PREFACE

In a bid to standardise higher education in the country, the University Grants Commission (UGC) has introduced Choice Based Credit System (CBCS) based on five types of courses viz. *core, discipline specific, generic elective, ability* and *skill enhancement* for graduate students of all programmes at Honours level. This brings in the semester pattern, which finds efficacy in sync with credit system, credit transfer, comprehensive continuous assessments and a graded pattern of evaluation. The objective is to offer learners ample flexibility to choose from a wide gamut of courses, as also to provide them lateral mobility between various educational institutions in the country where they can carry acquired credits. I am happy to note that the University has been accredited by NAAC with grade 'A'.

UGC (Open and Distance Learning Programmes and Online Learning Programmes) Regulations, 2020 have mandated compliance with CBCS for U.G. programmes for all the HEIs in this mode. Welcoming this paradigm shift in higher education, Netaji Subhas Open University (NSOU) has resolved to adopt CBCS from the academic session 2021-22 at the Under Graduate Degree Programme level. The present syllabus, framed in the spirit of syllabi recommended by UGC, lays due stress on all aspects envisaged in the curricular framework of the apex body on higher education. It will be imparted to learners over the *six* semesters of the Programme.

Self Learning Materials (SLMs) are the mainstay of Student Support Services (SSS) of an Open University. From a logistic point of view, NSOU has embarked upon CBCS presently with SLMs in English / Bengali. Eventually, the English version SLMs will be translated into Bengali too, for the benefit of learners. As always, all of our teaching faculties contributed in this process. In addition to this we have also requisitioned the services of best academics in each domain in preparation of the new SLMs. I am sure they will be of commendable academic support. We look forward to proactive feedback from all stakeholders who will participate in the teaching-learning based on these study materials. It has been a very challenging task well executed, and I congratulate all concerned in the preparation of these SLMs.

I wish the venture a grand success.

Professor (Dr.) Subha Sankar Sarkar
Vice-Chancellor

Netaji Subhas Open Univerasity
Under Graduate Degree Programme
Choice Based Credit System (CBCS)
Subject: Honours in Education (HED)
Course: Evaluation in Education
Course Code: CC-ED-08

First Print : August, 2022

Netaji Subhas Open University
Under Graduate Degree Programme
Choice Based Credit System (CBCS)
Subject: Honours in Education (HED)
Course: Evaluation in Education
Course Code: CC-ED-08

: Board of Studies :

Members

Dr. Atindranath Dey
Director, SoE, NSOU, Chairman (BoS)

Mr. Swapan Kr. Sarkar
Associate Professor, SoE, NSOU

Dr. Debasri Banerjee
Professor, Dept. of Education,
University of Calcutta

Dr. Kutubuddin Halder
Professor, Dept. of Education,
University of Calcutta

Dr. Dibyendu Bhattacharyya
Professor, Dept. of Education,
University of Kalyani

Dr. Sanat K. Ghosh
Professor, SoE, NSOU

Dr. Sumanta Chattaraj
Professor, SoE, NSOU

Dr. Papiya Upadhyay
Assistant Professor, SoE, NSOU

Dr. Parimal Sarkar
Assistant Professor, SoE, NSOU

: Course Writer :
Dr. Keka Trivedi
Assistant Professor of Geography
Institute of Education (P.G.) for women
Chandannagar, Hooghly

Course Editor
Dr. Papiya Upadhyay
Assistant Professor, SoE, NSOU

Format Editing
Dr. Parimal Sarkar
Assistant Professor, SoE, NSOU

Notification

Kishore Sengupta
Registrar

Netaji Subhas Open University

UG : Education

(HED)

Course : Evaluation of Education
Course Code: CC-ED-08

Module - I
Evaluation

Module - II
Testing & Evaluation

Block-1
Evaluation

Unit-1 ❏ Assessment, Measurement & Evaluation

Structure

1.1 Objectives

After going through the unit the students will be able :

● To define measurement, assessment and evaluation in education.

● To understand the role and significance of measurement, assessment and evaluation in teaching learning.

● To classify different types of assessment, test, examination and evaluation.

● To categorize the purpose and scope of measurement and evaluation in today's teaching learning activity.

● To create a relationship between measurement, assessment and evaluation.

1.2 Introduction

The knowledge of measurement, evaluation and assessment is essential for the teachers to make the teaching learning an effective one. The concept of measurement has been there since days immemorial. In education, measurement and quantification are very important. Measurement involves the process of quantification. The application of the principles of measurement in the field of education is known as educational measurement. Assessment is the process of documentation of the measurement. It precedes the final decision making stage of evaluation. Evaluation is a broader concept. It does not deal with only the classroom examination system; rather it deals with the evaluation of the cognitive, affective and psychomotor domain of the student. As teachers become more familiar with data-driven instruction, they are making decisions about what and how they teach based on the information gathered from their students. In other words, teachers first find out what their students know and what they do not know, and then determine how best to bridge that gap. During the process of gathering information for effective planning and instruction, the words measurement, assessment and evaluation are often used interchangeably. These words, however, have significantly different meanings. This unit will cover the various aspects of measurement, assessment and evaluation in details. The interrelationship of measurement, assessment and evaluation will also be discussed.

