others. Comparison is essential to determine the reasons for individual difference.

4. Prediction:

Many decisions of our daily life involve prediction. We may be interested in the problem whether a test in visual recognition may predict success in perception in an aeroplane. A physician may be interested in the predictive value of a drug. Scores obtained in present test may indicate success in future. So prediction involves foretelling.

If a company has to employ some salesmen, it may apply a test and make selection on the basis of scores obtained in that test. Results of intelligence, aptitude and other tests may be evaluated in the context of their predictive value.

5. Diagnosis:

Diagnosis involves location of strengths and weaknesses. Educational diagnosis implies the use of various technical procedures. Weakness in the individual student, if identified through diagnostic tests, remedial steps can be taken up.

Thus Diagnosis is helpful not only in locating handicaps, but anticipating causes and remedies. Diagnosis has suggestive value. For example, if diagnosis tells us that students in the fourth grade are weak in arithmetic computation, it may suggest that this is due to wrong methods of instruction.

6. Improving Instructional Practices:

To make instructional programmes attractive many instructional practices or methods are adopted by the teacher. He is the sole executor and students are the sole respondents.

Therefore, measurement tools assist both to rectify and to improve their methods of teaching and learning. What method will be suitable in a particular situation to a particular group of students, a teacher knows better. Thus a teacher improvises his quality of teaching as well as his instructional practices.

7. Developing Curriculum:

Curriculum is constructed on the assumptions of fulfilling the concept of three R's—reading, writing and arithmetic. It should have enough scope to learn and to know so that learning may help the learners in dealing with their odds of life.

The measurement through its devices enables to know the authenticity, objectivity and usability of curriculum. If there is drawback, it can be improved. Actually, effectiveness of courses and programmes outlined in the curriculum are determined by measurement. It selects, clarifies and appraises the objectives decided in learning programmes through curriculum.

8. Counseling and Guidance:

Through measurement, a counselor can know the potentialities in his students and then he can suggest them to adopt a job of their own choice. In these days, counseling and guidance play an important role in an individual's life.

Proper counseling and guidance can only put a man in his best capacity to utilize full opportunity to grow. This can be done on measuring aptitude, interest and intelligence of a student by his teacher.

9. Helping Administration:

Various ways of measurement enable the authorities to do administration effectively and sincerely. The schools can serve the purposes of the community through maintaining a good relationship.

The closeness can be measured and maintained by organising functions, sharing the social values by the staff members, participating in the functions and ceremonies of the community and by forming organisation like, "Parent Teacher Associations".

10. Research:

Measurement helps in research. In psychological and educational research, we generally take two groups—control group and experimental group—and compare the performance of the two groups.

Research deals with the investigation of some problem. While doing so, we control all other factors and study one particular factor. Before starting such a study, we take a measuring programme to determine the points of similarities and differences between the two groups and within the group.

Methods of Measurement

For both physical and mental measurement, some tools and methods are necessary. The variation of method may be due to the nature of variable and purpose of measurement.

The methods of measurement are as follows:

(i) **Tests:** A test consists of a set of questions to be answered or tasks to be performed. Tests are used to assess the ability or trait in question. Psychological and educational tests are standardized procedure to measure quantitatively or qualitatively one or more than one aspect or trait by means of a sample of verbal or non-verbal behaviours. Items of a test are placed in increasing order of difficulty and its procedure of administration is standardized to ensure maximum objectivity. The psychological tests are used to know the ability of the students, to diagnose the weakness, to predict the future progress, and to provide educational and vocational guidance. The different types of tests are: achievement tests, intelligence tests, attitudetests, aptitude tests, personality tests, creativity tests etc.

(ii) **Inventories:** Different inventories are used for different traits. Interest inventories are used to measure interest; personality inventories are used to measure certain traits on personality etc.

(iii) **Observation:** There are certain traits like honesty, punctuality, persistence, truthfulness etc., which can hardly be measured objectively via tests. So here, observation is an important technique of measurement. The observation may be participant observation or non-participant observation for accurate and scientific observation. One may use observation schedule and otherinstruments.

(iv) **Interview:** Interview is a face-to-face interaction between one interviewee and one interviewer or more than one interviewers. There are certain things which an individual does not want to express and they can be only assessed through interviews. The interview

schedules may be used and the interviewer through a better personal support, and in congenial atmosphere, can succeed to bring out the inner feelings of the interviewee through carefully plannedinterviews.

Checklist: A checklist consists of a series of items which needs response from the respondent. The presence or absence of an item may be indicated. by 'Yes' or 'No' (by a '□' or 'X' against the items). Checklists are popularly employed for appraisal of studies, school buildings, textbooks, outcomes,instructional procedures etc.

(v) Rating scales: Psychological traits are relative concepts. So it is very difficult to make watertight compartments between them. Sometimes, the degree of a trait is necessary on the part of the rater. Rating scale is used to evaluate the personal and social conduct of the learner. We take the opinion of teachers or parents or friends or judges on a particular quality or trait of a pupil along a scale. The rating scale may be of 5 points, 7 points, 9 points or 11 points. For example, to assess particular trait, we can use a 5 point scale as: very good, good, average, below average, and poor. The trait in question is marked by the judges in any one of the five categories. Rating scales can be used to evaluate: personality traits, tests, school courses, school practices, and otherschool programmes.

(vi) Attitude scales: Attitude refers to the bent of mind or feelings of an individual towards an object, an idea, an institution, a belief, a subject or even a person. Attitude scales are used to measure this trait objectively with accuracy.

(vii) Projective techniques: Projective techniques are very ambiguous and subjective in nature. Through projective techniques, the sub-conscious and pre-conscious mind of an individual is reflected. For example, with the help of Thematic Apperception Test (TAT), we measure the personality of individuals.

Anecdotal record card and cumulative record cards are also used for educational measurement and evaluation.

Need for Measurement and Evaluation in Education

The process of evaluation consists of the following three aspects:

(i) Quantitative measurements

(ii) Qualitative assessment

(iii) Teachers' opinion

(i) Quantitative measurement

We often do the quantitative measurement of the performance of students through tests. It usually is the case that more time is spent on teaching a unit than testing it. The achievement of the students usually depends on the nature of tests such as teacher-made tests, standardized tests, etc. But the achievement also depends on the physical and mental health of the student, the level of their preparation and motivation during the tests. The marks obtained by the students in different subjects in terminal examinations and expressed in percentage of marks obtained as the aggregate is an example of applying quantitative measurement. This kind of measurement provides information about the overall achievement level of pupils, but does not help us in understanding the individual performance of a child and difficulties faced by him or her. Therefore, quantitative measurement has its limitations and thus it needs to be supplemented by qualitative assessment.

(ii) Qualitative assessment

The qualitative assessment description implies observation by teachers and records maintained by them pertaining to various facets of student personality and performance in the school. The systematic record of statement by teachers about the special achievement or exceptional ability in non-scholastic area or exceptional behaviour of students is one of the illustrations of the qualitative assessment description.

(iii) Teachers' opinion

Teachers form opinions about the students on the basis of conclusions drawn by them from quantitative measures and a qualitative description of the behaviour of students. The opinion of a teacher arrived at on the basis of qualitative and quantitative measurement gives a comprehensive picture of the students' progress. Though it is an evaluation in an informal setting, yet it conveys information about the students' performance on some vital issues.

Evaluation

Concept and Definitions of Evaluation

Evaluation is the application of a standard and a decision-making system to assessment data to produce judgments about the amount and adequacy of the learning that hastaken place."

"The process of determining to what extent the educational objectives are actually being realized" (Tyler)

"Evaluation is the process of determining merit, worth, or significance; an evaluation is a product of that process" (Scriven)

Evaluation is a systematic determination of a subject's merit, worth and significance, using criteria governed by a set of standards. It can assist an organization, program, design, project or any other intervention or initiative to assess any aim, realisable concept/proposal, or any alternative, to help in decision-making; or to ascertain the degree of achievement or value in regard to the aim and objectives and results of any such action that has been completed. The primary purpose of evaluation, in addition to gaining insight into prior or existing initiatives, is to enable reflection and assist in the identification of future change.

Evaluation is the structured interpretation and giving of meaning to predicted or actual impacts of proposals or results. It looks at original objectives, and at what is either predicted or what was accomplished and how it was accomplished. So evaluation can be formative, that is taking place during the development of a concept or proposal, project or organization, with the intention of improving the value or effectiveness of the proposal, project, or organisation. It can also be summative, drawing lessons from a completed action or project or an organisation at a later point in time or circumstance.

Nature of Evaluation

- Evaluation involves observing, documenting and measuring. It compares what happened with what is expected to happen.
- Evaluating the value or worth of something, is an activity that involves making judgements. "Value" is not absolute – people have different views about what is of value.
- Evaluation is the systematic assessment of the design, implementation or results of an initiative for the purposes of learning or decision-making.
- An evaluation should be as systematic and impartial as possible
- An evaluation is methodical, providing information that is credible, reliable, and useful to enable the incorporation of lessons learned into decision-making process of users
- Evaluation is based on empirical evidence and typically on social research methods.
- Evaluation focuses on grades and might reflect classroom components other than course content and mastery level. An evaluation can be used as a final review to gauge the quality of instruction. It's product-oriented. This means that the main question is: "What's been learned?" In short, evaluation is judgmental.
- Evaluation is purposive. Evaluation can be conducted for the purposes of decision

making, judgements, conclusion, findings, new knowledge, organizational development and capacity building in response to the needs of identified stakeholders leading to improvement, decisions about future programming, and/or accountability ultimately informing social action ameliorating social problems and contributing to organizational or social value

Characteristics of Evaluation

Evaluation has a specific set of characteristics which makes it a unit process, different from assessment and measurement. The following are the different characteristics of evaluation.

- Evaluation is a continuous process
- It is a comprehensive concept
- It includes academic and non-academic subjects
- It is not confined to classrooms
- It is of different types
- It contributes for the improvement of the product
- It is an integral part of teaching
- It is a systematic process
- It also includes evaluation of teachers and school
- It is a diagnostic appraisal
- It includes both qualitative and quantitative description.
- Evaluation includes two important elements. They are measurement and appraisal.

Importance of Evaluation

Evaluation proves to be fruitful to different members in different ways in educational settings It can be summarised as below:

- Evaluation is important to the classroom teachers, supervisors, and administrators, indirecting as well as guiding teaching and learning.
- Evaluation aids in devising more effective instructional materials and procedures ofinstruction.
- Evaluation helps to measure the validity and reliability of instruction.
- Evaluation stimulates students to study
- Evaluation helps teachers to discover the needs of the pupils
- Evaluation can be used to enforce external standards upon the individual class orschool
- Evaluation, helps to provide objective evidence for effective cooperation betweenparents and teachers.
- Evaluation helps parents to understand pupil-growth, interests and potentials.
- Evaluation is helpful in securing support for the school from government.

Evaluation is considered important as it serves many more purposes. They are as follows:

- It provides periodic tests, which give direction for improvement of work
- It serves as check
- It helps to determine the degree attained
- It enhances knowledge
- It helps to improve achievement
- It gives satisfaction to school personnel.
- It contributes for the improvement of instruction
- It provides the basis for guidance
- To improve student personality
- For clarification of objectives of teaching and related tasks
- To make needed changes in the curriculum

- To attain success in teaching
- To promote better learning

Principles of Evaluation:

Evaluation is a systematic process of determining to what extent instructional objectives has been achieved. Therefore evaluation process must be carried out with effective techniques.

The following principles will help to make the evaluation process an effective one:

1. It must be clearly stated what is to be evaluated:

A teacher must be clear about the purpose of evaluation. He must formulate the instructional objectives and define them clearly in terms of student's observable behaviour. Before selecting the achievement measures the intended learning out comes must be specified clearly.

2. A variety of evaluation techniques should be used for a comprehensive evaluation:

It is not possible to evaluate all the aspect of achievement with the help of a single technique. For the better evaluation the techniques like objective tests, essay tests, observational techniques etc. should be used. So that a complete' picture of the pupil achievement and development can be assessed.

3. An evaluator should know the limitations of different evaluation techniques:

Evaluation can be done with the help of simple observation or highly developed standardized tests. But whatever the instrument or technique may be it has its own limitation. There may be measurement errors. Sampling error is a common factor in educational and psychological measurements. An achievement test may not include the whole course content. Error in measurement can also be found due to students guessing on objective tests. Error is also found due to incorrect interpretation of test scores.

4. The technique of evaluation must be appropriate for the characteristics or performance to be measured:

Every evaluation technique is appropriate for some uses and inappropriate for another. Therefore while selecting an evaluation technique one must be well aware of the strength and limitations of the techniques.

5. Evaluation is a means to an end but not an end in itself:

The evaluation technique is used to take decisions about the learner. It is not merely gathering data about the learner. Because blind collection of data is wastage of both time and effort. But the evaluation is meant for some useful purpose.

Need of Evaluation

The needs of the Evaluation are as follows:-

(1) For test the educational importance of the activities of educational administrators, other personnel and guardians from time to time, and to suggest for improvement.

(2) For analyse the educational objectives, to test their utility, and to suggest for timely change.

(3) For find out the effect of the curriculum at different levels in the achievement of educational objectives, to give suggestions for improvement and to enlighten for research.

(4) For study the effect of teaching methods being used from time to time, to find out useful/ useless methods, to suggest for improvement and to enlighten the field for research.

(5) For find out the utility of textbooks in the achievement of educational objectives, to give suggestions for improvement and to guide for research.

(6) For study the effect of the use of various means in teaching as to their place and kind of use, and to suggest measures for improvement.

Functions of Evaluation:

The main aim of teaching learning process is to enable the pupil to achieve intended learning outcomes. In this process the learning objectives are fixed then after the instruction learning progress is periodically evaluated by tests and other evaluation devices.

The function of evaluation process can be summarized as following:

1. Evaluation helps in preparing instructional objectives:

Learning outcomes expected from class-room discussion can be fixed by using evaluation results.

What type of knowledge and understanding the student should develop?

What skill they should display?

What interest and attitude they should develop?

Can only be possible when we shall identify the instructional objectives and state them clearly in terms of intended learning outcomes. Only a good evaluation process helps us to fix up a set of perfect instructional objectives.

2. Evaluation process helps in assessing the learner's needs:

In the teaching learning process it is very much necessary to know the needs of the learners. The instructor must know the knowledge and skills to be mastered by the students. Evaluation helps to know whether the students possess required knowledge and skills to proceed with the instruction.

3. Evaluation help in providing feed back to the students:

An evaluation process helps the teacher to know the learning difficulties of the students. It helps to bring about an improvement in different school practices. It also ensures an appropriate follow-up service.

4. Evaluation helps in preparing programmed materials:

Programmed instruction is a continuous series of learning sequences. First the instructional material is presented in a limited amount then a test is given to response the instructional material. Next feedback is provided on the basis of correctness of response made. So that without an effective evaluation process the programmed learning is not possible.

5. Evaluation helps in curriculum development:

Curriculum development is an important aspect of the instructional process. Evaluation data enable the curriculum development, to determine the effectiveness of new procedures, identify areas where revision is needed. Evaluation also helps to determine the degree to what extent an existing curriculum is effective. Thus evaluation data are helpful in constructing the new curriculum and evaluating the existing curriculum.

6. Evaluation helps in reporting pupil's progress to parents:

A systematic evaluation procedure provides an objective and comprehensive picture of each pupil's progress. This comprehensive nature of the evaluation process helps the teacher to report on the total development of the pupil to the parents. This type of objective information about the pupil provides the foundation for the most effective co-operation between the parents and teachers.

7. Evaluation data are very much useful in guidance and counselling:

Evaluation procedures are very much necessary for educational, vocational and personal guidance. In order to assist the pupils to solve their problems in the educational, vocational

and personal fields the counsellor must have an objective knowledge of the pupils abilities, interests, attitudes and other personal characteristics. An effective evaluation procedure helps in getting a comprehensive picture of the pupil which leads to effective guidance and of counselling.

8. Evaluation helps in effective school administration:

Evaluation data helps the administrators to judge the extent to which the objectives of the school are being achieved, to find out strengths and weaknesses of the curriculum and arranging special school programmes. It also helps in decisions concerning admission, grouping and promotion of the students.

9. Evaluation data are helpful in school research:

In order to make the school programme more effective, researches are necessary. Evaluation data help in research areas like comparative study of different curricula, effectiveness of different methods, effectiveness of different organisational plans, etc.

Difference between Measurement and Evaluation

Measurement	Evaluation
1. In the measurement process, a trait of an object, person or activity is expressed in definite words, symbols or units.	1. In the evaluation process, the results obtained from the measurement of an object, person or activities are analysed.
2. Measurement is the first step of evaluation.	2. Under evaluation, the results are also analysed after having

	obtained results from measurement.
3. There are four steps in the measurement process.	3. There are six steps in the evaluation process.
4. Evidences are collected in the measurement process.	4. The evidences are analysed under the evaluation process.
5. No clear concept about an object, person or activity can be formed on the basis of the results of measurement.	5. Clear concept about an object, person or activity can be formed on the basis of the results obtained from evaluation.
6. Proper comparision between two or more objects, persons or activities cannot be done on the basis of results obtained from measurement.	6. Two or more objects, persons or activities can be properly compared on the basis of results obtained from evaluation.
7. They cannot be properly classified on the basis of results obtained from	7. They can be properly classified on the basis of results obtained from evaluation.

measurement. 8. Proper guidance cannot be given to the persons on the basis of results of measurement. 9. No prediction can be made about a person on the basis of results obtained from measurement	8. The persons can be given proper guidance on the basis of results obtained from evaluation. 9. Prediction about a person can be made on the basis of results obtained from evaluation.

UNIT II

TOOLS OF MEASUREMENT AND EVALUATION

Subjective and Objective Evaluation

We know that in evaluation, the analysis of results obtained from measurement is carried out on the basis of certain social, cultural or scientific standards (Norms). The first thing to keep in mind in this context is that the social and cultural standards (Norms) are not fully clear and definite in themselves; and secondly, the persons take them according to their own views and use them accordingly. It is clear then, that the evaluation based on social and cultural standards (Norms) is subjective. On the other hand, the scientific standards (Norms) are always clear and definite by themselves and as such the evaluation done on their basis is objective.

Formative Evaluation

Formative evaluation is such evaluation which is conducted before giving final Shape to an educational policy or programme, curriculum, teaching method, teaching aid or evaluation method. For it, the evaluator first of all prepares the preliminary draft of the proposed educational policy, planning or programme, curriculum, teaching method or evaluation method; then he analyses each of its steps and receives approval of the specialists. This approval is received with the help of interview, questionnaire or rating scale.