1.3 Evaluation: Concepts and Significance of Measurement, Assessment and Evaluation

Measurement, assessment and evaluations are very often used interchangeably in education. In reality, these words are not all synonymous. All these have different scope and characteristic features. These are as follows;

1.3.1 Concept and significance of Measurement

Generally, to measure and show the weight, length and volume of an object in definite units is called measurement; for example, to show the weight of a person in kilograms, length of cloth in metres and volume of water in litres. But the field of measurement is very wide. It includes defining any characteristic of any object or person or activity in words, symbols or units. As far as explaining the qualities of objects, persons and activities is concerned, it has been in vogue from very ancient times, of course, without any definite base of measurement. In the present times, the bases of most of

the qualities of objects, persons and activities have been defined; their standards and units have been specified; measuring tools and methods have been devised and methods to demonstrate the results of measurement in brief have been decided. Now, a characteristic of an object, person or activity is described in definite words, symbols and units in brief. Many scholars have attempted to delimit the definition of this process. According to J.P. Guilford (n.d) "measurement is the assignment of numerals to objects or events according to certain rules". Norman E. Gronlund has defined "measurement results are some score or numerical value and quantitative descriptions of the pupils". Most scholars are in agreement with the definition given by James M. Bradfield (n.d). In his words :Measurement is the process of assigning symbols to the dimension of phenomenon in order to characterise the status of phenomenon as precisely as possible. In this definition of measurement only the measurement of qualities of objects and activities has been included, and not the measurement of qualities of persons. Though the persons are included in the objects of the universe; however, the objects are meant to include only concrete materials, so it is necessary to show the measurement of qualities of the persons separately. This definition of Bradefield does not point to any such basis of measurement. Educationists opine that it should also be necessarily included in it and in that case measurement should be defined as: Measurement is the process by which a characteristic of an object, person or activity is perceived and understood on specific standards and is described in standard words, symbols or definite units.

Factors of Measurement: The above definition of measurement shows that there are four factors of measurement:

(1) The object, person or activity any of which characteristic has to be measured.

(2) The characteristic of that object, person or activity which has to be measured.

(3) The tools and devices of measuring such characteristics.

(4) The person who measures it.

Educational measurement is known as mental measurement. It is concerned with the way to measure learning. In the educational system, measurement is quantitative assessment of performance of the students in a given test. It can be used to compare performance between different students and to indicate the strength and weaknesses of the students. It helps in classifying students into homogenous groups to assign educational and vocational guidance and to provide remedial measures to the low

achievers. There are four common levels of measurement. These are as follows:

I.   **Nominal Scales:** Nominal scales are used to classify objects. It is the least precise or lowest order scale. Here, the numbers assigned to the individuals are not bound by any rules and there is no logical consideration. For example, the serial numbers given to the cricket players are only to identify the players. Nominal level is sometimes called the classification level and each class is represented by a letter, a name, a number or even a geometrical design. This scale is used for students' classification, job classification, for different types of classification etc.

II. **Ordinal scale:** In this scale ranking or grading takes place. It indicates that one observation represents more or less of a variable than the other, without indicating how much more or less. Thus ordinal scale corresponds to quantitative classification of a set of objects with reference to some attribute or rule. For example, when we serially arrange five bags of cereals according to weight, the heaviest bag is represented by numeral 1, the next by 2 and so on. Accurate and precise comparison is not possible with ordinal scale.

III. **Interval scale:** Interval scales have an arbitrary zero, not absolute zero. It is also known as equal interval scale because these scales show that a person or item is so many units larger or smaller, heavier or lighter etc. than the other one. The intervals are assumed to be equal over the scale.

IV. **Ratio scale:** the ratio scale is most concrete and refined among the four basic scales. It has an absolute zero point representing complete absence of the property being measured. Measuring the height and weight of students and making student wise records, is an example of ratio scale. If students are 20 kg and 30 kg of weight then the difference in their weight is 10kg and the weights are in the ratio 2:3.

Significance of Measurement: The significances of measurement are as follows:

● It measures students' achievement. Students' achievement can be determined whether he has reached the goals of the learning tasks or not through measurement.

● It evaluates instruction. The effectiveness or ineffectiveness of instruction is ascertained through measurement.

- It motivates learning. Upon knowing the results of the achievement test, the student's interest is aroused especially if he gets a high score; otherwise, if his score is low, he strives hard to get a higher score in the next examinations.

- Measurement is an essential element of research. No research work in education and psychology is possible without some sort of measurement.

- It helps parents to understand student's growth, interests and potentialities. The major responsibility of school and teacher is to help the parents understand their children. Understanding a youth means understanding his progress in the various areas of the curriculum, his desires and motives and behavior they lead to, his potentialities for learning, as well as his achievement.

- It helps the learners in diagnosis of the nature of difficulties. The weakness of the learner can be identified through measurement.

- Measurement in general evaluates educational goals and purposes, determines the extent to which accepted educational objectives are implemented.

- It predicts success. Success and failure of the student is predicted through measurement. For instance, a student who always gets high scores in all his subjects may mean that he is sure to pass and passing means success.

- There are individual differences in the traits and qualities of different individuals. These differences can be ascertained by means of comparison. Comparison leads us to definite conclusions. It helps us to find out which one is superior and which is inferior. Thus, it helps us to select suitable people for a particular course.

- It generates information about how to improve programmes that do not meet criteria.

1.3.2 Concept and significance of Assessment

Assessment is an ongoing process aimed at understanding and improving student learning. It involves making expectations explicit and public; setting appropriate criteria and high standards for learning quality; systematically gathering, analyzing, and interpreting evidence to determine how well performance matches those expectations and standards, and using the resulting information to document, explain, and improve performance. A critical Dictionary of Education (1986) explains the concept of Assessment as "students may be assessed for a different number of reasons and in a

number of different ways. We may for example, wish to test the student attainment, to gather data to aid enquiry into our teaching or for purposes of accountability, to make predictions concerning the students' future, to set standards, or to motivate students. The range of types of assessment runs from the conventional written essay, through oral tests or practical projects, to various types of objective tests."