Summative Evaluation

Summative evaluation is such evaluation which is conducted in order to test the utility of an already existing educational policy, planning or programme, curriculum, teaching method, teaching aid or evaluation method. For it, the evaluator constructs the most suitable measurement tool or method based on interview, questionnaire or rating scale, for evaluation of the educational policy, planning or programme, curriculum, teaching method, teaching aid

or evaluation method. After this, he tests its utility on the basis of related standards (Norms) and statistical calculations. Finally, he decides whether such educational policy, planning or programme, curriculum, teaching method, teaching aid or evaluation method should continue or not, and if it is to continue, what should be its form

Test

Meaning of Test

A set of questions or exercises evaluating skills or knowledge. One valid approach to assessment is to observe everything that is taught. In most situations this is not possible, because there is so much information to be recorded. Instead, one has to select a valid sample from the achievements of interest. Since school learning programmes are expected to provide students with the capability to complete various tasks successfully, one way of assessing each student's learning is to give a number of these tasks to be done under specified conditions. Conventional pencil-and-paper test items (which may be posed as questions) are examples of these specially selected tasks. However other tasks may be necessary as well to give a comprehensive, valid and meaningful picture of the learning. For example, in the learning of science subjects practical skills are generally considered to be important so the assessment of science subjects should therefore include some practical tasks. Similarly, the student learning music may be required to give a musical performance to demonstrate what has been learned. Test items or tasks are samples of intended achievement, and a test is a collection of such assessment tasks or items.

Single, discrete items may not be reliable (or consistent) indicators of achievement. However, when a number of similar items or tasks are combined as a test, we can look at patterns of success on the test. Such patterns tend to be more dependable indicators because they are based on multiple sources of evidence (the various separate assessment tasks).

Definition of Test

Test is a set of questions or exercises evaluating skills or knowledge of student individually

Factors affect the Test

The factors which affect the construction of a test are called predisposing factors. Some of the predisposing factors of test are as follows:

i)Number of Papers

The paper setter has first to decide whether only one or two question papers in a subject are to be set. If two, automatically the examination syllabus is halved and the sampling of content would be more adequate. More number of questions can be set; more possibility of testing different types of abilities will be there. Therefore decision about papers has to be taken in advance.

ii) Total Time

Time to different papers in different subjects is usually stipulated. It may vary from 2 to 3 hr in various subjects in different boards at secondary or senior secondary stage. For class tests it may vary from 15 min. to 1 hr. A term test could be of 1 hr to 3 hr, whereas an annual test may be of 3 hr. whatever may be the total time it must be decided in advance for planning the question paper. It is stipulated by examining agency for external examinations and by teachers for school examinations. Giving more time than required for a fixed set of questions may lead to collateral effects like guessing, copying, redundant and irrelevant answers besides affecting the measurement of speed in speeded tests.

(iii) Total Marks

A question paper carrying 100 or 50 marks has different implications. In the former it is possible to include more open-ended essay-type questions (time remaining the same) by increasing the number of short-answer and objective type questions. With less marks, as in the

latter, possibility of including long-answer questions decreases and less number of questions would be included, thereby lowering the reliability of the test. Inclusion of more short-answer and objective-type questions would become necessary for better coverage of syllabus. Total marks therefore have its impact on the quality of question papers.

(iv) Scheme of Sections

Depending upon the types of questions to be included in a question paper, we can have one or two sections, which could be content-wise such as Botany and Zoology or question-wise like objective and open-ended questions. For external examinations separate time limit of 30 min. or more can be imposed for objective questions. Whatever be the decision it must be taken beforehand by the board for external examinations and by teachers for their school examinations.

(v) Scheme of Options

If no option is provided and all questions are compulsory, comparability of students' marks improves. If free option is provided, comparability of students' scores is not possible because different students attempt different combinations of questions. Moreover, it leads to selective teaching and selective learning. Provision of internal balanced options of 'This OR This' type ensures more coverage of syllabus, besides opening opportunities to students to demonstrate their best. Therefore the scheme of options must be carefully decided while designing the question papers.

(vi) Medium of Testing

A decision about medium of testing and number of versions to be developed has also to be taken before commencing to write the questions. Framing of questions directly in a language and getting them translated from questions written in another language is quite different. Validity of many questions is lost during their translation from English to other languages. Unless the framer of the question is cognizant of this fact during construction of

questions some good questions that cannot be properly translated into another language have to be sacrificed at the altar of validity of the question.

Tools of evaluation

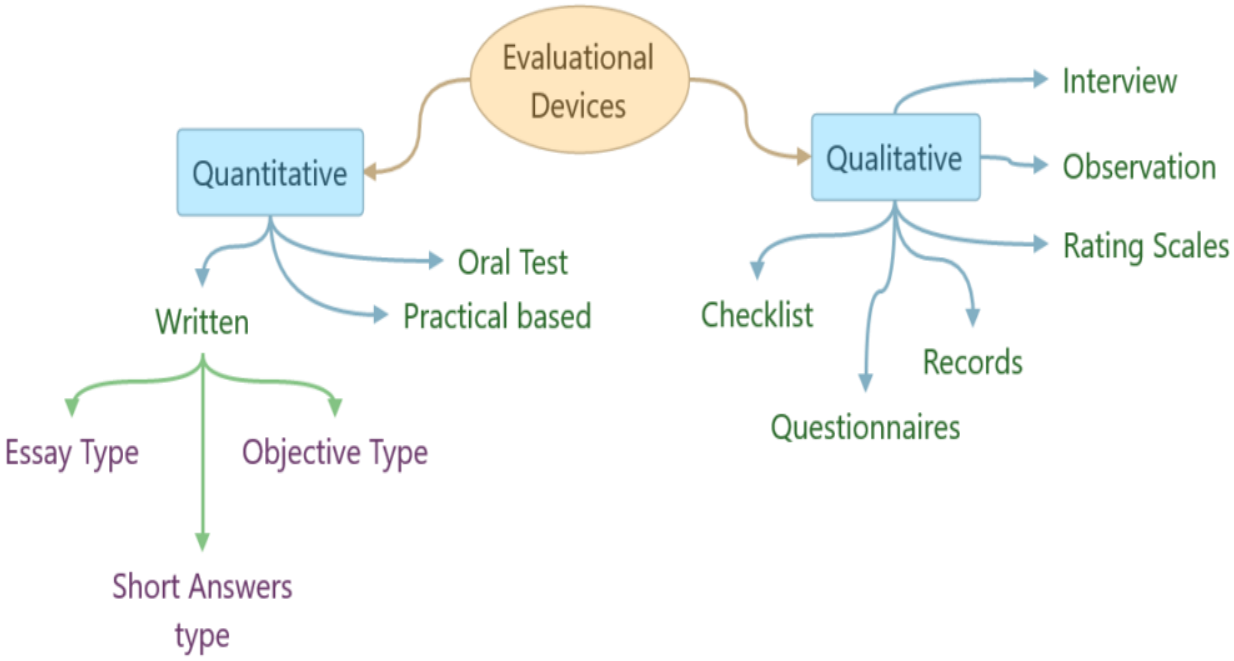
Educational evaluation can be considered as the process of determining the following aspects of the teaching and learning process. A teacher can employ any one of the available tools of education at his/her disposal to know about,

1. The extent to which educational objectives are being achieved.
2. The effectiveness of the teaching-learning experiences provided in the classroom situation.

Various evaluation techniques in Education

Different evaluation techniques are used by teachers to know all the changes that take place in the child as a result of teaching. There are various tools and techniques of evaluation. Generally, they are classified into two categories

1. Quantitative technique
2. Qualitative technique



Oral Tests

- The oral exam (also oral test or viva voce) is a practice in many schools and disciplines, where an examiner poses questions to the student in spoken form. The student has to answer the question in such a way as to demonstrate sufficient knowledge of the subject in order to pass the exam.
- Many programs require students to finish the program by taking an oral exam or a combination of oral and exams in order to show how well a student has comprehended the material studied in the program.
- Schools use oral exams just to test knowledge, but the ability to respond on the spot. Sometimes the oral exam is offered in schools as an alternative to a written exam for students with a learning disability, like dysgraphia, developmental coordination disorder, or non-verbal learning disorder. Often the parents of the students have to request that the oral exam be given to their child in lieu of the written exam.

Written evaluation techniques/devices

- Most commonly used tool of evaluation.
- Various types of written tests are conducted throughout the academic session.
- Written tests are tests administered on paper or on a computer.
- Here student appearing for the written test have to provide answers by writing or typing in the space given or on a separate sheet.

So, written tests are tests that are administered on paper or on a computer. A test taker who takes a written test could respond to specific items by writing or typing within a given sp[ace of the test or on a separate form or document. In some test; where knowledge of many constants or technical terms is required to effectively answer questions, like Chemistry or Biology – the test developer may allow every test taker to bring with them a cheat sheet. A test developer's choice of which style or format to use when developing for testing. Be that as it may, certain test styles and format have become more widely used than others. Below is a list of those formats of test items that are widely used by educators and test developers to construct paper. As a result, these tests may consist of only one type of test item format (e.g., multiple choice test, essay test) or may have a combination of different test item formats (e.g., a test that has multiple choice and essay items).

Written examination includes the following type of tests.

1. Essay type tests
2. Objective type tests
3. Short answer type tests.

Content of Written Test

This includes the following types of test items :

- Short-answer
- Long-answer
- Multiple-choice and matching
- True/false (alternative choice)

This is followed by a brief discussion on the use of graphics in test questions. It is important to use the appropriate types of tests in your evaluations.

(1) Essay type tests:-

In essay type test students give responses to many questions of the curriculum in some fixed devotion in the form of essay. This type of tests is commonly employed to check the power of memorization, expression, recognition, etc. So these tests are used to measure the levels of interpretation and evaluation of the student. Weidman (1941) categorised such tests into 11 types of questions ranging from the simple to the most complex using directional words like (i) what, when, who, which, where; (ii) list; (iii) outline; (iv) describe; (v) contrast; (vi) compare; (vii) explain; (viii) discuss; (ix) develop; (x) summarise and (xi) evaluate. Likewise, Monroe and Carter divided ETQ into 21 categories, ranging from selective recall to ...cause and effect...application of principles...criticism... new procedure to inferential thinking.

Principles of Constructing Essay Type Test

General Hints

(i) Use ETQ (Essay Type Question) where you must and avoid where you can.

(ii) Try to increase the number of questions by including more questions of restricted-response variety in preference to extended-response variety.

(iii) Avoid giving choice, especially free options (6 out of 9 type), in ETQ to discourage teaching and selective learning besides ensuring better comparison of students.

(iv) If only ETQ are included in a question paper, ensure wider range of difficulty level to cater to poor, average and bright students using lower-order, middle-order and higher-order questions.

(v) Frame each question keeping in view the stipulated time requirement.

(vi) Write explicit and clear instructions for examinees to enable them to attempt questions on similar lines for similar goal.

Specific Hints

(i) Pinpoint the specific assessment objectives, which becomes the basis of your question.

(ii) Select content clusters from one or more content areas of syllabus which are relevant to assessment objectives.

(iii) Use familiar and appropriate directional words that evoke the desired responses and demand exercise of intended mental process (list, describe, compare, discuss, justify, evaluate etc.). Avoid directional words like 'what do you know of, 'write short notes on', 'give an account of' etc.

(iv) Structure the question to pinpoint the area of response and delimit the scope of expected response, by proper wording of the question.

(v) Avoid semantic difficulties by using simple, precise and unambiguous language.

(vi) Set task in the question that require students to demonstrate command of the essential knowledge, not the factual information.

(vii) Indicate clearly part-wise marks for each question that has more than one parts, e.g. Define germination. What are the conditions necessary for germination ? Illustrate with the help of an experiment. (2 + 3 + 5)

(viii) Write model answer to test the efficacy of your questions. It helps improve the question if needed.

(ix) Work out the marking scheme, indicating major value points and their corresponding marks and the mode of deduction of marks, if any.

All the hints given above are not only useful in improving the quality of ETQ but also for ensuring greater objectivity in grading or scoring of questions. If care is taken to observe all the guidelines given above, most of the defects associated with ETQ can be done away with to make these questions a more valid measure of determining pupils' learning outcomes.

Merits of essay type tests:-

1. Essay type tests are easy to construct. Here question papers are short, can be prepared in a small time frame and at a low cost.
2. These tests are suitable for all subjects.
3. Through essay type tests students can express their original ideas.
4. Essay types tests are very simple information, conduct and operation.
5. The mental abilities like thinking, reasoning, expression, and criticism, etc. can only be measured by essay type questions.
6. Through essay type tests, the writing style of the students can be developed.
7. Here students are free to express their ideas in a logical way.
8. Essay type tests inspire students for extensive study.

Demerits of essay type tests:-

1. The essay type tests look at clearly defined objectives.
2. It promotes cramming and rote memorization.
3. The scores of there tests show variations and the results are also not consistent.
4. The personal views, opinions, ideas, etc. of examiner and examinee affect the response of a particular answer as well as evaluation.
5. Sometimes these tests become time-consuming.
6. Proper evaluation is not possible by these tests.

7. Candidate with good handwriting sometimes gets more marks than the one who gives an exact and accurate answer in bad handwriting.
8. Evaluation of these tests is a hard, lengthy and difficult task in which let of time is wasted.

(2) Objective type tests:-

- Objective type questions are answered by just writing one or two words, or numerals, fill in the blanks, choosing one cut of multiple responses given, etc.
- In objective tests, the achievement of subjective knowledge of pupils, their aptitude, attitudes, interests, inelegance, etc are measured.
- These tests consist of about 150 to 200 short and pointed questions based on the entire curriculum in a very short duration.
- These tests have objectivity and their measurements will not effect scorability of the pupils.
- Since the answers to these types of questions even one and the same, therefore it is quite reliable, objective, valid and can not vary.

Type of Objective tests:-

(1) True/False (alternative choice) questions

The choices offered in these types of questions may be True/False, Yes/no, Fact/Opinion, High/Low, Agree/Disagree, and so on. There must be only one correct response to the question.

Use True/False questions to test a student's ability to :

- Recognize a correct statement of fact or opinion
- Identify relationships (including cause)

- Identify attitudes, values, and beliefs
- Identify a new situation where known principles apply.

These type of tests are presented in the form of a simple declarative statement, to which the pupil responds indicating whether the statement is true or false.

Example:-

Direction – write 'T' if the statement is true and 'P' if the statement is false.

(a)(a+b)^2=a2+b2+2ab._____

(b)(a-b)^2=a2-b2._____

(c) The sum of the angles of a triangle is equal to two right angles._____

(2) Completion test:-

A fill-in-the-blank item provides a test taker with identifying characteristics and requires the test taker to recall the correct term. There are two types of fill-in-the-blank test. The easier version provides a word bank of possible words that will fill in the blanks. For some exams, all words in the word bank are used exactly once. If a teacher wanted to create a test of medium difficulty, they would provide a test with a word bank, but some words may be used more than once and others not at all. The hardest variety of such a test is a fill-in-the-blank test in which no word bank is provided at all. This generally requires a higher level of understanding and memory than a multiple choice test. Because of this, fill-in-the-blank tests are often feared by students. In these tests, the pupils are required to complete the given incomplete statement.

Direction: Fill in the blanks in the following sentences.

Examples:

(i)4:5::8:_____

(ii)$(a+b)^2+a^2+$_____+2ab

(iii) sum of the exterior angles of a polygon is_____.

(3) Multiple choice tests:-

A multiple choice is a form of assessment in which respondents are asked to select the best possible answer (or answers) out of the choices from a list. If guessing an answer, there's usually a 25 percent chance of getting it correct on a 4 answer choice question. Finding the right answer from multiple choices can be automated using multiple choice question answering systems. The multiple choice format is most frequently used in educational testing, in market research, and in elections, when a person chooses between multiple candidates, parties, or policies.

Although E.L. Thorndike developed an early multiple choice test, Fredrick J. Kelly was the first to use such items as part of a large scale assessment. While Director of the Training School at Kansas State Normal School (now Emporia State University) in 1915, he developed and administered the Kansas Silent Reading Test. Soon after, Kelly became the third Dean of the College of Education at the University of Kansas. These tests contain a number of items, each of which has three or four responses. One of the responses in each term is correct. The students are required to tick the correct suspense.

Example:

Direction: Choose the correct answer from four given responses.

The length of the hall is 16m, breadth is 12m and height is 10m. The area of one of its long wall(m^2)is

(a)16×12

(b)16×10

(c)12×10

(d)16x12x10

(4)Matchingtests:-

In these tests, the pupils match test items of one side to the test items of the other side. On the other side, answers are not written exacted in front of the questions but they are written un-sequentially. Students have to make a pair by mentioning the right sequence of answers in front of the questions.

Advantages of Objective Test

The objective tests have all the qualities of a good test, though all the abilities of the students cannot be measured by them.

1. **Validity :** Objective tests are generally made to measure the knowledge of the students. First, the questions asked pertain to the field, the knowledge of which has to be measured. Second, the number of questions is generally large which is spread out on the entire range of knowledge. Third, the evaluation of these questions is objective. So these tests are valid.

2. Reliability : These tests are reliable. All the questions contained in them are clear and bear only one meaning, and their answers too are definite. These are very comprehensive, and the measurement done by them is also objective. It is natural for them to be reliable.

3. Objectivity : As has been clarified above, the questions asked in these tests are clear and have a single meaning, their answers are also definite and the evaluator has not any liberty to mark them. Whoever the evaluator may be, the result of such tests is similar. It is evident that these, tests are objective and therefore reliable.

4. Comprehensiveness : As has been stated in the beginning, small questions are asked in these tests and the answers to these questions are given in one mark, number or word. From the viewpoint of availability of time, the time to be taken on to write one essay type answer or 46 short-answer-type answers can suffice to think and write the answers of 40-50 objective questions. Thus, these tests are comprehensive in nature.

5.Discriminative :Notes The questions in these tests are spread out on the whole course and are of several types. Some questions have to be answered using recall, some have to be answered by knowledge, and some questions have to be answered using logic and prudence; so the classification of the students done on their basis can be reliable.

6. Practicability : Making objective test questions is a difficult task. It takes time to construct so many questions (often 100-100, 200-200). Marking them too takes time. So now in order to save effort, only multiple choice questions are constructed and their answers are marked on an answer-sheet which can be evaluated by computer. Thus, the construction, administration and evaluation of these tests have become practicable.

Limitations of Objective Test

Despite having all the merits of a good test, these tests are not devoid of some shortcomings.

1. Measurement of Cognitive Aspect Only : These tests basically help to measure the cognitive aspect of the students; these are unsuitable for the measurement of the conative and affective aspects.

2. Measurement of Memory Power Only : These tests generally examine the power of memory, the students have to hardly use the higher mental faculties such as logic and thinking.

3. No Measurement of Language skill and Expression Power : These tests do not help in the measurement of language skill and power of expression of the students, so we cannot rely on them for language tests. It is considered to be the biggest demerit of these tests.

4. Difficulty in Construction of Questions : According to some scholars, it is very difficult to construct the objective questions for these tests. They consider it a demerit of these tests. We are, however, of the view that these questions can be constructed easily by taking a little interest and care. It depends on interest and practice.

5. Administrative Difficulty : Some scholars are of the view that these tests cannot be administered easily; according to them, these tests are expensive and present difficulties in evaluation. However, our experience is that these tests are less expensive than the essay-type tests and their evaluation can also be done easily. We have already written that these tests can be evaluated with the help of computer.

6. Fluke Answers : Some people opine that these tests can be answered in a fluke. It is true, but we have already invented a statistical method to control it, which is called negative marking.

7. Use of Unfair Means : It is often heard that the students can easily copy in these tests. In our view, this demerit has occurred due to inexperienced people. The fact is that if a student copy from a book or notebook, he cannot answer all questions in the prescribed time. Of course, the students can easily copy if someone prepares the answer-sheet for him. But it is not a demerit of the test, but that of the administrators of the test.

Characteristics of a Good Test

Validity • A said to be valid if it measures what it intends to measure.

There are different types of validity :

• Operational validity

• Predictive validity

• Content validity

• Construct validity

Operational Validity

• A test will have operational validity if the tasks required by the test are sufficient to evaluate the definite activities or qualities.

Predictive Validity

• A test has predictive validity if scores on it predict future performance

Content Validity

• If the items in the test constitute a representative sample of the total course content to be tested, the test can be said to have content validity.

Construct Validity

• Construct validity involves explaining the test scores psychologically. A test is interpreted in terms of numerous research findings.

Reliability

• Reliability of a test refers to the degree of consistency with which it measures what it indented to measure.

• A test may be reliable but need not be valid. This is because it may yield consistent scores, but these scores need not be representing what exactly we want measurer.

• A test with high validity has to be reliable also. (the scores will be consistent in both cases)

• Valid test is also a reliable test, but a reliable test may not be a valid one

Different method for determining Reliability

Test-retest method

A test is administrated to the same group with short interval. The scores are tabulated and correlation is calculated. The higher the correlation, the more the reliability.

Split-half method

The scores of the odd and even items are taken and the correlation between the two sets of scores determined.

Parallel form method

• Reliability is determined using two equivalent forms of the same test content.

• These prepared tests are administrated to the same group one after the other.

• The test forms should be identical with respect to the number of items, content, difficult level etc.

• Determining the correlation between the two sets of scores obtained by the group in the two tests.

• If higher the correlation, the more the reliability.

Discriminating Power

• Discriminating power of the test is its power to discriminate between the upper and lower groups who took the test.

• The test should contain different difficulty level of questions.

Practicability

• Practicability of the test depends up on...

• Administrative ease

• Scoring ease

• Interpretative ease

• Economy

Comparability

• A test possesses comparability when scores resulting from its use can be interpreted in terms of a common base that has a natural or accepted meanings

• There are two method for establishing comparability

  • Availability of equivalent (parallel) form of test

  • Availability of adequate norms

Utility

• A test has utility if it provides the test condition that would facilitate realization of the purpose for which it is mean.

1. **Simplicity** Simplicity means that the test should be written in a clear, correct and simple language, it is important to keep the method of testing as simple as possible while

still testing the skill you intend to test. (Avoid ambiguous questions and ambiguous instructions).

2**. Scorability**

Scorability means that each item in the test has its own mark related to the distribution of marks given by (The Ministry of Education)

Scales of measurement: Whenever we measure anything, we assign a numerical value. This numerical value is known as scale of measurement. A scale is a system or scheme for assigning values or scores to the characteristics being measured (Sattler, 1992). Like for measuring any aspect of the human being we assign a numeral to quantify it, further we can provide an order to it if we know the similar type of measurement of other members of the group, we can also make groups considering equal interval scores within the group. These are called as measurement in different scales. There are mainly four scales of measurement, namely;

- Nominal
- Ordinal
- Interval and
- Ratio

Nominal scale: In nominal scale, a number is assigned for characterizing the attribute of the person or thing. That caters no order to define the attribute as high-low, more-less, big-small, superior inferior etc. In nominal scale, assigning a number is purely an individual matter. It is nothing to do with the group scores or group measurement. It is therefore, measurement in nominal scale has limited meaning, even if some experts do not consider it as a measurement. As discussed earlier, any student achievement related scores (scores in subjects) or other measurement like height, weight, etc. are the examples of nominal

measurement. Statistics such as frequencies, percentages, mode, and chi-square tests are used in nominal measurement.

Ordinal scale: Ordinal scale is synonymous to ranking or grading. It includes the characteristics of a nominal scale and provides an order to the measurement, like; when we know the achievement scores of students in a group, we can arrange them either in ascending (lowest to highest) or descending (highest to lowest) order. We can also interpret the result like; who stood first, second, 10th in the group, even the last one in the group. In ordinal scale, we can use the statistics such as median (measures of central tendency), quartile and percentile measures, correlation in rank difference method, and non-parametric tests.

Interval scale: Interval scale carries all the characteristics of earlier scales like nominal and ordinal and added with an arbitrary zero point. That is, there is no absolute zero-point or true zero point. In this scale, we can group the scores into equal intervals like, scores within the intervals of: 0-5; 5-10; 10- 15; 15-20; 20-25; 25-30 etc. This is also called as equal interval scale as the size of the classes are equal, i.e. size of the class 0-5 is 5; 5-10 is 5; 10-15 is 5; 15-20 is also 5.As there is no absolute zero point in this measurement and the existing zero value is an arbitrary one, that's why all types of measurement done in education and social sciences are usually done by the interval scale. The statistics like mean, standard deviation, product moment correlation, t-test and f-test can be used in interval scale.

Ratio scale: Ratio scale is called as the highest scale in measurement. It carries all the characteristics of earlier discussed scales with a true or absolute zero point. As there is absolute zero point in this measurement, we can say that zero height means no height. But in the case of interval scale, we cannot say that zero intelligence means no intelligence. All types of measurements conducted in Physical Sciences such as Physics, Mathematics, etc. are done by ratio scale. Ratio scales are almost non-existence in psychological and educational

measurement except in the case of psycho-physical measurement. All mathematical applications – addition, subtraction, multiplication and division can be done in ratio scale. All statistical techniques are permissible with ratio scale.

Properties of Measurement Scales

Each scale of measurement satisfies one or more of the following properties of measurement.

• Identity : Each value on the measurement scale has a unique meaning.

• Magnitude : Values on the measurement scale have an ordered relationship to one another. That is, some values are larger and some are smaller.

• Equal intervals : Scale units along the scale are equal to one another This means, for example, that the difference between 1 and 2 would be equal to the difference between 19 and 20.

• Absolute zero : The scale has a true zero point, below which no values exist. Values assigned to variables represent a descriptive category, but have no inherent numerical value with respect to magnitude.

Gender is an example of a variable that is measured on a nominal scale. Individuals may be classified as “male” or “female”, but neither value represents more or less “gender” than the other. Religion and political affiliation are other examples of variables that are normally measured on a nominal scale.

Rating scales

Meaning

By rating is meant by comparison of one individual with the other members of the group. Rating scale is one of the important techniques of evaluation. The term rating stands for expression of opinion or judgement regarding some situation, objects or character. It refers to the degree of dimension of an attribute being observed. In short, by rating means the judgement of one by another.

Rating scales resemble - checklist that is a rating scale is an improvement over list. While a checklist simply records that something happened a rating scale adds another dimension: how much or how less it happened. It enables us to indicate the degree to which a trait is present in an individual. Rating scale provides systematic procedure for obtaining, recording and reporting the observes judgment.

Some definitions of rating scale

1. According to **Barr and Johnson**, "Rating is a term applied to expression of opinion or judgement regarding some situation, objects or character. Opinions are usually expressed on a scale of value."

2. According to **Wrightstone**, "A rating scale consists of a set of characteristics or qualities to be judged and some type of scale for indicating the degree to which each attribute is present."

3. According to **Ruthstrong**, "Directed observation is rating scale."

Three points are taken into consideration while using the rating scale technique:

1. Specification of traits.

2. Decision about the sale range that is it can be 3 point scale, 5 point scale, 7 point scale, 9 point scale or 11 point scale.

3. The persons are decided who are required to use the scale to quantify the trait regarding a specific person whose personality is to be measured.

Precautions while using a rating scale

The rating scale method can be made more effective if following precautions are taken:

1. The raters should be impartial.

2. A number of raters should give their ratings and these should be taken into consideration for framing final assessment on the trait or traits.

3. Generally 'Halo effect' is there in the ratings of raters for different traits. Halo effect should be checked. Halo effect is a tendency for transfer to occur so that a rater may rate the same individual high or low for many different traits. One impression may colour the rest. Suppose a rater rates a child high for his politieness and he transforms this good impression for other traits of the child example emotional stability, honesty perseverance etc.

4. Avoid to be too lenient or true strict in standard.

5. Avoid the tendency to either remain at the neutral point or extreme point in the rating scale.

Purpose of rating scale

According to **Freeman** the purposes of rating scales are as follows:

1. Rating scale is chiefly useful for knowing what impression an individual has made on persons with whom he has come in contact in respect to some specified trait or attitudes.

2. It is a device that rates social values, occupational efficiency, group status and the like in certain specified areas.

3. It reflects the impression the subject has made upon the persons who do the rating.

4. For the evaluation of an individual rating scales, these are submitted to teachers, counsellors, employers, colleagues, parents and others who have had sufficient contact with the person in question.

5. Rating scales may be devised for a variety of traits such as tact, generosity, leadership, cooperativeness, resourcefulness, punctuality, industriousness, honesty, emotional control, study habits and personal attractiveness etc.

Points for construction and use of rating scale

The following are the major aspects which are considered in the construction and use of rating scale according to **Freeman.**

1. Each trait should be clearly defined.

2. The degree of the trait should be defined. Each trait should be rated on a scale, most frequently of five or seven intervals.

3. The mean or the median of the ratings by different judges should be taken for having a reliable rating. Methods of studying reliability of rating scales most commonly used include the following:

1. Repeating judgements after a time interval.
2. Correlation between rating of two or more judges.
3. Relationship between judges ratings and self rating.

4. The principal indication of validity of some rating scale is the fact that persons are using them. Guidance, counsellors, employers and personnel officers find them helpful if the judges are carefully selected and if the ratings are properly made.

In rating scales overt traits are more reliability rated than covert traits. For example overt traits like emotional expression, social acceptability, manifest fear and anxiety, aggressive or impulsive acts are rated with greater reliability then the covert traits like persons inner life and feelings about oneself.

5. Degree of certainty of rating should be stated e.g. very strong, strong moderate.

6. Extroverted persons are more reliably judged than introverted.

Types of rating scales

There are four types of rating scales.

1. Numerical rating scale.

2. Graphic rating scale.

3. Descriptive graphic rating scale.

4. Percentage of group scales.

Numerical rating scales: These numbers are assigned to each trait. It is an even point scale, the number 7 represents the maximum amount of that trait in the individual, and 4 represents the average. The rater merely enters the appropriate number after each name to indicate the judgment of the person.

Method of Paired Comparisons: In this, the rater compares each person being rated concerning the trait of every other individual being rated in the general terms of equal, better, or worse.

Graphical rating scales: It is commonly used Examples are: He attends to important details

1	2	3
Always	Sometimes	Never

1	2	3	4	5
Aphthetic	Rarely	Sometimes	Usually	Highly
Cooperative	Cooperative	Cooperative	Cooperative	

Scorecards:

It is a type of scale in with whatever is being rated is analyzed into its parts. An expert assigns each part a maximum score. The rater assigns a value to each item, and he passes judgment, and these values are totaled, and a final score is pronounced.

Man to man Scales:

In this case, an individual is asked to rate the person to be assessed by comparing him to a person already rated and assigned a position on the scale. The rate is assigned to his job.

Errors in Rating:

The following are the common errors in rating

1 **Generosity Error**: This sometimes means the rater does not want to run down the rate, the latter being his favorite, by giving him low ratings.

2 **The Errors of Hollow Effect**: Sometimes, the rating is done in terms of general impression about the rates formed based on some past performance.

3 **The Errors of Central Tendency:** Sometimes, the ratertends to rate all or any of the rates near the midpoint on the scale.

Advantages of rating scales

- They are helpful in writing reports to the parents.
- They are helpful in finding students needs.
- These can judge various aspects of parent social developments as interest, classroom conduct and cooperation etc. rating scales help in excellent teaching techniques.
- It can be used with a large number of stimuli.
- These are supplementary to other resources of understanding about the students.
- The teacher who has little training for the purpose can also use it.
- They give a common scale on which judgements can be recorded.
- It directs our attention to the same aspect of achievement in all the students.
- The teacher can make a common frame of reference for comparing all the students on the same set of characteristics.

Limitations

Following are some limitations of rating scale.

1. **A subjective technique**

Rating is entirely a subjective technique. Even the most impartial judgement about one particular pupil made by a number of judges vary. This is because of the subjective factor or

the personal equation. Hence, rating cannot be relied upon in its entirety. It must be supplemented with personal observation and other techniques. Personal interviews will, to a large extent, smoothen the angularities found in rating. Even then, rating is a very useful technique employed by counsellors for individual and group guidance.

2**. Errors of leniency**

There is most of the time error of leniency. It is also known as generosity error. There is human and unwillingness to give unfavourable judgement of their fellows. The raters own feelings and sympathy towards a particular ratee compels him to be a little generous while rating a particular individual.

3. **Stringency errors**

Some raters have the tendency to rate all the ratees low.

4**. Halo error**

Another error that crops up in rating is the halo error. Halo means a tendency to rate in terms of general impressions about the ratee formed on the basis of some previous experience. For example, the rater believes that a person who is good in English must be good also in playing football.

5. **Error in Central tendency**

There is a tendency in some raters to rate all the ratees near the midpoint of the scale. They want to give rating which is average.

6**. Logical error**

This type of error occurs when two characteristics are rated as more alike or less alike than they actually are. For example, in rating intelligence, teachers desire to over-rate the intelligence of the students with high achievement because logically they expect that two characteristics go together.

7**. Proximity error**

It has been seen that adjacent traits on a rating tend to inter-correlate higher than the remote ones, their degree of whole similarity being approximately equal. These errors may be counteracted to some extent by placing similar traits further apart and different ones close together.

Guideline for developing and improving rating scales

1**. Defining the trait**

Traits which we want to rate should be defined properly. For example if we want to rate sociability of a person, first of all we should formulate definition of sociability and then try to rate it.

2. **Defining the scale**

The scale should be clearly defined that is we are rating 3,5,7,9 or 11 point scale.

3. **Straight forward and unambiguous trait**

Traits which are straight forward and unambiguous should be chosen.

4. **Number of traits**

Limit the number of traits or characteristics to be rated, say not more than four or five at a time.

5. **Different situations**

The rater should observe the trait in different situations.

6. **Directions**

The direction of using the rating scale should be clear and comprehensive.

7. **Providing some space**

In the rating scale cards, some space should be provided for the rater to write some supplementary notes.

8. **Use of the scale**

Use of the scale should be preceded by a period of training and practice in its use.

9. **Trained raters**

Experienced, trained and wise raters should be selected for training.

10. **Several raters**

Several raters should be employed to increase reliability of any rating scale.

11. **Independent judgement**

The raters should make independent judgement without consulting others.

12. **Uniform standard**

Uniform standard of rating should be observed.

Questionnaires

A questionnaire is a useful tool of evaluation. A large amount of data on various aspects of the theme in hand can be collected and the person upon whom the questionnaire is administered can be evaluated accordingly. In the other words of Good, "In general, the word questionnaire refers to a device for securing answers to questions by using a form which the respondent fills himself." It is a general device and can be used to measure all kinds of information. It may or may not be totally connected with measurement of personality.

A questionnaire is a valuable tool for collecting information directly given by a person. Such an information may consist of personal knowledge, likes and dislikes, values and preferences, attitudes, beliefs, experiences and present status of things or events. This information can be both qualitative (verbal description, comments or views) as well as quantitative (numbers or scores).

An assembly of questions cannot be called a questionnaire. It must reflect an objective, a design and a frame work. It is a way of obtaining data about persons by asking them rather than watching them behave or by sampling a portion of their behaviour. A good questionnaire embodies a beginning, a middle and an end. There is an introductory section comprising a note on the purpose of the questionnaire, an appeal and a direction for responding to the

various questions and a basic data chart which is to be filled in accurately by respondents. In the middle of the questionnaire are placed the questions organised or arranged in to different sections according to the relevance of information sought. At the end, the questionnaire provides concluding remarks and acknowledgements to the respondents or any other source of authorities concerned.

Designing of a questionnaire

Following points should be taken into consideration for designing a questionnaire:

1**. Objectives**.

The information which is needed for research study becomes the basis for the formulation of objectives. List of possible information to be collected is prepared. This information is further sub-divided into categories and sub-categories.

2. Formulation of Questions and Responses.

The questions and types of responses should be decided in advance. A draft is prepared. A questionnaire may include the following types of questions and responses.

1. **Types of Questions**

(a) Direct and Indirect Questions. Use of indirect questions is preferred.

(i) Direct questions. Do you favour Nationalisation of Banks?

(ii) Indirect questions. Indicate your opinion on nationalisation of Banks.

(b) Specific and non specific questions. Specific questions ask for particular information. On the other hand, non specific questions elicit the information in a roundabout manner e.g.

(i) Specific question. How many hours you are required to teach?

(ii) Non specific question. How many hours you actually teach?

(c) Fact and Opinion questions. A factual question requires the facts while opinion question ask for opinion.

(d) Statements. Statements as well as questions may be used in a questionnaire. e.g.

(i) New scheme should be dropped. agree/disagree

(ii) Do you think that the new scheme should to dropped? Yes/No

2. Types of responses to questions

a. **Unstructured Responses**

b. **Fill in Response**

c. **Tabular Type Response**

d. **Scaled Response**

e. **Ranking Response**

f. **Check List Response**

g. **Categorical Response**

3. Examples of Questionnaires for Self-Appraisal of Personality

1. Attitude Questionnaire
2. Attitude Scale
3. Biographical Data Blank
4. Projective Questionnaire
5. Cattle's 16 Personality factor questionnaire

<u>Characteristics of a Good Questionnaire</u>

The following are characteristics of good questionnaires:

- It should consist of a well-written list of questions.
- The questionnaire should deal with an important or significant topic to create interest among respondents.
- It should seek only that data which cannot be obtained from other sources.
- It should be as short as possible but should be comprehensive.
- It should be attractive.
- Directions should be clear and complete.