Examinations and essays along with speeches and projects are forms of assessment. Assessment is a critical step in the learning process. It determines whether or not the course's learning objectives have been met. A learning objective is what students should know or be able to do by the time a lesson is completed. Assessment affects many facets of education, including students' grades, placement, and advancement as well as curriculum, instructional needs, and school funding.

Significance of Assessment

Assessment is an integral part of instruction, as it determines whether or not the goals of education are being met. Assessment inspires us to ask these hard questions: "Are we teaching what we think we are teaching?" "Are students learning what they are supposed to be learning?" "Is there a way to teach the subject better, thereby promoting better learning? Hence, the significances of assessment are as follows:

- Today's students need to know not only the basic reading and arithmetic skills, but also skills that will allow them to face a world that is continually changing. They must be able to think critically, to analyze, and to make inferences. Changes in the skills base and knowledge our students require new learning goals; these new learning goals change the relationship between assessment and instruction.

- Teachers need to take an active role in making decisions about the purpose of assessment and the content that is being assessed. Assessment and its associated feedback are essential to student learning. However, one may find that more time is taken up with the areas of assessment associated with quality assurance, rather than its potential to support students' learning. Well designed assessment has numerous benefits aside from the obvious one of providing a measure of students' progress as it can be a means to engage students with their learning. Ideally then, teacher should aim to support active learning rather than assessment of learning to ensure that the assessment process is an integral part of students' education. A student undertaking any form of study will be subject to assessment in one form or another. Similarly,

any member of the teaching staff will be engaged at some point in assessment related work. For students it can be a significant determinant of what, when and how they learn. Getting assessment 'right' is therefore essential, both for students and the teacher.

● Well-designed assessment can encourage active learning especially when the assessment delivery is innovative and engaging. Peer and self-assessment, for instance, can foster a number of skills, such as reflection, critical thinking and self-awareness as well as giving students responsibility and insight into the assessment process.

● Discussing the ways in which the teacher is assessing with students can also help to ensure that the aims and goals of assessments are clear. Utilising assessment that makes use of technology, such as the use of online discussion forums or electronic submission of work, can teach students new skills. Well designed assessment can prevent plagiarism by reducing the ways in which students can gather and report information.

● At the end of the day, taking some time to think about why, what and how the teacher is going to assess students is a worthwhile investment of time. It can ensure assessing the skills and knowledge that are intended and it could open up new possibilities for different ways to assess students, some of which may be more efficient and effective than the current methods.

● Assessment is a key component of learning because it helps students learn. When students are able to see how they are doing in a class, they are able to determine whether or not they understand course material. Assessment can also help in motivating students. If students know they are doing poorly, they may begin to work harder. For example, Johnny is a chemistry student. He just took his first examination in his class. He earned a 56%; he needs a 79% to pass the class. The low test score lets Johnny know that he missed something important he should have learned. Perhaps, he did not understand the material, or maybe he did not study long enough or did not give best effort. Whatever the case, the assessment results let Johnny know that he did not successfully learn the material and that he must try something new in order to earn a better score.

● Just as assessment helps students, assessment helps teachers. Frequent assessment allows teachers to see if their teaching has been effective.

Assessment also allows teachers to ensure students learn what they need to know in order to meet the course's learning objectives. For example, a teacher after finishing one unit, gives a 50-point multiple-choice test. Upon grading the examination, he or she realized the average class grade was a 68%, far below the cutoff line for passing. The teacher can easily see that students didn't fully learn the unit. This tells the teacher that he or she needs to re-plan and re-visit the unit and determine why students failed the examination. Perhaps the teacher may need to try a different teaching strategy, or perhaps she did not spend enough time on difficult material.

1.3.3 Concept and significance of Evaluation:

We are aware that measurement is used to express a trait of an object, person or activity in standard words, symbols or units. In evaluation, these results are analysed and this analysis is done on the basis of certain social, cultural or scientific standards (Norms) and by this analysis, the relative condition of the trait of the object, person or activity is clarified. James M. Bradefield has defined this process of evaluation in the following words: "Evaluation is the assignment of symbols to phenomenon in order to characterise the worth or value of the phenomenon usually with reference to some social, cultural and scientific standards" (James M. Bradfield, n.d). If we want to further clarify this definition in simple terms, we can do it in the following words: Evaluation is the process in which the analysis of the result obtained from measurement of a trait of an object, person or activity is done on the basis of certain social, cultural or scientific standards (Norms), and the relative position of the object, person or activity is determined as relative to that trait.

In the field of education, generally, the measurement of educational achievements of the students is called evaluation. The first thing in this context is that the measurement is the first step of evaluation; it is not evaluation in itself. In evaluation, the results of measurement are analysed according to predetermined standards (Norms). Secondly, not only the measurement of educational achievements of the students is done, but their intelligence, interest, aptitude and personality, etc. are also measured and evaluated. Besides the traits of students, the activities of other people concerned with education such as administrators, teachers, other personnel and guardians are also measured and evaluated. The measurement and evaluation of educational policy, the aims of education, the curriculum at various levels and teaching methods are also carried out and suitable suggestions are given. Therefore, the evaluation in the field of education should be defined as follows:

Educational evaluation is the process in which measurement of the decisions related to education and of the traits and activities of the persons concerned with education is carried out and the results of such measurement are analysed on the basis of predetermined standards (Norms) and on its basis the relative results are declared and suggestions for improvement in them are given. Generally, people use the words educational evaluation and measurement in identical meanings. However, they have the difference of the whole and the part. Measurement is the first step of evaluation, evaluation is a process to analyse the results of measurement. From the functional steps viewpoint, the process of measurement has only three steps - (i) selection or construction of suitable measurement tool or device, (ii) use of measurement tool or device, and (iii) recording of administration and result. Evaluation has three additional steps besides the above. These are; (i) Selection and use of suitable standards (Norms) for analysing the measurement results, (ii) analysis of measurement results on the basis of these standards (Norms), and (iii) prediction, suggestions or guidance on the basis of this analysis. The analysis of results of educational measurement under educational evaluation cannot be done as of the results of physical measurement, because the physical measures are whole measures, and educational measures are relative in nature. Now statistical methods are used for their analysis.