- It should be represented in good psychological order proceeding from general to more specific responses.
- Double negatives in questions should be avoided.
- Putting two questions in one question also should be avoided. Every question should seek to obtain only one specific information.
- It should be designed to collect information which can be used subsequently as data for analysis.

Format of Questions in Questionnaire

The questions asked can take two forms:

- **Restricted questions**, also called closed-ended, ask the respondent to make choices — yes or no, check items on a list, or select from multiple choice answers.
- Restricted questions are easy to tabulate and compile.
- **Unrestricted questions** are open-ended and allow respondents to share feelings and opinions that are important to them about the matter at hand.
- Unrestricted questions are not easy to tabulate and compile, but they allow respondents to reveal the depth of their emotions.
- If the objective is to compile data from all respondents, then sticking with restricted questions that are easily quantified is better.
- If degrees of emotions or depth of sentiment is to be studied, then develop a scale to quantify those feelings.

Advantages of Questionnaires

Some of the many benefits of using questionnaires as a research tool include:

- **Practicality:** Questionnaires enable researchers to strategically manage their target audience, questions and format while gathering large data quantities on any subject.

- **Cost-efficiency:** You don't need to hire surveyors to deliver your survey questions — instead, you can place them on your website or email them to respondents at little to no cost.
- **Speed:** You can gather survey results quickly and effortlessly using mobile tools, obtaining responses and insights in 24 hours or less.
- **Comparability:** Researchers can use the same questionnaire yearly and compare and contrast research results to gain valuable insights and minimize translation errors.
- **Scalability:** Questionnaires are highly scalable, allowing researchers to distribute them to demographics anywhere across the globe.
- **Standardization:** You can standardize your questionnaire with as many questions as you want about any topic.
- **Respondent comfort:** When taking a questionnaire, respondents are completely anonymous and not subject to stressful time constraints, helping them feel relaxed and encouraging them to provide truthful responses.
- **Easy analysis:** Questionnaires often have built-in tools that automate analyses, making it fast and easy to interpret your results.

Disadvantages of Questionnaires

Questionnaires also have their disadvantages, such as:

- **Answer dishonesty:** Respondents may not always be completely truthful with their answers — some may have hidden agendas, while others may answer how they think society would deem most acceptable.
- **Question skipping:** Make sure to require answers for all your survey questions. Otherwise, you may run the risk of respondents leaving questions unanswered.

- **Interpretation difficulties:** If a question isn't straightforward enough, respondents may struggle to interpret it accurately. That's why it's important to state questions clearly and concisely, with explanations when necessary.
- **Survey fatigue:** Respondents may experience survey fatigue if they receive too many surveys or a questionnaire is too long.
- **Analysis challenges:** Though closed questions are easy to analyze, open questions require a human to review and interpret them. Try limiting open-ended questions in your survey to gain more quantifiable data you can evaluate and utilize more quickly.
- **Unconscientious responses:** If respondents don't read your questions thoroughly or completely, they may offer inaccurate answers that can impact data validity. You can minimize this risk by making questions as short and simple as possible.

Schedule

Schedule is a device consisting of set of questions which are asked and filled in by an interviewer in a face to face situation with another person. It differs from the questionnaire in that the former is administered personally to a respondent or a group of respondents while the later can be mailed. The schedule has many advantages over the questionnaire. It helps the researcher to explain the nature and purpose of investigation and to make the meaning of the questions clear to the respondents if they misinterpret or give incomplete or indefinite responses. It also economizes time and expenses of the investigation. Sometimes, it may not be possible to contact personally all 11 the respondents either individually or in a group and it is one of the limitations of the schedule. The procedure of constructing a schedule, analysis and interpretation of data gathered through it is almost same as that of a questionnaire.

Inventory

Inventory is a criterion referenced test designed to determine if an individual has an accurate working knowledge of specific field. It is generally a multiplechoice test in order to aid interpretability and facilitate administration in large classes. Inventories are used in the research purposes. In research inventory aims to ascertain the range of what individuals think a particular question is asking and the most common responses to the questions. An inventory is a kind of self-report instrument. The individual checks responses to certain questions or statements. Inventories yield scores which are assumed or have been shown to measure certain tendencies or traits. Interest inventories attempt to yield a measure of the types of activities that 7 an individual has a tendency to like and to choose. In 1907, Hall, a psychologist, standardized “Recreational Interest Inventory” and in 1934, Kuder, another psychologist, standardized “Occupational Interest Inventory”.

Types of Inventory

1. **Direct Questioning** - It presents questions like - 'Do you like to play cricket?'

2. **Direct Observation** - Behaviour of the individual is observed and it is found out in which activities the individual is more interested.

3**. Tested Interest** - The individual is exposed to various experiences related to different types of interests and it is observed in which area the individual is showing more interest.

4**. Interest Inventory** - This is most preferred way of measuring interest. It consists of various statements and the individual is asked to arrange them in order of preference. Mostly interest inventories are of verbal type.

5. A Personality Inventory attempts to measure certain personality traits such as individual need, adjustment, etc. It consists of questions of 'Yes/No' type.

Advantages and Limitations

Interest inventories are useful in educational and vocational guidance. They are helpful to the teachers to provide educational facilities to the students as per their interests. Inventories can be administered of individuals as well as on groups. Personality inventories have been proved valuable in diagnosis in clinical situations for the psychiatrists. An inventory is not very much timebound. Any can administer the inventory without any training to get expertise. But, because of individuals' inability or unwillingness to report their own reactions accurately or objectively these instruments may be of limited value. This limitation may be attributed to the inadequate theories of personality upon which some of these inventories have been based.

Observation

Observation is the process in which individuals observe the events and occurrences in the real life situations and classify and record those happenings according the prior plan. It is used to evaluate the overt behaviours, events, and the contexts surrounding the events and behaviours in controlled and uncontrolled situations. Patton (1990) has suggested five dimensions along which observation varies:

Observational methods have occupied an important place in the field of education and research. These include the observation of the setting or physical environment, social interactions, physical activities, non-verbal communications, planned and unplanned activities and interactions. Observation needs proper planning, expert execution and adequate recording and interpretation so as to prove itself a good technique. Let’s discuss these steps one by one to understand the process of observation better.

Fifth: Breadth of Focus

Fourth: Duration of the observation.

Third: those being observed may be given full explanations, partial explanations, no explanations, or given a false explanation.

Second: the observer may conduct the observations covertly, with full knowledge of those being observed, or with only some of those being observed aware of the observation.

First: observer's role may vary from full participant to a complete outsider without disturbing the subjects who are observed.

Planning for Observation

The planning for observation includes definition of specific activities or units of behaviour to be observed, the nature of the groups of subjects to be observed, the scope of observation-individual or group, length of each observation period, interval between different periods, deciding about the tools to be used in making the observation and recording, training of the observer to develop expertise, interpretation of observation, etc.

Execution of Observation

Execution of the observation can be done appropriately if the observer is aware of the type of the observation being utilized for the purpose. Observation may be classified as participant or non-participant and unstructured or structured observation. In the participant observation, the observer becomes one of the members of the group under observation. Here the observer remains in the site of the person being observed and actually take part in some activity within the group. Participant observer has to play any role in the group may be as a

stranger, an attentive listener, an eager learner, etc. In non-participant observation, observer takes a position where it does not disturb the group. In this type of observation, a one-way vision screen that permits the observer to see the subject but prevents the subject from seeing the observer, is useful.

Unstructured observation is associated with participant observation. It is an exploratory technique. In this type of observation, it is not possible to categorize behaviour in advance. Instead of using predetermined categories, the observer considers aspects of behaviour in terms of their context or the situation of which they are part. On the other hand, structured observations are formal and designed to provide systematic description to test casual hypotheses. It is executed in controlled situations like classroom or laboratory settings. Interaction analysis of the classroom verbal behaviour of a teacher is an example of structured observation. Structured observation starts with relatively specific formulations. There is much less choice with respect ot the content of observation. The observers set in advance categories of behaviour 10 in terms of which they wish to analyze the problem, and keeps in mind the time limit under which the observations are to be made.

Recording and Interpreting Observation

The recording of the observation data may either be simultaneous or soon after the observation. In viewing, classifying and recording behaviour, the observer must take utmost care to minimize the influence of his biases, attitudes and values on the observation report. In order to overcome the biases introduces by the human observer, various mechanical instruments are used to obtain a more accurate record of events. The use of camera, tape recorder, stop watch, one-way vision screen or mirror, etc. allows behaviour to be measured to a degree of accuracy which could not be achieved by the human observer. It is advisable to develop an observation form or schedule while making observations.

Observation provides a direct procedure for studying various aspects of human behaviour, which may be the only effective way to gather data in a particular situation. It enables the observer to code and record behaviour at the time of its occurrence. The limitations of this method are that, sometimes a subject may intentionally attempt to exhibit an artificial behaviour when knows about being observed. Also, observation method is time consuming and sometimes too costly.

Types of observation

There is a spectrum of styles of observing: from the non-structured to the highly structured, and from the observer as participant in the activity, to the observer as non-participant, perhaps even using a video camera. Two types of observation are generally used in an evaluation effort. They are:

i) Non-structured observation:

Non-structured or open or natural observation allows the collection of a rich variety of information. In such situations the observer, while being present at the site, does not control or manipulate anything. It is a technique of many social anthropologists. Here the observer enters the observation site with as open a mind as possible. He/she can see the things and record them in a natural setting. However, practice and skills are needed by the observer in drawing his or her experience and judgement to focus upon, and record events considered to be important. It must be realised that however experienced the observer, only a fraction of the interactions and events occuring will be seen and recorded. For this reason, it is often valuable to focus upon certain aspects of the event, object, etc., being observed. For example, behaviour in a counselling session -- behaviour of the counsellor, behaviour of the learners,

and the interaction between the counsellor and learners - can best be assessed through natural observation.

ii) Structured observation:

An evaluator observes the phenomenon under structured conditions, with the knowledge of the person(s) being observed. The observation situation may also b'e simulated and observed. This technique allows the evaluator to observe particular behaviours. For example, a teacher or a trainee observes the role play in a teacher-parent conference, etc. The major disadvantage of this type of observation is that it is not natural, and the behaviour exhibited by people may not be the behaviour that would occur in a natural setting. People may behave the way they think they should behave rather than the way they normally would. A checklist to be used during the observation process is valuable to have accurate records of the behaviour expressed.

Steps of observation

- Planning
- Execution
- Recording and interpretation

Planning

The characteristic, topic or the thing to be observed is decided in this stage. Whether it is group observation or personal observation when and how many times the observation would be done, the tools useful for recording the observation etc. is also decided here. The specific type of training is helpful during interpretation. Who will be doing observation or the observer is also pre-decided.

Execution

The arrangement for observation is made. The necessary method for the observation as such the natural or artificial arrangement is made. Then after that environment of that opportunity is given so that the person is motivated to behave in some manner and that behaviour is observed.

Recording and interpretation

If the tools or instruments are ready, then observation turns a fast process. The observation or recording is hereby evaluated and interpreted.

Uses and limitations of observation

Some of the uses and limitations of observation are given below:

- It helps us to get first hand information;
- Certain traits like honesty, punctuality, truthfulness, etc., can be observed;
- Structured observation yields objective and accurate data;
- The observer codes and records the overt behaviour at the time of its occurrence;
- However, there is the possible tendency of an observer to let overall 6 feelings towards an individual or an initial impression affect subsequent observations; and
- A 'subject' may intentionally attempt to exhibit artificial behaviour.

Interviews

By interview we mean face to face method of collecting information. The oldest form of obtaining knowledge has always been an important tool in the hands of certain professional men such as lawyers, doctors and newspaper reporters. In education its chief use lies in diagnosis and guidance. Much can be observed and recorded from the gestures, voice and several models of expression of the interviewer. The interview is an important tool used to supplement the ordinary objective evidence about a pupil, such as is afforded by his school record and by a first-hand knowledge of such things as his feelings and points of view. More over when the teacher seeks to collect some confidential information about the pupil, interview seems to be the best technique.

Some definitions of interview

1. According to **Brown and Gheselli**, "The term interview stands for the generic concept which includes a variety of procedure used in collecting data through a person to person contact between the interviewer and respondent."

2. According to **Good and Hatt**, "Interview is basically a process of social interaction."

3. According to **P.V.Young**, "Interview is an interactional process, it is a mutual views."

Purposes of Interview

According to **Hahn and Mclean**, interview technique can be used for the following purposes:

1. To collect new information and simplify or interpret the information already gathered.

2. To permit counselee to think aloud in the presence of a sympathetic listener.

3. To carry necessary information to the counselling.

4. To find socially acceptable and personality satisfying alternatives with and counselee.

Types of interview

According to function

- Diagnostic

- clinical
- research

According to number of participants individual

- group
- single interviewer
- panel interviewer

According to methodology

- structured
- unstructured

1**. Diagnostic interview**

In this type of interview, an effort is made to diagnose the problem or symptoms. In other words the diagnosis interview is arranged with a view to identify the problems and collect necessary information so that required treatment may be provided.

2. **Clinical interview**

clinical interview is used for the purpose of intensive study of individuals, gather data from multiple sources. They complete the picture of combining the information gathered through interview.

Interview from the point of view of participants

1. **Individuals interview**

In this interview, only one individual is interviewed at one time by an interviewer.

2. **Group interview**

Hear a group of individuals is interviewed at one time by an interviewer.

3. **Single interviewer**

In this type only single interviewer conducts the interview of individuals as well as groups.

4. **Panel interviewer**

Such type of interview is held by a panel of interviewers composed of different experts but related with different fields. Usually panel interviewers hold the interviews for selection and treatment purposes.

Methodology or Role

1. **Structured or formal interview**

Here the structure of the interview is already setup such as number of questions, time limit of interview and working of questions etc. Everything in the interview is standardized.

2**. Unstructured or informal interview**

Such type of interview is flexible in nature. The interviewer is free to use any type of questions. The numbers of questions, time limit of interview, wording of questions are not predetermined. The discussions with the respondents are merely informal in nature.

Techniques of conducting interview

The techniques of conducting interview effectively involves the following points

1. **Congenial physical conditions**

An interviewer should see that physical conditions are congenial and that there is not much external interruption caused by noise etc.

2. **Rapport with the interviewer**

Rapport with the interviewee by asking friendly questions is essential for the success of an interview.

3. There should be no evidence of fatigue.

4. **Keeping in view the purpose of interview**

The purpose of interview must be kept in view and all interview should be directed towards that end.

5. **Cordial nature of interviewer**

The interviewer should be of friendly and cordial disposition. Interview should not be interrupted by the interviewers critical or condemnatory remarks.

6. **Winning the confidence of interviewee**

Before the interview is closed the interviewee should get the assurance that he/she can confide in the interviewer. He/she must go satisfied with some plan of action.

7. **Personal qualities of the interviewer**

In an interview, the personal qualities of the interviewer are as important as the use of techniques.

Advantages of interview as a technique of evaluation

1. **More responses**

In this technique more responses can be obtained as compared to responses obtained by post in a questionnaire.

2. **Confidential information**

The respondents give even the confidential information if proper rapport is setup.

3. **Misunderstanding are cleared**

In face to face interview the responding can I clear misunderstanding about the interview and the use of data.

4. **Exchange of ideas**

Even the ideas can be exchanged between the interviewer and the respondent by asking supplementary questions from both sides.

5. **Reliable information**

The information obtained through interview is more reliable.

Limitations of interview

1. **Time consuming**

Interview process is time consuming as time is wasted before the start of interview.

2. The interview is a very costly process as it needs a large number of field workers trained for the purpose.

3. **Lack of expertise**

The interviewer may be lacking the expertise needed.

4. **Application of the method is limited**

This method fails with deaf, dumb, shy people and infants.

Performance Test

Performance Testing is a software testing process used for testing the speed, response time, stability, reliability, scalability and resource usage of a software application under particular workload. The main purpose of performance testing is to identify and eliminate the performance bottlenecks in the software application. It is a subset of performance engineering and also known as "Perf Testing".

The focus of Performance Testing is checking a software program's

- Speed – Determines whether the application responds quickly
- Scalability – Determines maximum user load the software application can handle.
- Stability – Determines if the application is stable under varying loads

Types of Performance Testing

- **Load testing –** checks the application's ability to perform under anticipated user loads. The objective is to identify performance bottlenecks before the software application goes live.
- **Stress testing –** involves testing an application under extreme workloads to see how it handles high traffic or data processing. The objective is to identify the breaking point of an application.

- **Endurance testing –** is done to make sure the software can handle the expected load over a long period of time.
- **Spike testing –** tests the software's reaction to sudden large spikes in the load generated by users.
- **Volume testing** – Under Volume Testing large no. of. Data is populated in a database and the overall software system's behavior is monitored. The objective is to check software application's performance under varying database volumes.
- **Scalability testing** – The objective of scalability testing is to determine the software application's effectiveness in "scaling up" to support an increase in user load. It helps plan capacity addition to your software system.

Common Performance Problems

Most performance problems revolve around speed, response time, load time and poor scalability. Speed is often one of the most important attributes of an application. A slow running application will lose potential users. Performance testing is done to make sure an app runs fast enough to keep a user's attention and interest. Take a look at the following list of common performance problems and notice how speed is a common factor in many of them:

- **Long Load time –** Load time is normally the initial time it takes an application to start. This should generally be kept to a minimum. While some applications are impossible to make load in under a minute, Load time should be kept under a few seconds if possible.
- **Poor response time –** Response time is the time it takes from when a user inputs data into the application until the application outputs a response to that input. Generally, this should be very quick. Again if a user has to wait too long, they lose interest.

- **Poor scalability –** A software product suffers from poor scalability when it cannot handle the expected number of users or when it does not accommodate a wide enough range of users. Load Testing should be done to be certain the application can handle the anticipated number of users.
- **Bottlenecking –** Bottlenecks are obstructions in a system which degrade overall system performance. Bottlenecking is when either coding errors or hardware issues cause a decrease of throughput under certain loads. Bottlenecking is often caused by one faulty section of code. The key to fixing a bottlenecking issue is to find the section of code that is causing the slowdown and try to fix it there. Bottlenecking is generally fixed by either fixing poor running processes or adding additional Hardware. Some **common performance bottlenecks** are
- CPU utilization
- Memory utilization
- Network utilization
- Operating System limitations
- Disk usage

Performance Testing Process

The methodology adopted for performance testing can vary widely but the objective for performance tests remain the same. It can help demonstrate that your software system meets certain pre-defined performance criteria. Or it can help compare the performance of two software systems. It can also help identify parts of your software system which degrade its performance.

Below is a generic process on how to perform performance testing

1. **Identify your testing environment** – Know your physical test environment, production environment and what testing tools are available. Understand details of the hardware, software and network configurations used during testing before you begin the testing process. It will help testers create more efficient tests. It will also help identify possible challenges that testers may encounter during the performance testing procedures.
2. **Identify the performance acceptance criteria –** This includes goals and constraints for throughput, response times and resource allocation. It is also necessary to identify project success criteria outside of these goals and constraints. Testers should be empowered to set performance criteria and goals because often the project specifications will not include a wide enough variety of performance benchmarks. Sometimes there may be none at all. When possible finding a similar application to compare to is a good way to set performance goals.
3. **Plan & design performance tests –** Determine how usage is likely to vary amongst end users and identify key scenarios to test for all possible use cases. It is necessary to simulate a variety of end users, plan performance test data and outline what metrics will be gathered.
4. **Configuring the test environment** – Prepare the testing environment before execution. Also, arrange tools and other resources.
5. **Implement test design –** Create the performance tests according to your test design.
6. **Run the tests –** Execute and monitor the tests.
7. **Analyze, tune and retest** – Consolidate, analyze and share test results. Then fine tune and test again to see if there is an improvement or decrease in performance. Since

improvements generally grow smaller with each retest, stop when bottlenecking is caused by the CPU. Then you may have the consider option of increasing CPU power.

Oral Tests

Oral tests are used extensively in the lower classes as well as in the high school. Most mathematical teachers make use of oral tests to develop the mental skills necessary for increasing speed and accuracy and habit formation.

How can make better your Oral Test

• Preparation- reviewing concepts you'll be tested on. Identify main details within lecture notes, textbooks, and other sources.

• Listen Carefully to the Questions

• Stay Focused on a Topic

• Think in Threes-Provide three supporting points to support your position,(supported with effective arguments.)

Advantages of Oral Tests

In the past, the classroom teacher relied very heavily on the oral work of his pupils in order to arrive at an estimate of the extent to which they mastered the work of his course. The value of the oral examination is quite apparent. Unlike the situation with written examination, which assumes that the examine understands the questions, the oral examiner can pose a question and modify the same if it is not understood. Because of this flexibility, the examiner can probe the depth of student's understanding better through an oral test.

The probing results not only in a more specific answer but also gives some indication of the thought process used by the students in answering the questions. The teacher can ask

for clarification, is needed oral tests help the teacher to evaluate how well a pupil can synthesise, integrate and organizethe materials that he learnt. It becomes a valuable tool for the diagnosis of pupil's difficulties. Skilful questioning by the teacher to may help the pupil to apply known mathematical principles and formulae to a new situation. Oral tests are very valuable for a teacher of mathematics.

Disadvantages

Nevertheless there are serious of oral examination which in habit its use. Probably the most prominent weakeness is the unreliability of the oral tests. Such factors as lack of precision in actually conducting an oral examination. Failure to pre-plan the questions, prejudices of the teachers and so on serve to detract from the usefulness of this technique. Moreover, it is time consuming and it may be difficult to test each student over the entire are under study.

Diagnostic tests

Meaning

Educational diagnostic testing is a form of assessment that occurs before instruction begins. The purpose of administering diagnostic tests is to try to determine what students already know about the concepts and skills to be covered by instruction. The tests are not graded. The tests can determine if differentiated instruction is a need and discover students' preferred learning styles as well as their strengths, weaknesses, and misconceptions.

Educational Diagnosis

Educational diagnostic testing is a form of assessment that occurs before instruction begins. The purpose of administering diagnostic tests is to try to determine what students

already know about the concepts and skills to be covered by instruction. Diagnostic tests are not graded.

Diagnostic tests can also be called pre-assessments, predictive assessments, or diagnostic pre-tests. Educational diagnostic testing can be used to evaluate students' level of academic achievement and learning potential and can sometimes lead to additional assessments if warranted by students' responses. Standardized tests and instructor-constructed assessments can be used as diagnostic tools to determine students' level of prior knowledge. One-on-one interviews may also be used as a diagnostic tool. Using diagnostic testing can help instructors plan differentiated instruction, lesson plans, teaching strategies, and other classroom techniques to help all students in the classroom achieve their academic potential.

The corrective diagnosis can be made at the following levels.

• Classification

• Finding the nature of difficulties

• Finding the causes of difficulties

• Providing remedial measurement

Classification

It is the process of sorting out students into groups, particularly of underachievers and lower achievers. If the students have not reached their expected level of performance, they will be the ones who need remediation. However, if they have crossed their level, some enrichment programs may be planed to help them improve their achievement future.

Finding the nature of difficulties

In this level of diagnosis, the specific areas of difficulties have to be identified. Achievement tests, unit tests, etc. can be used for this purpose as they cover as many learning points as possible. If most of the students do poorly on a particular learning point, it is an indication that something is wrong with the instruction relating to that point.

Finding the causes of difficulties

This is the most difficult stage in diagnosis. The main difficulty of this lies in the fact that test appraises only the products of learning and not the process of learning. They may establish where the breakdown in learning has taken place but can seldom reveal anything about the causes of it. Most of the causes of difficulties may be located in the areas of academic aptitude, retardation of basic skill, study habits, physical factors, and emotional factors.

Remediation

After identifying the causes of the difficulties, the next stage is planning and applying remedial measures. There is, however, no set pattern and no cut and tried formulae for remediation. In some cases, it may be a simple matter of review and re-teaching. In others, an extensive effort to improve motivation, correct emotional difficulties, and overcome deficiencies in work and study habits may be required.

Prevention

Prevention is better than cure in education elsewhere. A program of diagnostic testing should help an imaginative teacher in getting an insight into types of possible errors that are likely to occur in learning, their possible causes, and the ways of preventing them in the future. Thus educational diagnosis does not and should not end at remedial measures but

also should become a means of improving instruction, modifying curriculum, and also for refining instructional materials and strategies.

Construction of a Diagnostic Test

Before the construction and use of a diagnostic test by a teacher, students who struggle to learn the subject content are to be identified. Generally, by analyzing the achievement test scores, the teacher could easily identify the students who lay behind significantly in their learning achievement.

After arranging the achievement test scores of students in the descending order, the bottom 10% of students could be identified as backward in learning. It could be easily inferred that they are having learning difficulties.

Following are the five steps in the construction of a diagnostic test that helps to identify the students with learning difficulties, the content areas which they find it difficult to learn and the nature of their learning difficulties.

1. Planning

2. Writing the test items

3. Assembling the test items to form the question paper

4. Providing clear instructions for answering the test

5. Providing the scoring key and marking scheme.

Uses of Diagnostic Test

The following are the uses of diagnostic tests

• Point out inadequacies in specific skills.

• Locate areas in which individual instruction is required.

• Furnish continuous information so that learning activities may be most productive of desirable outcomes.

• Serve as a basis for improving instructional methods, instructional material, and learning procedures.

Remedial Measures

Remedial-Teaching is conducted to eradicate the shortcomings which are found out by the diagnostic tests. The success of remedial measures depends on the fact how widely the teacher knows of the students' shortcomings. The plan for individual or collective remedial measures is adopted according to the nature of shortcomings. It is necessary that the remedies of shortcomings are done readily and immediately, else the shortcomings may become permanent. For remedial measures, the students are divided into different groups, as dull, common and intelligent, etc. The collective and individual remedial measures for these different groups are prepared differently for each of these groups. The forms of remedial measures are determined by the teachers, but this task must be conducted immediately. The following factors should be kept in view while conducting remedial measures :

(1) The weak students should be asked to sit on the front seats in the class.

(2) The development of the subject matter should be done with the help of examples and illustrations.

(3) The attention of the students should be drawn to those concepts, principles and activities related to the subject matter in which they commit errors.

(4) The fundamental concepts of mathematics and other subjects, such as factors, percentage, unit, square root, etc. should be taught carefully.

(5) The students should be provided sufficient opportunity for thinking and reasoning in the class.

(6) The concepts should be made clear to the weak students by using models, charts and other audio-visual aids.

(7) The matter written on the blackboard should be clear, correct, orderly and useful.

(8) The exercises on each sub-topic should be such which the students can think about themselves.

(9) The correction in the written work of the students should be done in their presence.

(10) The students should be given individual counselling even after the class, to help them in learning.

UNIT III

PSYCHOLOGICAL TESTING

Psychological tests

Meaning of psychological tests

1. According to **J.L. Mursell**, "A psychological test is a pattern of stimuli, selected and organised to elicit responses which measure certain psychological characteristics in the person who takes them."

2. **Lee J. Cronbach's** view, "A psychological test is a systematic procedure for comparing the behaviour of two or more persons".

3. According to **Anastasi Amne**, "A psychological test is a standardized instrument designed to measure objectively one or more aspects of total personality by means of samples of verbal or non-verbal or by means of other behaviours".

Types of psychological tests

Psychological tests may be divided into two categories

1. Individual Tests

(a) Individual verbal tests

(b) Individual non-verbal tests

2. Group Tests

(a) Group verbal tests

(b) Group non-verbal tests (Cultural tree test and paper pencil tests)

Uses of psychological tests

1. Analysis and description of individuals characteristics.

2. Evaluation, prediction guidance.

3. Determination of vocations and vocational preparation and selection for a job.

4. Diagnosis for various purposes.

5. Research for various purposes.

Basic principles regarding construction of a psychological test

There are three basic principles of construction of a psychological test

1. Objectivity

2. Representative population sample

3. Standardization

1. Objectivity

The purpose of standardizing a test is to give it objectivity i.e. to devise an instrument that so far as possible will be free from subjective judgement regarding the regarding the ability, skill, knowledge, trait, potentiality to be measured and evaluated. There are several elements or aspects that make a test objective. The elements are as under.

1. Every one who administers the test does so according to uniform and specified set of instruction.

2. The responses to test items are uniformly scored according to specific answers or specimen answers provided in a manual.

3. Norms of performance are based upon a population sample that has been scientifically selected for the purpose of the particular test.

4. Mental activities or personality traits to be tested are defined and specified and psychological rationale is given.

5. The activities or traits to be tested have been selected on the the basis and analysis of the operation or behaviour to be evaluated upon the views of a number of exports and upon information available from previous research.

6. The content of the test under construction is subjected to analysis by means of established techniques of test standardization.

2. Representative population sample

Every test is designed for use with a population or group, for instance

1. A test of Intelligence test may be standardized for use with individuals from the age of 2 years through adulthood (standard -Binet).

2. A list of scholastic achievement in a particular school subject may be intended for grades eight through twelve.

3. Tests of specific aptitudes similarly are designed for specified populations.

4. Rating scales, personality inventories and projective tests are intended for use with a specified segment of the total.

Types of sampling in use for constructing a psychological test

Two kinds of sampling are actually involved in constructing a psychological test (1) Sampling for gross variable and (2) Sampling for operational level.

First, the most relevant constituents of the gross variable that is the broad, comprehensive trait, must be selected. For example, gross variable is general intelligence. The constituent parts of the test may be vocabulary, verbal comprehensive, arithmetical problems, reasoning with practical problems, verbal and other analogies, perceptual organization etc.

Second, the operational level that is the actual items, must be selected example which arithmetic processes and at what levels, what kinds and which level of words, what types and ranges of situations, which perceptual figures etc.

3. Test standardization

A standardized test is one in which the procedure, apparatus and scoring have been fixed so that precisely the same test can be given at a different times and places.

According to **Anastasi,** "Standardization implies uniformity of procedure in administering and scoring the test".

According to **Freeman**, "The fundamental purpose of standardization of a psychological test is to establish its reliability and its validity at as high a level as possible.

Characteristics of Good Psychological Test

Characteristics of a good or sound psychological test are

1. Reliability
2. Validity
3. Objectivity
4. Item-analysis
5. Practicability in use and economy.

Steps involved in the construction of a psychological test

1. Planning
2. Preparation
3. Tryout
4. Evaluation

Aptitude Test

Meaning and Definition of Aptitude

Everybody possesses certain hereditary abilities, capacities and talents which help him to get success in a particular field in the future. As these abilities find suitable environment, they become refined. If these abilities and capacities of a person are educated or trained, they become refined all the more and the person succeeds in the field even more. Such hereditary abilities, capacities or talents of a person are called aptitude.

Dictionary of Education – Aptitude is defined as a "Pronounced innate capacity for or ability in a given line of endeavour such as a particular art, school subject or vocation."

Van Dusen has defined the term in a rather strict manner. "Aptitude is a measure of the probable rate of learning which results in interest and satisfaction and is relatively specific and narrow."

Traxler – "Aptitude is a present condition which is indicative of an individual's potentialities for the future."

The first thing about aptitude is that it is a hereditary ability which is developed by environment, education and training. Secondly, it indicates a person's interest, inclination and ability which is necessary for his future success in a particular field. What a child will become in future — a musician, a doctor, an engineer, or somewhat else depends on his aptitude. A child makes progress in the field for which he has aptitude. Of course, it is necessary to give him required environment, teaching and training. So, aptitude should be defined as a hereditary ability.

Nature and Characteristics of Aptitude

Bingham has clarified the nature of aptitude in the following points :

(1) Aptitude of a person is the total sum of his present qualities which points towards the future capacities.

(2) A particular aptitude of a person indicates his fitness for doing a particular task.

(3) Aptitude is an abstract noun and not a concrete object or ability, which expresses a specific quality in the whole personality of a person.

(4) Aptitude exists in the present but it indicates the future capacities.

(5) There is close relation among aptitude interest, inclination and satisfaction.

(6) Aptitudes of different people are different and they are more or less in degree.

(7) Aptitude is the chief cause of individual differences.

(8) Generally, an aptitude is stationary, very few changes can be effected in it.

Measurement

Aptitude tests have been used to measure aptitudes of a person. In the words of Freeman :

An aptitude test is one designed to measure a person's potential ability in an activity of a specialized kind and within a restricted range.
-Freeman

Aptitude tests can be broadly categorised under two heads viz.:

(i) Differential Aptitude Test Battery, and

(ii) Special Aptitude tests.

(i) Differential Aptitude Test Battery:

This is a comprehensive and carefully developed Battery. It has been developed by George K. Bennett, Harold G. Seashore and Alexander G. Wesman. It was developed principally for use in educational and vocational counseling of high school students.

It was designed for grades 8 through 12. The Battery consists of 7 subtests.

These subtests are:

(i) Verbal reasoning test,

(ii) Numerical ability test,

(iii) Abstract reasoning test,

(iv) Space relations tests,

(v) Mechanical reasoning test,

(vi) Clerical speed and accuracy test, and

(vii) Language usage test.

(ii) Special Aptitude Tests:

We shall discuss the different Special Aptitude Tests under following headings:

(A) Mechanical Aptitude Test:

Mechanical ability is an ability involved in manipulating concrete objects, such as tools, and in dealing mentally with mechanical movements.

A number of tests are available for measuring mechanical aptitude for a fairly large field of occupations rather than for a single occupation.

1. Minnesota Mechanical Assembly Test.

2. Minnesota Spatial Relations Test.

3. Minnesota Paper Form Board.

4. Johnson O'Connor's Wiggly Blocks.

5. Sharma's Mechanical Aptitude Test Battery.

6. Stenguist Mechanical Aptitude Tests.

(B) Clerical Aptitude Tests:

Different opinions are held regarding the term clerical aptitude. Super opines that it refers to the ability of routine clerical work. Bills points out clerical duties **"include the gathering, classification, and presentation of data of all sorts, and analysis and use of these data in planning, executing and determining the results of operation."**

A number of tests are available for measuring clerical aptitude:

1. Minnesota Clerical Aptitude Test.

2. General Clerical Aptitude.

3. The Detroit Clerical Aptitude Examination.

4. P.R.W. Test.

5. Orissa Test of Clerical Aptitude.

6. Clerical Aptitude Test.

(C) Tests of Artistic Aptitude:

Some tests have been devised to measure the artistic aptitudes.

Some such tests are listed below:

1. Graphic Arts Test:

These tests measure the art and aesthetic aptitudes.

2. Musical Aptitude Tests:

These tests measure the various components of musical talent.

3. Literary Aptitude Tests:

Some examples of such tests are Abbot Traube Test, Rigg Poetry Judgement Test.

(D) Professional Aptitude Tests:

These tests primarily measure aptitude for different professions. Such tests are administered before admission into professional institutions like medical, legal, engineering institutions.

There are many tests to measure aptitude in medicine, science, mathematics, law, engineering, teaching etc.

(E) Scholastic Aptitude Tests:

These tests measure the scholastic aptitudes. Some examples of such tests are Scholastic Aptitude Tests of C.E.E. Board, Graduate Record Examination.

(F) Other Tests like Motor Dexterity Tests:

Other Tests like Motor Dexterity Tests, Sensory Tests, Visual Tests and Auditory Tests.

Uses of Aptitude Tests:

(i) Aptitude tests are used for the purpose of prediction for future success both in educational and vocational careers.

(ii) These tests are used to help the students in the proper choice of courses, subjects and careers.

(iii) They are used as they supplement other psychological test.

(iv) Aptitude tests are used to help the students in the improvement of certain special traits.

v) They are used for the purpose of admission and selection of students in the college and technical institutions.

Attitude

Meaning and Definition of Attitude

It is clear that an attitude is not hereditary, a person develops it on the basis of his experiences. Another fact in this regard is that a person can form positive or negative feelings toward an object, person, activity or thought. Thus, an attitude towards an object, person,

activity or thought may be positive or negative. Thurston has defined an attitude as a measure of positive and negative thought factor.

An attitude is the degree of positive or negative effect associated with some psychological object. — **Thurston**

An attitude is a dispositional readiness to response to certain situations, persons or objects in a consistent manner, which has been learned and has become one's typical mode of response.

—**Freeman**

Nature and Characteristics of Attitude

On the basis of foregoing discussion, the nature and characteristics of attitude may be summaried as follows :

(1) Attitude is a mental talent of a person towards an object, person, activity or thought.

(2) Attitude can be either positive or negative.

(3) Attitude is related to one's own experiences.

(4) Emotions prove helpful in the development of attitude.

(5) Development of attitudes occurs due to social relations.

(6) Attitudes of a person are stable so long he does not experiences otherwise; they are subject to change on the basis of environment and experiences.

(7) Attitudes influence the personality of a person.

Types of Attitudes

Attitudes are generally classified as follows:

1. Positive Attitude :

When a person likes and accepts an object, person, activity or thought and is attracted towards it and tries to adjust himself according to that object, person, activity or thought. it is called his positive attitude towards that object, person, activity or thought.