Factors of Evaluation

Two processes have to be undertaken in evaluation; first, the measurement and the second, analysis of the information or data obtained from measurement. We have already discussed that there are four factors of measurement; now two more factors of analysis of the data or result received from the measurement are added for evaluation. Hence the factors of evaluation are as follows:

(1) The object, person or activity any of which characteristics has to be measured.

(2) The characteristic of the object, person or process which has to be measured.

(3) The tools and devices of measuring such characteristics.

(4) And the person who measures it.

(5) Those standards (Norms) on the basis of which the results of measurement are analysed.

(6) Those devices (logical, mathematical or statistical) with the use of which such analysis is carried out.

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 an 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. 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.

Significance of Evaluation

The following significance of the Evaluation are observed:-

(1) To 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 analysis of the educational objectives, to test their utility, and to suggest for timely change.

(3) To 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 studying the effect of teaching methods being used from time to time, to find out useful/or useless methods, to suggest for improvement and to enlighten the field for research.

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

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

(7) It helps teachers to discover the needs of the students. The purpose of any programme of evaluation is to discover the needs of the pupils being evaluated and then to design learning experiences that will satisfy these needs.

1.4 Measurement versus Assessment in Education: Purpose and Scope

Very often assessment and measurements are used interchangeably. According to M.Q.Panton (1985), "measurement implies assigning a numerical quantity. While instruments such as rulers' stopwatches can be used to determine height, speed and so on, any intellectual capacities or other quantities of educational interest must be measured indirectly. Thus tests are typically used on measures such as dimensions and levels of intelligence. It may be included in an assessment." Whereas, he has opined that "as far as possible the term assessment should be reserved for application to people. It covers activities included in grading (formal and non-formal) examining, certifying and so on. Students' achievement on a particular course may be assessed."

Purpose of Measurement

The purpose of measurement are as follows:-

(1) To measure the ability of students, finding out their interests and aptitude at the time of admission, and admit them on its basis.

(2)   To measure their intelligence and personality after admission and accordingly divide them into specific classes, and to assist in their personality development.

(3)   To find out from time to time the effect of teaching on the students (educational achievements or change of behaviour), and to guide the students on its basis and to inspire them to learn.

(4)   To measure and evaluate the educational achievements of students from time to time and to provide them feedback.

(5)   To find out the hindrances in the educational progress of the students and remedy them.

Scope of Measurement

Generally, it is understood that in the field of education only educational achievements of the students are measured, and are expressed in terms of marks. However, in the field of education, in present times, various variables, such as intelligence, interest, aptitude, attitude and personality of the students are also measured. Moreover, the aims of education, usefulness of the curriculum and of the teaching methods, basis of educational policy, the educational activities of the persons concerned such as administrators, teachers, and guardians, are also measured. There is no area in the field of education which is not subject to measurement in present times and the result of which is not expressed in more and more objective terms. Therefore, we can define educational measurement as follows: Educational measurement is that process by which the usefulness of various factors of educational process, the usefulness of various activities of persons concerned, and the intelligence, interest, attitude, aptitude, personality and educational achievements of the students are measured on the basis of definite standards and are expressed in definite words, symbols or units. Measurement is an extensive process which has several tools and devices. The physical aspects of a person, such as height, weight and temperature are measured with the help of scale, weights and thermometer respectively and his interest, aptitude, intelligence, personality and educational achievement are measured with the help of Interest Tests, Aptitude Tests, Intelligence Tests, Personality Tests and Achievement Tests respectively. Evidently, measurement is a process to measure the characteristic of an object, person or activity; and testing is only a device or method of measurement. They are different like the whole and its part. Thorndike and Hagen (1979) have provided six propositions that seem to use us to provide the foundations for a contemporary view of educational and psychological measurement procedures and their role in our schools and our society.

These propositions are as follows:

1. Various types of decisions have to be made

2. The more relevant and more accurate the information, the better is the decision.

3. Measurement instruments and procedures provide an important set of tools for improving the information available for decision making.

4. The use of any type of information for decision making needs to know what that information signifies and how far it can be trusted.

5. The facts and values involved in any decision are complex.

6. The wisdom of the decision maker is crucial.

Moreover, according to Thorndike and Hagen measurement in any field involves following three common steps.

A. Identifying and defining the quality or attribute that is to be measured.

B. Determining a set of operations by which the attribute may be made manifest and perceivable

C. Establishing a set of procedures or definitions for translating observations into quantitative statements of degree or amount.