2. Negative Attitude :

When a person dislikes and rejects an object, activity or thought and tries to make distance from that object, person, activity or thought, it is called his negative attitude towards such object, person, activity or thought.

3.General Attitude :

The attitude which is expressed in general or collective term towards an object, person, activity or thought is called general attitude, such as motherly affection towards women.

4. Specific Attitude :

The attitude which is specifically expressed towards an object, person, activity or thought is called specific attitude; such as, special love and devotion to one's own mother.

Measurement

Measurement of attitude has been undertaken since ancient times in our country, and we have been dividing it into attitude of sat (goodness), attitude' of raj (passion), and attitude of tam (darkness). However, the history of measurement of attitude towards an object, person,

activity or thought is not much long. Measuring attitude in this way was started by Thurston in 1927. Later several psychologists developed different techniques. These techniques are classified into two types— Behavioural techniques and Psychological techniques.

(I) **Behavioural Techniques**

Those techniques are included in behavioural techniques by which day-to-day behaviour of a person is observed or questions are asked directly from him to measure his attitude. These techniques are further divided into two parts :

(i) Method of Direct Observation

(ii) Method of Direct Questioning

(II) **Psychological Techniques**

(i) Thurston's Method of Paired Comparison

(ii) Method of Equal Appearing Intervals

(iii) Method of Summated Rating

(iv) Suffir's Method of Successive Difference

(v) Guttman's Scalogram Method

(vi) Scale Discrimination Technique

There are numerous types of aptitude tests. The most common among them is self-report inventory. A large number of statements are given here, and the person is directly asked what his attitude toward a particular account is. They are technically called an attitude scale.

There are two basic underlying assumptions behind all attitude scales. An individual behaviour concerning objects or events will be consistent from one situation to another. Attitude cannot be measure directly. It is inferred from the statements or actions of the person, which means it is referred from verbal and nonverbal behaviour of the subject.

Thurstone Scale and Likert Scale:

Direct statements regarding the objects of attitude are given to the examinees here.

Thurstone Scale:

In psychology and sociology, the Thurstone scale was the first formal technique to measure an attitude. It was developed by Louis Leon Thurstone in 1928, as a means of measuring attitudes towards religion. It is made up of statements about a particular issue, and each statement has a numerical value indicating how favorable or unfavorable it is judged to be. People check each of the statements to which they agree, and a mean score is computed, showing their attitude.

Following are the steps in constructing this type of test

Step 1

Several hundred statements expressing various degrees of negative and positive attitude towards the objects or events are collected. These statements can be obtained from experienced people or popular literature on the issues.

Step 2

Each statement is written on a separate slip of papers or cards. These statements are given to different judges to pile them on 11 point scale ranging from favorable to unfavorable.

Step 3

This is the scale construction stage. These statements are put into different scale points based on consensus of judge those statements where judges do not agree are eliminated here.

Step 4

Final statement for the scale is selected here. Then the subject is asked to mark those statements with which he is fully agreed, and he is then given then scored.

Likert Scale:

Definition:

A Likert Scale is a scale used to measure the attitude wherein the respondents are asked to indicate the level of agreement or disagreement with the statements related to the stimulus objects.

The Likert Scale was named after its developer, Rensis Likert. It is typically a five response category scale ranging from “strongly disagree” to “strongly agree.” The purpose of a Likert scale is to identify the attitude of people towards the given stimulus objects by asking them the extent to which they agree or disagree with them.

Often, the respondents are presented with questionnaires containing the set of statements to rate their attitude towards the objects. For example, the respondents might be asked to rate their purchase experience with shoppers stop by assigning the score as (1= strongly disagree, 2= disagree, 3= neither agree nor disagree, 4= agree, 5= strongly agree) to the series of statements given below:

The data obtained from the Likert Scale are typically treated as the interval. Thus, we can say that the Likert scale possesses description, order, and distance characteristics. Description means the unique labels or tags designated to each value of the scale; Order means the relative position of the descriptor, and Distance implies that the absolute differences between the descriptors are known and can be expressed in units.

For analysis, each statement is allotted a numerical score ranging from either 1 to 5 or -2 to +2. The analysis could be done item-wise, or a total score can be computed by summing up all the items for each respondent. One of the advantages of a Likert scale is that it is easy to construct and administer.

The major limitation of this scaling technique is that it is time-consuming and requires much more time as compared to other itemized scaling techniques. This is because each respondent is required to read every statement given in a questionnaire before assigning a numerical value to it. Another limitation of a Likert scale is that it could be misunderstood at times, especially when the responses are unfavorable.

Use, Need and Importance of Attitude Test

(1) Attitude tests measure students' attitude towards different objects, persons, activities and thoughts, etc.

(2) Attitude tests can be used to find out their favourable and unfavourable attitudes.

(3) Favourable attitudes of the students are developed.

(4) Attempt is made to diminish the unfavourable attitudes.

Personality test

Meaning and definition of personality

Etymologically the origin of Personality is Persona. It means mask used by actors on the stage. Personality covers the whole nature of an individual and hence it is very difficult to define it.

According to **Woodworth**, "Personality is the integration of all traits which determine the role and status of the Person in Society".

Allport, "Personality is the dynamic organization with in the individual of those Psycho-Physical systems that determine his unique adjustments to his environment". or "to the characteristic behaviour and thought".

According to **H.C.Warren**, "Personality is the entire organization of human being at any stage of his development".

Characteristics of personality

1. Dynamic. Personality is dynamic, ever changing and ever evolving. If personality stops growing, there has come about the end of civilization.

2. Psycho-Physical. Personality is both physical and mental i.e. outer and inner.

3. Organised and integrated. Personality is the organisation and integration of various systems.

4. Determination. Personality determines our thoughts and actions and gives direction to our specific goals.

5. Uniqueness. There is some specific feature or uniqueness in the personality of every individual.

6. Adjustment to environment. Personality is continually adjusting itself to its environment to one's inner culture.

7. Self- conscious. Personality is self-conscious. We do not attribute personality to a crow and even a child cannot be described as a personality because it has only a vague sense of personal identity. In children it is still shaping itself.

8. Social. Personality is through and through social. It develops through self-interaction. Society plays an important role in shaping the personality of an individual.

9. Personality functions as a unified whole.

Dimensions, Traits or Aspects of personality

1. Physical dimensions or traits. Physical traits include the physical body, size, shape, structure, colour, appearance, weight, voice, glandular, and nervous system.

2.Intellectual dimension or aspects. Intellectual traits of the personality are memory, imagination, observation, attention, judgement, perception, reasoning, thinking ability to make adjustments to various situations.

3. **Emotional dimensions**. Under this heading we can include the emotional reactions, temperaments like calm or excitable, cheerful or gloomy, courageous or timid, submissive or dominating, sentiments, desires, attitudes, complexes and various types of abnormalities.

4. Social traits. Social traits of personality are social behaviour, sociability, social acceptance, social effectiveness, social ideas and social adjustment.

Concept of Mature Personality

According to **G.W. Allport**, the following are the requirements of a mature personality.

1**. Extension of the self.**

The first requirement of a mature personality is an extension of the self. The developed person is the one who has a variety of autonomous interests that is he can lose himself in work, in contemplation, in recreation and in loyality to others. Egocentricity is not the mark of a personality. In a mature personality pleasures and pains of the moment set backs

and defeats, and the impulse for self-justification fade into the background, so that they do not obscure the chosen goals. These goals represent an extension of the self.

2. Self-objectification.

The second requirement of mature personality is self objectification. It is a peculiar detachment of the mature person when he surveys his own pretensions in relation to his abilities, his present objectives in relation to possible objectives for himself, his own equipment in comparison with the equipment of others, and his opinion of himself in relation to the opinions others hold of him. The capacity for self-objectification is insight. The insight is bound with sense of humour.

3. Unifying Philosophy of life

Unifying Philosophy of life is another requirement of a mature personality. A mature personality participants and reflects, lives and laughs, according to some embracing philosophy of life, developed in his own satisfaction and representing to himself his place in the scheme of things, gradually developing from the stage.

Integrated personality

An integrated personality is one in whom the various aspects of personality (physical, intellectual, emotional and social) are working in a harmonious and effective manner. In such a person, his ambitions and aspirations are in accord with his mental capacities and his objectives are realistically attuned to the required physical energy involved in personal accomplishment. In other words, there are two marks of Integra personality.

1. Balance between mental processes.

2. Harmonious adjustment to social environment.

Assessment of personality

As is clear from the definitions of the personality, personality is not made up of any one trait or the other. It is the one dynamic organisation of innate and acquired traits which lead to the individuals unique adjustment to his environment.

Therefore, to assess personality means to know everything about the person. It will include his mental, physical, emotional and social behaviour. It will include both conscious and unconscious mind. For this purpose, only one test cannot justify to assess the whole personality. There are three main kinds of techniques.

1. Subjective techniques
2. Objective techniques
3. Projective techniques

INTELLIGENCE

Intelligence means the manner with which an individual deals with facts and situations. Intelligence is the aggregate or the global capacity of the individual to act purposefully to think rationally and to deal effectively with the environment.

Definitions of Intelligence

1. **Woodworth and Marquis**. "Intelligence means intellect put to use. It is the use of intellectual abilities for handling a situation or accomplishing any task".

2**. Terman**. "An individual is intelligent in proportion as he is able to carry on abstract thinking".

3. **Wagon**. "Intelligence is the capacity to learn and adjust to relatively new and changing conditions".

4.Jean piaget. "Intelligence is the ability to adapt to one's surroundings. Intelligence is the individual's capability by means of which the individual respond to the environment.

Intelligence is an ability which integrate many skills like adjustments, learning, abstract thinking and comprehension".

NATURE OF INTELLIGENCE

People do well academically or succeed at intellectual tasks, we consider them to be intelligent. Intelligence is a person's ability to learn and remember information and recognize concepts and their relations and to apply the information and recognition by behaving in an adaptive way.

Intelligence is

1. **Ability to adjust**- Ability of an individual to direct his behavior towards a goal.

2. **Ability to learn**- Intelligence is the learning ability.

3. **Ability to do abstract reasoning**- An individual is intelligent in proportion as he is able to carry on abstract thinking.

Characteristics of intelligent person

1. Learns fast and with ease

2. Good adjustment in behavior and capacity to solve problems.

3. He is capable of successfully completing difficult and complex task with high efficiency.

4. He perceives the appropriate solutions exist among objects or ideas.

5. He generally acts in pursuit at the welfare of the society.

The Concept of Mental Age

Binet conceived the idea of mental age to measure intelligence. Tests were made for different age levels.(If a test was passed by 60-70% of children of a given age, Binet accepted it for the age level). A child who successfully perform all the tasks thus accepted for the five years old. Child is taken to have a mental age of five; whatever his chronological age may be. If he is five years old chronologically to he is rated as normal or average. A seven year old

child having a. M.A of six is considered as mentally deficient, but if he has an M.A of eight then he is aid to be above average in intelligence.

Intelligence Quotient (I, Q)

Intelligence Quotient is an index of intelligence. It is comparative index showing how a particularly child performs when compared with others of the same chronological age. It is computed as a ratio between Mental age and chorological age

IQ = $\left(\frac{MA}{CA}\right)$X100

To avoid fractions we multiply the result with 100.

IQ value 100 refers to normal intelligence values above 100 indicate high intelligence and value blow 100 refers low level of intelligence.

Classification of intelligence

I.Q - 140	and above Genius
120 – 139	very superior
110 – 119	superior
90 – 109	average
80 – 89	dull
70 -79	border line
Below 70	feeble minded
50 – 59	Morons
25 – 49	Imbeciles
Below 25	Idiots

Classification of intelligence tests

Intelligence tests are of various kinds. They have been classified from different points of view.

2. Classification on the basis of the number that take the test at a time.

On the basis of the mode administration intelligence tests have been divided into two classes

Individual tests

These tests are administered to one individual at a time. They include performance test as well as tests that require linguistic ability. These tests are particularly appropriate for testing individuals, but they are not financially feasible as they are expensive. Intelligence test are time consuming, Administration of such tests require more expert training. Example Stanford Binet intelligence test, Wechsler Bellrnre Intelligence Test, Koh's Block Design test etc.,

Group test

In this category the tests are administered upon a large number of examinees simultaneously. Group test are more objective and less expensive. But in these test no rapport between the subjects and examines. These tests are more useful for the purpose of educational and vocational guidance.

Eg., Army Alpha Test, Army Beta Test.

3. Classification based on the medium used

In terms of medium used, intelligence tests are divided into two categories

1. Verbal tests:-

In this questions are asked in oral or in written form of a given language. It cannot be used for testing individuals who lack language ability and small children. It is less expensive, standardization, administration and scoring are easy.

2. Nonverbal test:-

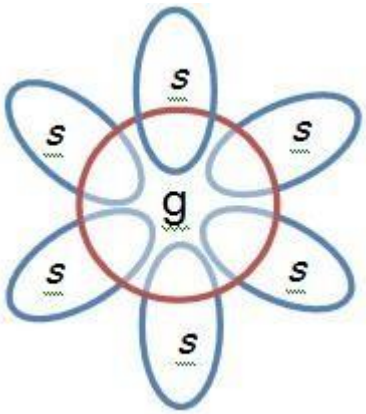
These test do not require the use of language but could be manipulated using figures, pictures, blocks or even other objects. It is more expensive. These tests can be used for testing subjects who lack language ability.

Two factor theory Spearman

This theory was advocated by Spearman in 1904. According to him every different intellectual activity involves a general factor ‘g’ which is shared with all the intellectual activities and specific factor ‘s’ shares which it shares with none. G belongs to the general intelligence which is common thread that run through all tasks that the individual performs through the course of his life. The amount of this g depends on the amount of cerebral cortex energy present. The individual uses this energy depends upon the environment, education, motivation and friend circle etc.

Apart from general factor ‘g’ an individual may have a specific ability to deal with a particular situation/ problem. This specific ability can be given as s1, s2, s3…etc. The ability of intelligence “A” of an individual can be expressed by following equation

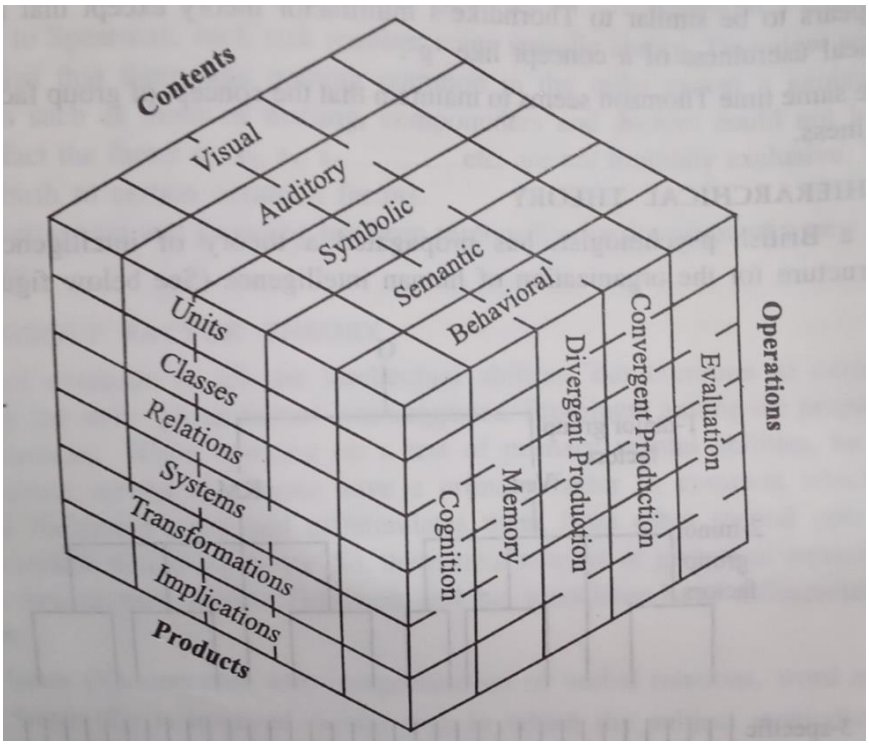
g + s1 + s2 + s3+….. = A



For example, an individual who is an IIT graduate could also be a good writer. An Engineer has a general factor intelligence ‘g’ and a specific ability ‘s1’ to crack the CET. An astronaut could have specific singing ability. Astronaut has a general factor intelligence ‘c’ and specific ability ‘s2’ for singing.

According to Spearman, each task requires some specific ability. This view was not proper as it implied that there was nothing common in the task except a general factor and profession such as those of nursing, compounders, doctors could not be put in one group. In fact, the factor s1, s2, s3….. etc. are not mutually exclusive. They overlap and give birth to certain common factors.

GUILFORD'S STRUCTURE OF INTELLIGENCE:



Guilford's structure of intellect model is a multiple intelligence theory. He believed that intelligence wasn't a monolithic, global attribute but a combination of multiple abilities, that were relatively independently. He applied factor analytical method to learn these mental abilities. According to him intelligence consists of several, intellectual abilities It was developed in 1956 (first version), intelligence as a collection of abilities or functions for processing different kinds of items of information in several ways.

Guilford's model included five types of mental contents, they are figural, symbolic, semantic and behavioral (1977). This dimension contains the broad areas of information in which operation are applied. It has divided in to five categories.

1. Figural

a. Visual =information arising from stimulation on the retina in the form of an image.

b. Auditory = information arising from stimulation of the cochlea of the ear as an image.

2. Symbolic- Information is perceived as symbols or signs that have no meaning by themselves. For example: the letters of an alphabet.

3. Semantic- Information perceived in words or sentences, whether oral, written or silently in one's mind.

4. Behavioral- As an act of an individual/individual's information is perceived.

OPERATIONS DIMENTIONS

This dimensions contains five kinds of operations or general intellectual process.

1. Cognition – Ability to understand, comprehend, discover, and become aware.

2. Memory – Ability to memorize information.

3. Evaluation – Process of judging whether an answer is accurate, consistent or valid.

4. Divergent production – Process of producing multiple solutions to a problem.

5. Convergent Production – Process of concluding a single solution to a problem.

PRODUCTS DIMENSION

This dimension contains result of applying particular operations to specific content. They are six types of products

1. Unit – Represents a single item of information.

2. Class – A set of items that share some attributes.

3. Relation – Represents a connection between items or variables; might be linked as opposites or in associations, sequences, or analogies.

4. System - An organization of items or networks with interacting parts.

5. Transformation – Change's perspectives, conversions, or mutation to knowledge; for example, reversing the order of letters in a word.