In the present age, measurement has influenced the progress in education and psychology too. Today, the age of theoretical education is over, and effort is being made to make education and psychology more and more practical. Under education and psychology different human behaviours and problems are studied. For this, it becomes necessary to measure human behaviours. Educational measurement is not a new concept. The teacher has been testing students since times immemorial in order to know their progress in studies, and to see what type of behavioural changes have occurred in students, if they are optimal and in what direction these behavioural changes have taken place. A teacher wants to know the shortcomings of the method of teaching he uses, for which these tests have been found to be very important. These tests have become all the more important in the present age, and it is expected from the future teacher or pupil-teacher that he or she would gain skill in constructing several types of tools of measurement and evaluation. The introduction of evaluation in the educational world is comparatively new. In fact, it was introduced in this field in order to get rid of the serious shortcomings of measurement. From the beginning

of the twentieth century, three types of important progresses were noted down in the field of educational measurement, they were testing, measurement and evaluation. This progress in education is known as progress in measurement. However, in the present times, evaluation is considered to be an important contribution of psychology and education. Recently, the attention of educationists and psychologists has been drawn towards evaluation.

Purpose of Educational Assessment

The purpose of Educational assessment is as follows:

    (a)   To prepare students for the new pattern of competency-based teaching and competency-based assessment.

    (b)   To make evaluation an integral part of the teaching-learning process.

    (c)   To make evaluation a cooperative and democratic process, involving various stake-holders of the evaluation programme.

    (d)   To collect more valid and reliable evidence about total development of the learners.

    (e)   To form sound judgements and take timely decisions about learners' instructions and learning environments.

Scope of Assessment

Comprehensiveness is the significant factor in the assessment of the whole chi1d vis-a-vis his total development to form the basis of assessment. Accordingly, the following areas of assessment will be included in the CCE. Assessment can be of two types; assessment of scholastic aspect and non scholastic aspects of human behavior.

Assessment of Scholastic Aspects

Use of the following tools will be made in assessment of scholastic aspects. Different types of tests required for assessing the scholastic aspects are given below:

    (a)   **Readiness tests:** For Standard V the prerequisite learning is indicated by Standard IV learning competencies. Before taking up teaching of a unit, a pretest based on pre-requisite learning competencies will be used and inadequacies in students' entry behaviours would be analysed. Based on learning gaps remedial teaching is undertaken. This will also apply to Standards VI and VII, VIII, IX, X etc.

(b) Unit-tests: Since unit teaching approach is envisaged, unit-test should be used at the end of each learning unit. It could be written or oral for review and reinforcement purposes. Each unit-test will be based on specific content elements of each unit, used for testing students on pre-determined learning outcomes formulated for each unit in terms of competencies. Focus of unit-tests will be on diagnosis of students' inadequacies in learning and improvement of their achievement through remedial action to enable most of the students to acquire mastery of concepts and skills. Sample criterion-referenced tests will be developed and supplied to schools and teachers oriented in unit testing.

Assessment of Non-Scholastic Aspects: This would include assessment of the following aspects which may be introduced in the second phase.

1. Physical health, covering basic understanding about nutrition and health, physical fitness, development of positive attitudes etc. Habits like health habits, study habits and work habits.

2. Interests in artistic, scientific, musical, literary and social service activities.

3. Attitudes towards students, teachers, class-mates, programmes, school property etc.

4. Character-building values like cleanliness, truthfulness, industriousness, cooperation, equality etc.

5. Participation in games, sports, gymnasium, literacy, scientific, cultural, social and community service activities.

Issues of Present System of Assessment

(a) Unless protagonists of reform develop a deep conviction and full faith in teachers' assessments and conceptualise dichotomy observable in theory and practice, establishment of a pedagogically sound system of assessment would be difficult.

(b) Unless a theoretically sound, indigenous, growth-oriented and functionally operative model of assessment is developed and made the basis of continuous and comprehensive evaluation; we are not likely to achieve the intended objectives of assessment to the desired extent.

(c) Until and unless within a given state, board of school education or the

concerned examining agency where curriculum requirements are the same for a particular stage of school education, a common blueprint of CCE based on the same design of assessment is adopted in all schools under the same board, comparability of attainment targets and pupils' achievement across the schools would not be possible, which in turn would pass it difficult to pass valid norm-referenced judgments.

(d) No system of school-based assessment would succeed unless we have full faith in the competency of practising teachers and provide them continued motivation, encouragement and recurrent training for empowering them to implement school based assessment systems effectively to achieve the desired attainment targets.

(e) Extreme subjectivity in internal assessment reigns supreme in school based assessment, which indeed is the stumbling block in implementing this scheme. Unless multiple scoring cooperative testing and quality question-banks are developed and used for assessment of scholastic achievement, and illuminative approach reflecting multiple views of participant observers are used in assessing non-cognitive outcomes and personal social qualities, it would not be possible for teachers to make objective assessment, form sound judgements and take right decisions.

(f) Unless regular feedback of formative evaluation (which is an integral part of teaching) and functional research support are provided continually for diagnostic purpose and remedial teaching, further improvement in pupils' achievement and instructional strategies would not be possible.

Current Trends in Educational Assessment

1. Visualise the context and background of the institution

2. Take cognizance of time frame, students' intake, input resources etc.

3. Study curriculum objectives of the prescribed courses of studies.

4. Formulate clearly subject-wise objectives in achievable and testable terms, covering both cognitive and non-cognitive domains.

5. Specify subject-wise objective in terms of intended learning outcomes, i.e. Essential Learning competencies and Desirable Learning Competencies.

6. Clarify the purpose and objectives of School-based Assessment.

7. Delineate the areas and components of assessment for each subject.

8. Choose the relevant mode of assessment to be used at different stages of instruction and evaluation of pupils and programmes.

9. Identify relevant techniques and tools of assessment for measuring both scholastic and co scholastic attainments of pupils.

10. Ensure needed validity, reliability and usability while constructing measuring instruments and other assessment devices.

11. Apply relevant tools and techniques to collect needed evidence about pupils' attainments on cognitive and non-cognitive outcomes of learning.

12. Analyse evidence in terms of specified intended learning outcomes.

13. Interpret the evidence in terms of assessment objectives.

14. Form appropriate judgments about pupils' performance in terms of self-referenced, criterion referenced and norm-referenced judgements.

15. Summarise results and other evidence into meaningful individual and group profiles for different areas and aspects.

16. Report meaningfully and regularly to students, parents and other functionaries concerned, for timely feedback and correctives.

17. Take appropriate decisions about further diagnosis (if needed), remedial measures, grading, promoting, certification, programme budgeting and improvement of instructional, evaluation and implementation strategies.

18. Initiate development of institutional question banks for improving the quality of test material and maintain performance standards.

19. Use summative evaluation at school, school complex, block and district levels through cooperative testing and participative strategy for maintaining and comparing performance standards.

20. Undertake annual institutional self-evaluation involving teachers, principals, students, parents, and community and evaluation experts to assess the efforts, efficiency and effectiveness of the school in terms of the scheme of SBA and review it for making the scheme more functionally operative.

1.5 Interrelationship Among Assessment, Measurement and Evaluation in Education

Assessment is the process of documentation, usually in measurable terms, knowledge, skills, attitude and beliefs. It refers to the task of processes of estimation of something. As the dictionary meaning of the term is to 'estimate the value or quantity of something', it is used to estimate the amount of learning through any process may be instructional or otherwise. Assessment has a narrower meaning than evaluation but broader meaning than measurement. Since the word assess means 'to assist the judge'. So it seems appropriate in evaluation studies to the process of gathering the data and fashioning them into an interpretable form. Judgement can then be made on the basis of the assessment. It precedes the final decision making stage in evaluation.

Relationship between Assessment and Measurement: Assessment, measurement, evaluation and research are part of the processes of science and issues related to each topic often overlap. Assessment refers to the collection of data to better understand an issue, while measurement is the process of quantifying assessment data and evaluation refers to the comparison of that data to a standard for the purpose of judging worth or quality. Whereas, research refers to the use of that data for the purpose of describing, predicting, and controlling as a means toward better understanding the phenomena under consideration. Measurement is done with respect to "variables" (phenomena that can take on more than one value or level).

The collecting of data (assessment), quantifying those data (measurement) and developing understanding about the data (research) always raise issues of reliability and validity. Reliability attempts to answer concerns about the consistency of the information (data) collected, while validity focuses on accuracy or truth. The relationship between reliability and validity can be confusing because measurements (e.g., scores on tests, recorded statements about classroom behavior) can be reliable (consistent) without being valid (accurate or true). However, the reverse is not true: measurements cannot be valid without being reliable. The same statement applies to findings from research studies. Findings may be reliable (consistent across studies), but not valid (accurate or true statements about relationships among "variables"), but findings may not be valid if they are not reliable. At a minimum, for a measurement to be reliable a consistent set of data must be produced each time it is used; for a research study to be reliable it should produce consistent results each time it is performed. For

example, the variable "gender" has the values or levels of male and female and data could be collected relative to this variable. Data on variables are normally collected by one or more of four methods: paper/pencil, systematic observation, participant observation, and clinical. Three issues are important for classroom assessment (data collection with regards to student learning that is under the control of the teacher.) The first relates to what data teachers will use for making judgements (qualitative or quantitative); a second issue revolves around when they will collect data (formative vs. summative assessment.) A third issue revolves around the reference to be used for making evaluations (criterion- versus norm-referenced); a fourth relates to how teachers will communicate their judgements to others (authentic assessment, portfolios, and grading).
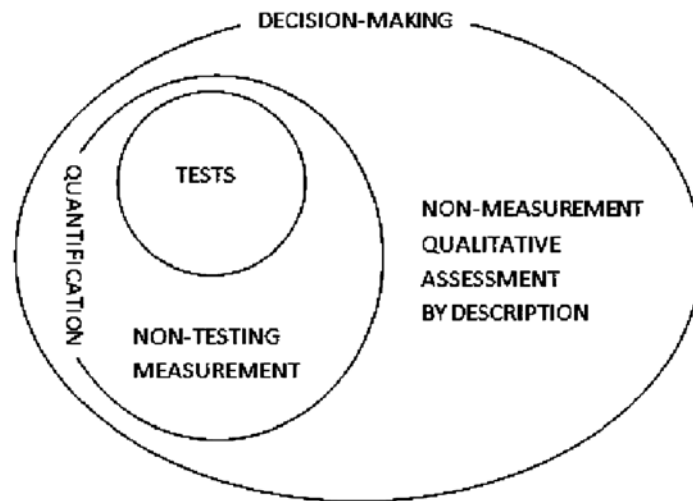

Fig 1. Relationship between Measurement and Assessment

As the above diagram shows, tests constitute only a small set of options, among a wide range of other options, for a language teacher to make decisions about students. The judgment emanating from a test is not necessarily more valid or reliable from the one deriving from qualitative procedures since both should meet reliability or validity criteria to be considered as informed decisions. The area circumscribed within quantitative decision-making is relatively small and represents a specific choice made by the teacher at a particular time in the course while the vast area outside which covers all non-measurement qualitative assessment procedures representing the wider

range of procedures and their general nature. This means that the qualitative approaches which result in descriptions of individuals, as contrasted to quantitative approaches which result in numbers, can go hand in hand with the teaching and learning experiences in the class and they can reveal more subtle shades of students' proficiency. This in turn can lead to more illuminating insight about future progress and attainment of goals. However, the options discussed above are not a matter of either-or (traditional vs. alternative assessment) rather the language teacher is free to choose the one alternative (among alternatives in assessment) which best suits the particular moment in his particular class for particular students.