6. Implication – Predications, inferences, consequences, or anticipations of knowledge.

This model proposes that intelligence consists of 150 independent abilities that results from the interaction of five types of operations and six types of products. Guilford, 1982.

In this way, according to Guilford's model of intellect there are 150 factors operating in once intelligence. Each one of these factors has a trigram symbol, i.e. at least one factor from each category of three parameters has to be present in any specific intellectual activity of mental task.

THURSTON'S GROUP FACTOR THEORY

Louis Leon Thurston proposed his multiple- factor theory of intelligence in 1938. His multifactor theory of intelligence is based on the generalized concepts, formulas, and methods used by Spearman in his two –factor theory of intelligence. He stated that every individual possesses different levels of seven primary mental abilities, i.e. Word fluency, verbal factor, numerical factor, inductive reasoning, memory, spatial visualization and perceptual speed and these levels do not depend on each other and each of these abilities can be evaluated separately.

1. The Numerical factor

It involves the ability of an individual to do quick and accurate numerical computations. It can be measured by checking the accuracy and speed of person in solving various arithmetic problems.

2. The Verbal factor

It refers to the ability of the person to understand and use various words, sentences, language, or other verbal content pieces. This ability can be assessed through vocabulary tests, jumble word tests and verbal or reading comprehensions tests.

3. The Space factor

It refers to the spatial visualization of the person. This ability comes into play when the person tries to understand the manipulation of various real/imaginary objects in space. The test that measures this ability involves solving various kinds of puzzles, understanding various geometric figures, and identifying the correct mirror image of the object, or choosing the correct image of the object when it is rotated by different angles.

4. Memory

It refers to the ability of the person to quickly memorize the various concepts or phenomena and retaining them for a longer period of time. Various memory tests like asking the participants to learn nonsense syllables, and their ability to remember them could a measure of their memory factor. The ability to recall the learned concepts by the students during the exams is largely depend on this primary mental ability factor.

5. The verbal fluency / Word fluency factor

When a person is asked to rapidly speak several isolated words or sentences, then verbal fluency comes in to play; a person with high verbal fluency may excel in this task, while the person with low verbal fluency may have difficulty in this task. This factor is responsible for the communication skills of the person. The tests to measure this factor may involve asking participants to rapidly think of words that begin or end with a specific letter.

6. The Inductive Reasoning factor

This reasoning involves the ability to deduce a general principle from a specific concept. This ability is measured through various tests like number series, words series and classification of words or numbers. The inductive reasoning tests may involve selecting an

appropriate number or image according to the sequential order of the given numbers or image series.

7. Perceptual speed factor

It involves the ability of a person to rapidly recognize and compare the specific image, numbers, or letters and to accurately proofread various types of content. Tests like picture recognition, rapidly crossing specific letters from the series of numbers, and finding particular words in the paragraphs are used to measure the perceptual speed factor of the person.

8. The Deductive Reasoning factor

It involves the ability to accurately understand a specific phenomenon or concept from the generalized principle. Various aptitude test are available to test the deductive reasoning of a person that includes the various set of statements, and person has to choose the best possible logical solution according to the given statements.

9. Problem solving ability factor

It refers to the ability of a person to solve various general of aptitude problems. The factor can be assessed by analyzing the individual's responses to various hypothetical problem and their ability to reach the conclusion.

Thorndike's Multi-factor Theory

Some psychologists believe that several kinds of intelligence may be distinguished from one another. E.L.Thorndike is considered the pioneer of this theory who distinguished three kinds of intelligent activity.

i) Social Intelligence. Or ability to understand and deal with persons.

ii) Concrete Intelligence. Or ability to understand and deal with other things.

iii) Abstract Intelligence. Or ability to understand and deal with verbal and mathematical symbols

According to **Freeman**, “The merit of this classification of types of intelligent activity, for psychological testing and diagonis is that it indicates several realms in which persons might be functioning and implies that separate and sufficiently specialized tests might be devised to measure how effectively persons are functioning in each”.

Freeman further says, “Thorndike’s multifactor theory of intelligence is at one extreme of the interpretations regarding the nature of mental organization. As the name of the theory indicates, intelligence is said to be constituted of a multitude of several factors, or elements, each one being a minute element of ability. According to this throry, there is no such factor, as general intelligence, there are only many high specific acts, the number of such depending upon how refined a classification we might wish to make and are capable of making”.

Uses of Intelligence Tests:

(i) The classification of the students on the basis of intelligence test scores.

(ii) Measurement of the learning readiness of students at different age levels.

(iii) Selection of subjects, courses and careers.

(iv) Diagnosis of reading inability and educational backwardness.

(v) Prediction of future progress of a student.

(vi) Selection of candidates for officers training in defense services.

(vii) Ascertaining the magnitude of individual difference.

(viii) Prediction for vocational success of student in occupational life.

(ix) For the preparation of case study report.

(x) The comparative study of the students may also be made.

(xi) Providing educational, vocational and personal guidance to the students.

UNIT IV

STATISTICAL CONCEPTS

Meaning of Z-Score

Z-scores are those converted scores of raw scores of which the mean (M) is zero (0) and standard deviation (σ) is one (1). These scores are obtained by linear transformation of raw scores, so they fall in the category of linear standard scores. The unit of Z-scores is similar to the standard deviation (σ). Its value is generally from –3 σ to 3 σ + . The positive (+ve) sign of a Z-score of a raw score indicates that it is more than the Mean (M) of the raw score; and the negative (–ve) sign indicates that it is less than the mean (M) of the raw score.

Calculation of Z-Scores

The following formula is used for converting raw scores into Z-scores :

Z-score, $\mathbf{Z = (X – M)/ \sigma}$

In which, Z = Z-score

X = Raw score

M = Mean

σ = Standard deviation

Example 1

In a test, the mean (M) of scores is 65 and the standard deviation (σ) is 10. In this test, Student A has obtained 90 marks and Student B, 35. Convert the scores of students A and B into Z-scores. **Soln**

Z scores of Student A, $Z = (X – M) / \sigma$

$= (90–65) / 10$

$= 25 / 10$

$= 2.5$

Z scores of Student B, $Z = (X - M) / \sigma$

$= (35–65) / 10$

$= (–25) / 10$

$= – 3$

If raw scores are to be found out from Z-scores, the following formula is used :

Raw scores, $X = M + (Z \times \sigma)$

Example 2

In a test, the mean (M) of the scores is 65, and standard deviation (σ) is 10. In this test, the Z score of Student A is 2.5 and that of Student B, -3. Find out raw scores of students A and B.

Soln

Raw score of Student A, $X = M + (Z \times \sigma)$

$= 65 + (2.5 \times 10)$

$= 65 + 25$

$= 90$

Raw score of Student B, $(X) = 65 + (– 3 \times 10)$

$= 65 – 30$

$= 35$

Utility of Z-Scores

In the field of education, the Z-scores are used in meaningful analysis of scores of a student, comparison of scores of two students and for comparison of marks of a student in-two subjects, Generally, the meaning of Z-scores is considered as follows :

Table

Meaning of Z-scores

Z-Score	Meaning
More than + 2.00	very good
+ 1.00 to + 2.00	good
+ 0.50 to + 1.00	above average
– 0.50 to + 0.50	average
– 1.00 to – 0.50	below average
– 2.00 to – 1.00	low
– 2.00	lowest

In comparison two qualities or abilities of a student in the reference group, the Z-scores of the scores in the two subjects are calculated and then they are compared.

Example 3

The marks obtained by Student A in the mathematics and language tests of maximum marks 100, each are 80 and 70 respectively. If the mean (M) of mathematics scores of the class students is 60 and standard deviation () σ 8, and in the language test 40 and 10 respectively; then find out in which subject the student A is more able as compared to other students.

Soln

Z-score of mathematics marks of Student A, $Z = (X – M) / \sigma$

$= (80–60) / 8$

$= 20/ 8$

$= 2.5$

Z-score of language marks of Student A, $Z = (X – M) / \sigma$

$= (70–40) / 10$

= 30 / 10

= 3

Analysis

Because Student A's Z-score in language is more than that of mathematics, so he is more able in language in the class as compared with mathematics.

The first thing to be seen here is that his obtained score in mathematics is more than that of language, but Z-scores reveal that he is more able in language than mathematics in the class. Secondly, the table reveals that the student falls in Very Good category in both subjects. However, Z-scores tell that despite being Very Good in both subjects, he is comparatively better placed in language.

Z-scores are both positive and negative and they are calculated to two numerals of fraction, so they are a little difficult to use. However, from the viewpoint of the above qualities and utility, they are most used in the field of education.

T-Scores (Transformed Scores)

Meaning of T-scores

T-scores are those converted scores of raw scores of which the mean (M) is 50, standard deviation (σ) is 10 and the distribution is normal. These scores are always positive (+ve) and their value is often from 20 to 80.

Calculation of T-Scores

The following formula is used for converting raw scores into T-scores :

T-Score, $T = 50 + 10\left(\frac{X-M}{\sigma}\right)$

In which, T = T-score

X = Raw score

M = Mean

σ = Standard deviation

And if Z-score of the raw score is known, the following formula is used :

T-score, $T = 50 + 10Z$

Example 4

Look at example 3. The score in mathematics = 80 and Z-score = 2.5; and the score in language = 70 and Z-score = 3. Convert these scores into T-scores.

Soln

T-score of mathematics raw score 80, $(T) = 50 + 10Z$

$= 50 + 10 \times 2.5$

$= 50 + 25$

$= 75$

And, T-score of language raw score 70, $(T) = 50 + 10Z$

$= 50\ 10\ 3 +\times$

$= 50 + 30$

$= 80$

It is clear that the student A is more able in language as compared with mathematics.

Utility of T-Scores

The mean (M) of T-scores is 50 and standard deviation (σ) is 10. so, if the range of general distribution is taken to be 100, the range of T-scale becomes from 0 to 100 as a result, measurement of scores can be done more accurately. The meaning of T-scores is derived from the following table :

Table

Meaning of Various T-Scores

T-Scores	Standard Deviation position	Percentage of Students securing low score	Percentage of Students securing high scores
80	+ 3.0 σ	99.87	0.13
75	+ 2.5 σ	99.38	0.62
70	+ 2.0 σ	97.72	2.28
65	+ 1.5 σ	93.32	6.68
60	+ 1.0 σ	84.13	15.87
55	+ 0.5 σ	69.15	30.85
50	0.1 σ	50.00	50.00
45	– 0.5 σ	30.85	69.15
40	– 1.0 σ	15.87	84.13
35	– 1.5 σ	6.68	93.32
30	– 2.0 σ	2.28	97.72
25	– 2.5 σ	0.62	99.38
20	– 3.0 σ	0.13	99.87

T-scores are generally large, so they are not suitable for statistical calculations. They are least used in the field of education.

Percentile Scores

A percentile score is another type of converted score. Your raw score is converted to a number indicating the percentage of the norm group who scored below you. For example, a score at the 60th percentile" means that the individual's score is the same as or higher than the scores of 60% of those who took the test. The 50th percentile is known as the median and represents the middle score of the distribution.

Percentiles can not be averaged nor treated in any other way mathematically. However they do have the advantage of being easily understood and can be very useful when giving feedback to candidates or reporting results to managers.

If you know your percentile score then you know how it compares with others in the norm group. For example, if you scored at the 70th percentile, then this means that you scored the same or better than 70% of the individuals in the norm group.

This is the score most often used by organizations when comparing your score with that of other candidates because they are so easily understood they are very widely used when reporting results to managers.

The characteristic way that test scores tend to bunch up around the average and the use of percentiles in the interpretation of test results, has important implications for you as a job candidate. This is because most aptitude tests have relatively few questions and most of the scores are clustered around the mean. The effect of this is that a very small improvement in your actual score will make a very substantial difference to your percentile score.

To illustrate this point, consider a typical aptitude test consisting of 50 questions. Most of the candidates, who are a fairly similar group in terms of their educational background and achievements, will score around 40. Some will score a few less and some a few more. It is very unlikely that any of them will score less than 35 or more than 45. Looking at these results in terms of percentiles is a very poor way of analyzing them and no experienced statistician would ever use percentiles on this type of data. However, nine times out of ten this is exactly what happens to these test results and a difference of three or four extra marks can take you from the 30th to the 70th percentile. This is why preparing for these test is so worthwhile as even small improvements in your results can make you appear far superior candidate.

Qualitative Interpretation of Test Scores

Assume that three boys, all of the same age, have been tested. Suppose that their intelligence quotients are 50, 100, and 150. Since these are numerical ratios (MA/CA × 100), it is natural to assume that they have a quantitative significance. So they do for they indicate rate of mental development. But these quotients also have a qualitative significance for, among other things, they indicate each boy's position in the "hierarchy of intelligence." If the measure of intelligence is valid, the boy having the IQ of 50 is seriously retarded and is in the lowest one percent of the population in respect to the psychological functions being tested; the boy with the IQ of 100 is the typical or average individual, midway up (or down) in the distribution of intelligence; and the boy having the IQ of 150 is very superior and belongs in the top percentile rank of the group.

Qualitative significance of the intelligence quotient can be illustrated further by asking this question : Is the brightest of these three boys one and one-half times as intelligent as the average boy, and three times as intelligent as the retarded one ? This question cannot be answered in terms of numbers; it is impossible to say how many "times" more capable or less capable one is than the others, because the IQ is not a percent. But each of these quotients has certain connotations. In this example, the qualified school or clinical psychologist will be able to draw important inferences from each boy's IQ regarding rate and quality of school leveling, extent and level of educability, vocational possibilities and levels, and probable types of interests.

The boy with an IQ of 50 probably will not be able to complete more than the second grade; the boy having the IQ of 100 should be able to complete twelve grades; the boy with an IQ of 150 will be able to progress in education as far as his interests and motives indicate. Obviously, too, the kinds of occupations that will be open to the first boy are very limited; those open to the second will be numerous; those open to the third will be practically

unrestricted, so far as mental capacity is concerned. And the same may be said of the range of interests in general that will be within the scope of each. These facts are of educational and clinical significance, but at present there are no psychological or statistical means whereby one can calculate how many times more or less capable one person is than another.

Caution is necessary at this point. The inferences drawn in the preceding paragraph cannot be based solely upon the numerical IQ value without reference to the clinical features in the test performances or other factors not shown by the numerical index. We have assumed that there are no complicating factors and that the IQs are valid measures of the capacities and performances of the three boys. The boy with 150 IQ, however, might be an unstable personality who is failing in many or all of his school subjects. The boy with 100 IQ might have been penalized on the test by a language handicap. And the boy with an IQ of 50 might show a "scatter" (inconsistency and variation) of performance indicating emotional disturbance rather than intellectual impoverishment. Occasionally, also, it will be found that a high test rating may be attributable to an inconsistently high level of performance on one or a few types of subtests (for example, memory span or word knowledge), just as, conversely, it occasionally happens that a person's IQ is depressed by an inconsistently low performance on one or a few subtests. "Inconsistent" means that the individual's levels of performance on these few subtests differ markedly, in one direction or another, from the general and more consistent levels of his scores on the other subtests.

It is to be noted that the possible vitiating factors mentioned in the preceding paragraph are of the type to which the experienced and qualified psychological examiner will be alert. These precautions do not signify that all or most intelligence test ratings are affected by these and other contingencies.

CORRELATION

Correlation is a statistical technique that can show whether and how strongly pairs of variables are related. For example, height and weight are related; taller people tend to be heavier than shorter people.

Positive Correlation

A correlation in the same direction is called a positive correlation. If one variable increases the other also increases and when one variable decreases the other also decreases. For example, the length of an iron bar will increase as the temperature increases.

Negative Correlation

Correlation in the opposite direction is called a negative correlation. Here if one variable increases the other decreases and vice versa. For example, the volume of gas will decrease as the pressure increases, or the demand for a particular commodity increases as the price of such commodity decreases.

Karl Pearson's correlation method:

Pearson's correlation is the most widely used correlation statistic to measure the degree of the relationship between linearly related variables. The following formula is used to calculate the Pearson r correlation: $r = \frac{N\sum xy - \sum x \sum y}{\sqrt{[N\sum x^2 - \frac{(\sum x)^2}{N}/N][\sum y^2 - \frac{(\sum y)^2}{N}]}}$

r = Pearson r correlation coefficient

N = number of value in each data set

$\sum xy$ = $sum\ of\ the\ products\ of\ paired\ scores$

$\sum x$ = $sum\ of$ x scores

$\sum y$ = $sum\ of$ y scores $\sum x2$ = $sum\ of\ squared\ x\ scores$ $\sum y2$ = sum of squared y scores

Problem: Find the correlation between the two sets of variables suing Karl Pearson's method.

x	y
24	13
20	19
8	12
17	20
15	11
12	16
10	5
8	2
6	7
4	1

Solution

x	y	X^2	Y^2	xy
24	13	576	169	312
20	9	400	81	180
8	12	324	144	216
17	20	289	400	340

15	11	225	121	165
12	16	144	256	192
10	5	100	25	50
8	2	64	4	16
6	7	36	49	42
4	1	16	1	4
$\Sigma x = 134$	$\Sigma y = 96$	$\Sigma x^2 = 2174$	$\Sigma y^2 = 1250$	$\Sigma xy = 1517$

Spearman's Rank Correlation

The presence of outliers gives a distorted picture of the association between two random variables, the Spearman's rank correlation is a non-parametric test that can be used instead of the Pearson's correlation coefficient.

The data used must be ordinal, interval or ratio

+1 = a perfect positive correlation between ranks

- 1 = a perfect negative correlation between ranks

0 = no correlation between ranks

The following formula can be used to calculate this coefficient,

$$\rho = 1 - \left(\frac{6\,(\sum D^2 + C\,)}{N(N^2 - 1)} \right)$$

Where: ΣD^2 is the sum of the squared differences between the pairs of ranks and n is the number of pairs.

Problem:

Find the relationship between the two sets of variables given below using Spearman's rank correlation method.

x	24	20	8	17	15	12	10	8	6	4
y	13	9	12	20	11	16	5	2	7	1

iii) It is used in regression equations and factor analysis

Solution

x	y	R_1	R_2	$D(R_1-R_2)$	D^2
24	13	1	3	-2	4
20	9	2	6	-4	16
8	12	7.5	4	3.5	12.25
17	20	3	1	2	4
15	11	4	5	-1	1
12	16	5	2	3	9
10	5	6	8	-2	4
8	2	7.5	9	-1.5	2.25
6	7	9	7	2	4
4	1	10	10	0	0
					$\Sigma D^2 = \mathbf{56.5}$

$$\rho = 1 - \left(\frac{6\,(\sum D^2 + C\,)}{N(N^2-1)}\right)$$

$= 1 - \left(\frac{6(56.5)}{10(10^2-1)}\right)$

$= \frac{339}{990}$

$= 0.342$

UNIT V

NEW TRENDS IN EVALUATION

GRADING

Grading in education is the process of applying standardized measurements of varying levels of achievement in a course. Another way the grade point average (GPA) can be determined is through extra-curricular activities. Grades can be assigned as letters (generally A through F), as a range (for example 1 to 6), as a percentage of a total number of questions answered correctly, or as a number out of a possible total (for example out of 20 or 100).