Based on the above discussion, grading could be considered a component of assessment, i.e., a formal, summative, final and product-oriented judgment of overall quality of worth of a student's performance or achievement in a particular educational activity, e.g., a course. Generally, grading also employs a comparative standard of measurement and sets up a competitive relationship between those receiving the grades. Most proponents of assessment, however, would argue that grading and assessment are two different things, or at least opposite poles on the evaluation spectrum. For them, assessment measures student growth and progress on an individual basis, emphasizing informal, formative, process-oriented reflective feedback and communication between student and teacher. Ultimately, the conception probably depends more on teaching philosophy than anything else.

Assessment and measurement are closely linked concepts in education. Both can be used for a variety of purposes, including:

- Reporting to students, parents and the public;

- As a method for supporting students' understanding of educational goals and their own progress;

- As a means for teachers to understand the effectiveness of their own practice;

- Strengthening school-wide collaboration;

- Providing the data necessary to discern whether the system is working equitably for all students.

- Assessment and measurement play a key role in informing educators' practice in classrooms, students' understanding of their own learning, parents' capacity to support their children's success in school, and policy-makers' decision-making.

- Competencies in creativity, social-emotional learning, citizenship, and health should be assessed for the same reasons that reading, writing, and mathematics are assessed to provide relevant, specific information about student learning in these vital areas.

- Assessment of these competencies is complex, and we cannot rely on the tools and strategies typically used to assess other skills or knowledge.

- At the classroom level, a range of assessment strategies can be used to understand students' growth in these areas.

- It is possible to assess these competencies at a jurisdictional level; however standardized assessments or surveys can only give information of limited quality about complex competencies.

Relationship between Assessment and Evaluation: Assessment is a methodical way of acquiring, reviewing and using information about someone or something, so as to make improvement where necessary. The term is interpreted in a variety of ways, i.e. educational, psychological, financial, taxation, human resource and so on. In general, assessment is an ongoing interactive process, in which two parties (assessor and assessee) are involved. The assessor is someone who assesses the performance based on the defined standards, while assessee is someone who is being assessed. The process aims at determining the effectiveness of the overall performance of the assessee and the areas of improvement. The process involves, setting up goals, collecting information (qualitative and quantitative) and using the information for increasing quality.

The term 'evaluation' is derived from the word 'value' which refers to 'usefulness of something'. Therefore, evaluation is an examination of something to measure its utility. Therefore, evaluation is a systematic and objective process of measuring or observing someone or something, with an aim of drawing conclusions, using criteria, usually governed by set standards or by making a comparison. It gauges the performance of a person, completed project, process or product, to determine its worth or significance.

The evaluation includes both quantitative and qualitative analysis of data and undertaken once in a while. It ascertains whether the standards or goals established are met or not. If they are met successfully, then it identifies the difference between actual and intended outcomes.

The significant differences between assessment and evaluation are discussed in the points given below:

1. The process of collecting, reviewing and using data, for the purpose of improvement in the current performance, is called assessment. A process of passing judgement, on the basis of defined criteria and evidence is called evaluation.

2. Assessment is diagnostic in nature as it tends to identify areas of improvement. On the other hand, evaluation is judgmental, because it aims at providing an overall grade.

3. The assessment provides feedback on performance and ways to enhance performance in future. As against this, evaluation ascertains whether the standards are met or not.

4. The purpose of assessment is formative, i.e. to increase quality whereas evaluation is all about judging quality, therefore the purpose is summative.

5. Assessment is concerned with process, while evaluation focuses on product.

6. In an assessment, the feedback is based on observation and positive & negative points. In contrast to evaluation, in which the feedback relies on the level of quality as per set standard.

7. In an assessment, the relationship between assessor and student is reflective, i.e. the criteria are defined internally. On the contrary, the evaluator and student share a prescriptive relationship, wherein the standards are imposed externally.

8. The criteria for assessment are set by both the parties jointly. As opposed to evaluation, wherein the criteria are set by the evaluator.

9. The measurement standards for assessment are absolute, which seeks to achieve the quintessential outcome. As against this, standards of measurement for evaluation are comparative, that makes a distinction between better and worse.

So, after reviewing the points above, it would be clear that assessment and evaluation are completely different. While evaluation involves making judgements, assessment is concerned with correcting the deficiencies in one's performance. Although, they play a crucial role in analysing and refining the performance of a person, product, project or process.

Relationship between Measurement and Evaluation: Distinguishing between measurement and evaluation, Wrightstone has written: "The emphasis in measurement is upon single aspect of subject matter achievement or specific skills and abilities, but the emphasis in evaluation is upon broad personality changes and major objectives of an educational programme. These include not only subject matter achievement but also attitudes, interests, ideals, ways of thinking, work-habits and personal and social adaptability, e.g. by testing a child in mathematics we may measure his mathematical ability and nothing else. We may not have any idea about the interests, abilities etc. of the child in mathematics by administering this single test."

Difference between Educational Evaluation and Testing: Educational evaluation is an extensive concept which includes the measurement and analysis of educational policy, educational planning educational objectives, educational curriculum, teaching methods, educational administration, activities of teachers and guardians, students' intelligence, interest, aptitude and personality, etc. and their educational achievements are carried out and on the basis of the results so received suggestions are given and predictions are made. On the other hand, testing is only one of the methods of measurement.