In some countries, all grades from all current classes are averaged to create a Grade Point Average (GPA) for the marking period. The GPA is calculated by taking the number of grade points a student earned in a given period of time of middle school through high school.[1] GPAs are also calculated for undergraduate and graduate students in most universities. The GPA can be used by potential employers or educational institutions to assess and compare applicants. A cumulative grade point average is a calculation of the average of all of a student's total earned points divided by the possible amount of points. This grading system calculates for all of his or her complete education career.

Evaluation and Grading System

Grades and the Basis for Assessment''

Letter Grades and Grading Standards

“A” (80-100%)

B” (70-79%)

“C” (60-69%)

“D” (50-59%)

“F” (0-49%)

"A" indicates Exceptional Performance: comprehensive in-depth knowledge of the principles and materials treated in the course, fluency in communicating that knowledge and independence in applying material and principles.

"B" indicates Good Performance: thorough understanding of the breadth of materials and principles treated in the course and ability to apply and communicate that understanding effectively.

"C" indicates Satisfactory Performance: basic understanding of the breadth of principles and materials treated in the course and an ability to apply and communicate that understanding competently.

"D" indicates Minimally Competent Performance: adequate understanding of most principles and materials treated in the course, but significant weakness in some areas and in the ability to apply and communicate that understanding.

"F" indicates Failure: inadequate or fragmentary knowledge of the principles and materials treated in the course or failure to complete the work required in the course.

"I" indicates Incomplete.

"W" indicates Withdrawal with permission.

Aegrotat Standing

Aegrotat standing (credit granted with incomplete course work) will be considered only in exceptional circumstances (usually only in cases of very serious illness) and if term work has been of high quality.

Determination of Final Grades

a. A student's grade in each course will be based upon the year's work and the final examination (if required);

b. The instructor will discuss with the class the basis for assessment specifying the relative weight of each examination, test, in-class activity and written assignment;

c. The instructor will also specify which assignments must be completed in order to receive a grade in the course; penalties for late assignments should be clearly stated in each course;

d. The method of determining final grades is to be discussed with students within the time frame Senate has approved for late course registration;

e. The instructor must inform students of their standing prior to the date for honourable withdrawal from the course. If no written term work has been evaluated by that date, the information shall be given in the form of a written statement of the student's standing;

f. The instructor is required to return to students all written work, other than final examinations, which has been submitted for evaluation purposes. Students may discuss with their instructor the work presented, the comments made, and the grade assigned;

g. Final evaluation submissions are not returned to students but are kept on file by the Office of the Registrar for six months after the publication of Grade Reports;

h. For each course, faculty will enter their final grades into the Student Information System (SIS) within seven (7) days of the date a final exam was written for an on-campus course, and within fourteen (14) days of the date the final exam was written for alternative delivery courses. For courses where no final exam was scheduled, final grades must be submitted within seven (7) days of the last day of classes in the term. Once the Dean has approved the marks submission, grades will be available for viewing on the SIS. The Dean's signature of approval indicates that the marks submission is consistent with existing practices and policies of the Faculty. Revisions to any previously assigned grade are submitted in writing for the approval of the Dean, together with the reasons for such revisions. Grades are not official until they have been approved by the Dean and released by the Office of the Registrar;

i. The final marks issued by the University are the only ones accepted as binding.

Incomplete Grades

In exceptional circumstances, a student may request consideration for an Incomplete grade (I). Such request must be submitted, through the Instructor to the Dean, together with the reasons for the request.

Any student assigned an Incomplete grade must normally complete all course requirements within 30 days after the end of the examination period or the last day of the course if there is no final examination for that course. If after that period the course is not completed, a grade of zero will be assigned to those components not completed.

Semester System

Generally, a session of a class in our country is of one year duration and the examination is conducted at its end. Some universities divide the session into two sessions of six months each, and conduct the examinations at the end of each six months session. Thus, a three-year degree course is divided into six equal sessions, and two-year postgraduate course is divided into four sessions. Such a system is called **semester system**. In this system, a student failing in one subject in one semester is not declared to be failed, rather is admitted to the next semester and is given an opportunity.

Merits of Semester System

This system has proved effective to some degrees in the field of higher education. This system has the following merits :

(1) The chief merit of this system is that no student (examinee) is declared to be failed in it. A whole year is not lost for the failure candidate to appear again at the examination, and there comes a difference of only six months if such incident occurs in the final semester. It eradicates stagnation in the field of education.

(2) The second merit of this system is that the study work goes on in the summer vacations. The teaching of those students continues in the summer vacations who have failed in a subject in the previous semesters.

(3) The third merit of this system is that the students have to complete limited and comparatively less course in a semester and their attention is turned towards it, so they are able to understand and assimilate it better.

(4) The students attend the classes in this system regularly and involve themselves in the studies continuously, and the problem of indiscipline in students is diminished.

(5) The teachers too have to work hard throughout the year, they are duty bound to complete the task within the limited time

Demerits of Annual System

The semester system has its own limitations and demerits :

(1) The semester system can only be used in the field of higher education where the number of students is limited. This system cannot be used at primary and secondary levels.

(2) It is a difficult proposition to determine curriculum of each semester and administer examinations at the end of each semester.

(3) In this system, the number of students re-appearing at the examinations with a desire to either clear the previous paper or papers or to obtain more marks in a particular paper or papers is comparatively large. To make arrangements for them in the examination programme and to prepare their results according to the new results is a difficult and complex task

(4) The students are always under the hammer of examinations.

Annual System VS Semester System

Comparison between semester and annual systems is very often done. Both the systems have its merits and demerits. Annual system is the traditional system. Annual system

covers more syllabuses at a stretch and compels the student to remember all this till the end of the year. Sometimes two or more topics will be included in the same paper (very often, a paper will have to be set by two examiners under such situations), when specializations are there. Otherwise, certain topics will be omitted and the syllabus diluted. Since at the end of the year only the public examinations are conducted University gets enough time to prepare question papers and value answer papers. Number of examiners and examinations also can be reduced, which become more economical for Universities. Results can be announced in time and the schedule can be kept.

In semester system, the students get more advantage; since examinations are held within months what is studied will remain afresh in their brain). The syllabus load also will be less. Different topics need not be combined in the same paper. Students get more chances to improve also.

Since examinations come within a few months student unrest also will be less in a semester system. There were many challenges to be faced by the under graduate colleges in preparing the students for the semester system. The semester system is a very proactive system as it engages both the faculty and the students throughout the year in academic activity. While, in the annual system once the student enters the college he feels free and thinks about studying only during the exam time. Semester system not only involves students more throughout the year but also reduces examination burden. The semester system is the need of hour and a very effective one. The semester system allows greater interaction with teachers and the students will be more focused on preparing throughout the year.

CONTINUOUS AND COMPREHENSIVE EVALUATION (CCE)

C–Continuous

C–Comprehensive

E–Evaluation

Continuous stands for assessment of a student throughout the year, not just at the end of a term. It may be done formally or in an informal way using different techniques of evaluation. Comprehensive takes care of assessment of all round development of a child's personality. A child will be assessed not only in terms of his knowledge about a subject but his participation in other activities also. Broadly, we assess a child's growth in two areas – Scholastic and Co-scholastic. The term Scholastic refers to those aspects, which are related to intellect or the brain. It is related to the assessment of learners in curricular subjects. It includes assignments, projects, practical etc.

Continuous and Comprehensive Evaluation (CCE) refers to a system of school-based evaluation of students that covers all aspects of students' development. It is a developmental process of assessment which emphasizes on two fold objectives. These objectives are continuity in evaluation and assessment of broad based learning and behaviourial outcomes on the other.

In this scheme the term `**continuous**' is meant to emphasise that evaluation of identified aspects of students' `growth and development' is a continuous process rather than an event, built into the total teaching-learning process and spread over the entire span of academic session. It means regularity of assessment, frequency of unit testing, diagnosis of learning gaps, use of corrective measures, retesting and for their self -evaluation.

The second term `**comprehensive**' means that the scheme attempts to cover both the scholastic and the co-scholastic aspects of students' growth and development

Need of continuous comprehensive evaluation

Continuous and Comprehensive Evaluation is intended to provide a holistic profile of the learner through assessment of both scholastic and non-scholastic aspects of education spread over the total span of instructional time in schools.

It helps to identify those positive attributes of the learner which are not usually assessed during the examinations conducted by the Board

As it is spread over a period of two years in class IX and X it provides several opportunities for the school to identify the latent talents of the learners in different contexts.

This document is supportive to the statement of marks issued by the Board after the examinations conducted by it.

Essential aspects of continuous comprehensive evaluation

1. To provide a holistic profile of the learner through assessment of both scholastic and nonscholastic aspects of education

2. To identify the latent talents of the learners in different contexts.

3. To identify strategies for raising Student Achievement

4. To plan a Comprehensive Evaluation Program to Improving Schools

5. To suggest suitable tools and techniques for achieving continuous comprehensive evaluation.

6. Use Evaluation for Continuous School Improvement

7. Using evaluation as a tool for continuous improvement of the school and the students.

8. To suggest ways of strategies of sensitizing school administrators.

Question Bank

Meaning of Question Bank

Question banks are large database of suitable questions that are coded by subject area, instructional level, instructional objectives measured, and various other pertinent question characteristics (e.g. difficulty level and discriminating powers). Questions in the question banks are often called ‘items’.

Planning a Question Bank

Planning for a question bank involves defining processes for preparation of individuals, preparatory work for the question bank and identifying what has to be established with the question bank. One should be prepared to answer technical questions that might arise. Computer expertise is an essential requirement. One should be capable of modifying computer programs, establishing a database system, and capable of running packaged programs. For planning a question bank, evaluation pattern of the program has to be specified e.g. details about courses and their contents, objectives of assessment, weightage of internal and external assessment, conflation and reporting procedure. Course-wise blue-print for the term-end examination has to be defined on primarily three dimensions viz. question types, content Blocks/Units and learner abilities. Preparations of question banks for courses of an educational program requires a lot of cooperative efforts. Expertise has to be tapped from all the available sources (from within and outside the university) and pooled together. Writers and reviewers of questions for the bank should have, besides their expertise in the subject content and teaching experience, sufficient grounding in evaluation methodology. Even persons selected to act as paper setters, moderators or evaluators should have, not only prescribed experience of teaching the subject, but also adequate background of modern evaluation methods. Having identified such personnel, subject/coursewise question bank task groups may be formulated. Every task group will be guided/headed by one faculty from the school and consist of 4-6 persons selected from among course-writers, teacher counsellors and experienced item-writers available from other institutions.

Advantages of Question Bank

Some examination bodies have got the question banks constructed and got the question paper set accordingly. It brought out the following merits and demerits. The question banks have the following merits :

(1) The teachers are aware what types of questions are to be asked in the examinations, so they conduct the teaching task accordingly.

(2) The students also know what types of questions will be asked in the examination, so they study accordingly.

(3) The examiners are at great ease in setting the question papers, because they have a guide before them. Questions being out-of-course cannot occur.

(4) All teachers have participation in the construction of question banks, so they understand the questions clearly, the objectivity in evaluation is then obvious.

Disadvantages of Question Bank

(1) If the number of teachers and students is large, it is not possible to seek their participation in it. Participation of the students is an impracticable concept.

(2) The scholars are not of uniform opinion, whether the question banks should be known to all teachers and students or not.

(3) The aim of an examination is not to pass or fail the students; it is rather to evaluate their educational achievements and guide them. Then what is the use of acquainting the teachers and students with question banks!

(4) Another demerit of question bank is that there is no originality in the construction of question for question papers and there is no opportunity to present a new problem before the examinees and it is, therefore, not possible to know of the ability and logic of the students under new circumstances.

The Benefits of Computer-Based Testing

In many educational online programs, we love many things about online learning, but oftentimes assessment is not one of those things! There is often concern about accuracy, validity, security, integrity, and quality of online assessment. Yet, computer-based testing offers so many benefits to online learners, instructors, and programs, mainly in terms of

administration, grading, and scale that they deserve a much closer look. As part of a multi-post series on assessment, this article makes a push for why you may want to embrace (more) computer-based/online testing.

1. Multiple-Test Administrations

Learners can take multiple, short, reliable assessments administered throughout the life of an e-learning program. The data gathered from these assessments can be correlated with national or regional standards so that learners can be measured on these standards. Data gathered over time—longitudinal data—can be used by online programs and instructors to identify trends and for program improvement. These data can also be easily exported from most LMSs into a database where they can be analyzed for evaluation and research purposes.

2. Dynamic And Individualized Assessments

Tests can be personalized and tailored to individual students. The level of difficulty of each question can be modulated depending on the learner's previous responses. (For more information on this, see this article on computer adaptive testing). Assessments can be further individualized by using programs like Skype or Face Time which allow the instructor to reach out and offer individual oral assessments to students or by using Recap or Flip Grid which allow students to share ideas, questions, etc.

3. Immediate Grading

Select-response tests (like multiple choice or True/False) can be scored instantly, allowing learners to instantly see how they did on an assessment and online instructors to

make real-time instructional changes based on assessment evidence. This immediate grading frees up instructors to focus on feedback, which is critical for learning and is also time consuming. Online instructors —or the assessment itself, depending on its design—can help and guide learners on what they need to do in order to improve. A simple example of this is Quizmaker in Articulate 360 which allows for substantive feedback and allows the course designer to redirect students who have done poorly to a specific slide or set of resources.

4. Helps With Open-Ended Assessments

As any instructor knows, open-ended assessments (like portfolios or projects) that use rubrics are extremely time-consuming to grade. Simple rubric extensions, like Orange Slice or Google Sheets add-ons, automates scoring of rubrics and communication to students. This allows instructors to spend more time on feedback to students.

5. Feedback

Voice feedback tools, like Kaizena, allows instructors to provide voice feedback which makes feedback both easier for the teacher and more personalized for the learner.

In particular, video-based feedback can provide students with individualized and personalized feedback on performance. Because (once you get the hang of it) video feedback is faster to create (by the instructor) and review (by the learner), it can provide rich, qualitative information on student performance and do so in a caring and personalized way.

6. Vertically and Horizontally Aligned Assessments

Tests can be vertically aligned—anchored to test the same core knowledge at increasing levels of difficulty (criterion-based testing). They can also be horizontally aligned— scored in such a way that learners can be compared against one another (norm-referenced), which is critical for sorting and choosing students for teaching posts, scholarships, and so forth. Using computer-based testing, raw test scores could be given phase wise[1] or as a total (Williams, 1999). Learners can receive a letter grade or percentile score to determine their relative position vis-à-vis other learners.

7. Value-Added Growth Measures

Tests measure individual growth over time, so programs are able to benchmark where learners should be at the end of the year based on tests from the beginning of the year. All of these data can be analyzed using statistical software packages so online programs can track learner growth over time.

8. Uncover Student Thinking

Games and branching scenarios can help instructors “uncover” student thinking and measure more higher-order thinking skills. By having learners play content-focused digital learning games and using “think aloud” protocols to explain their game-based decisions and rationale for such decisions, online program designers can design future learning experiences and assessments based that target specific learner competencies.

9. Engaging

The use of quiz-based video programs (like Edupuzzle), videos and video-notation tools and branching scenarios can make assessments more engaging than standard multiple-choice or essay tests. Branching scenarios, in particular, add a game-like feel to assessment, making assessment engaging and interactive and help learners learn from mistakes. Virtual reality can "immerse" students in an environment that seems real and both assesses and guides students through decisions and their consequences.

10. Analytics for The Instructor And Learner

Back-end data from LMSs, such as the number of log-ins, time on task, and number of discussion posts, can be linked to hard assessment data such as examinations or performance-based data to provide a fuller assessment of a learner's effort and progress in an online course. Additionally, "student-facing" analytics can help learners track progress and performance more easily and get notifications about assignment due dates and "early warnings" if they are in danger of failing. Diagnostics can enable students to focus on areas of weakness before a final assessment. Instructors can also use analytics to assess the quality and usefulness of course resources, the percentage of assessment items students answer correctly, and track student participation, thus prompting the instructor to intervene.

11. Greater Amount Of Test Items

This is particularly important for high-stakes assessments that determine whether or not a learner graduates, moves to the next level, or receives certification. For such critical assessments, more test items are necessary than for low-stakes assessments. Computer-based

assessments, because they draw from a back-end database of test items, typically comprise more test items than fixed paper-and-pencil exams.

12. Help Learners with Disabilities

If computer-based assessments are universally designed, they may form less of a physical impediment to test taking than is the case with paper-based tests. For example, screen readers, magnification tools, and text-to-voice or voice-to-text applications can help learners with visual, auditory, and motor impairments; learners with dyslexia; and learners who simply need more time to complete a test.

13. Incorporate Other Types Of Technology

Computer-based or online assessments offer a wealth of authentic assessment opportunities for online learners, both synchronous and asynchronous, Web-based and non-Web-based, and multiple platforms (phone, tablet, and laptop). For instance, learners can create electronic portfolios, digital representations, and collections of their work in an online-based course. They can upload images, create videos ad podcasts, dashboards, spreadsheets within a Learning Management System or upload from other programs to the learning management system. Online discussions and social media conversations can be a rich source of assessment data.

14. Improves Writing

Extensive writing via word processing or a digital writing tool—in which learners put forth a thesis statement, support their idea with evidence and supporting ideas, and come to a

conclusion—has been shown to be better than writing by hand if learners go through the complete writing cycle of drafting, editing, revising, and rewriting (Warschauer, 2009). Developing blogs, wikis, and websites, particularly with hyperlinked resources, can demonstrate learners' understanding of an issue, their appreciation of its complexity, and their knowledge of appropriate resources that address this issue. Audio- and Web-conferencing tools allow learners to present information to one another and the instructor and to engage in debates about a particular teaching-related or content-based issue.

15. Can Secure Testing

Cheating, in online and computer-based testing, is always a concern but there are a number of programs that make CBT safer than it has been, and in many cases, safer than paper-and-pencil testing. Plagiarism checkers, like Turnitin, Safe Exam Browser, Respondus, DyKnow, and Chrome browser testing apps like Edulastic can also help to lock down browsers and make cheating more difficult. Note that some may only work with Chromebooks.