Difference between Educational Evaluation and Examination: Educational evaluation is a broad concept. Under it, the measurement and analysis of educational policy, educational planning, educational objectives, activities of educational administrators, teachers and guardians, and students' intelligence, interest, aptitude and personality, etc. is carried out, and in the end, predictions, suggestions and guidance are provided. On the other hand, examination is used for measurement and evaluation of only educational achievements. Though there is a difference between measurement and evaluation, because measurement is a part of evaluation and is the first step of evaluation, so now it is used as a joint concept in the field of education as 'measurement and evaluation'. Secondly, there is no significance of measurement in the field of education until its results are not analysed. When the results of measurement are analysed, it is called evaluation. Thirdly, whether we join measurement with evaluation or not, the process of measurement is included in it. Generally the term 'measurement' is confused with 'Evaluation', while they are quite apart from each other. Measurement is a quantitative form of an object, while evaluation presents its qualitative together with its measurement. In brief, measurement is numerical and evaluation is quantitative. Measurement tells us how much of an object, while evaluation tells us how good it is. Besides, evaluation pays attention to the fact of realization of

objectives, while by measurement is meant only the number of specific objectives that have been realized. Measurement is incomplete without evaluation.

Difference between Measurement and Evaluation:

1. Measurement does not express any clear assumption about a student. Clear assumptions about a student can be formed on the basis of evaluation.

2. Measurement does not require much energy and time. Evaluation requires more energy and time.

3. Measurement is objective, evaluation is subjective by large.

4. Measurement is concerned with only the amount, quantity or frequency of a variable. Evaluation matches such an amount, quantity or frequency with relevant criteria for the purpose of making some value judgement about measured amount.

5. Measurement is static at the most periodical. Evaluation is a continuous and never ending process.

6. At a time measurement involves a single aspect of achievement, a single variable that is mathematical ability or scientific skill etc. whereas, evaluation is a more comprehensive concept. It involves not only achievement but also the attitudes, interests, ideas, way of thinking, work habits, personal and social adaptability.

7. The scope of measurement is limited, only some dimensions of personality can be tested under measurement. The scope of evaluation is wide, in it; the entire personality of a student is tested.

8. Measurement is content-oriented. Evaluation is objective-oriented.

9. Measurement is a means, and not an end in itself. Evaluation is an end in itself.

10. The purpose of measurement is to gather evidence. Evaluation is to deduce inferences from evidence, that is, its work is appraisement of evidence.

11. Measurement may not be an essential part of education. Evaluation is the integrated or necessary part of education.

12. Measurement answers the question 'how much', such that Sunanda has scored 56 marks in English. Therefore, measurement is quantitative.

Evaluation answers the question 'what value' like Sunanda has scored 50% marks in English and passed second division in the class. Therefore, evaluation is qualitative.

13. Prediction cannot be made meaningfully on the basis of measurement. Evaluation can predict meaningfully.

14. Measurement acquaints with a situation. This is isolated from the entire situation. Evaluation acquaints us with the entire environment. It judges its worth or value.

15. Measurement indicates those observations which are displayed numerically. Evaluation comprises of both quantitative and qualitative observations.

16. Measurement can be conducted at any time. Evaluation is a continuous process.

17. Measurement is absolute, evaluation is relative.

18. Measurement does not look before and after. Evaluation on the other hand takes into account the past achievements and also the future goals.

19. The unit of measurement is fixed and constant throughout measurement. In evaluation, the unit of measurement is not fixed and varies during the process of evaluation.

20. Measurement is more accurate because of its true zero pint. Evaluation is less accurate because of its arbitrary zero point.

21. Measurement is direct, evaluation is indirect.

Relevance between Measurement and Evaluation: Measurement and Evaluation both are used for assessing the internal qualities like I.Q., aptitude, attitude and intelligence of students. The main purposes of measurement and evaluation are:-

● Placement of students which involves bringing students appropriately in the learning sequence and classification or streaming of students according to ability or subjects.

● Selecting the students for courses; general, professional, technical, commercial etc.

● Both help to certify that a student has achieved a particular level of performance.

- Both help in stimulating learning. This can be motivation of the student or teacher, providing feedback, suggesting suitable practice etc.

- Both help in improving teaching by reviewing the effectiveness of teaching arrangements.

The word measurement, as it applies to education, is not substantially different from when it is used in any other field. It simply means determining the attributes or dimensions of an object, skill or knowledge. We use common objects in the physical world to measure, such as tape measures, scales and meters. These measurement tools are held to standards and can be used to obtain reliable results. When used properly, they accurately gather data for educators and administrators. Some standard measurements in education are raw scores, percentile ranks and standard scores. One of the primary measurement tools in education is the assessment. Teachers gather information by giving tests, conducting interviews and monitoring behaviour. The assessment should be carefully prepared and administered to ensure its reliability and validity. In other words, an assessment must provide consistent results and it must measure what it claims to measure. Creating valid and reliable assessments is critical to accurately measuring educational data. Evaluating the information gathered, however, is equally important to the effective use of the information for instruction. In education, evaluation is the process of using the measurements gathered in the assessments. Teachers use this information to judge the relationship between what was intended by the instruction and what was learned. They evaluate the information gathered to determine what students know and understand, how far they have progressed and how fast and how their scores and progress compare to those of other students.

According to educator and author, Graham Nuthall, in his book The Hidden Lives of Learners, "In most of the classrooms we have studied, each student already knows about 40-50% of what the teacher is teaching." The goal of data-driven instruction is to avoid teaching students what they already know and teach what they do not know in a way the students will best respond to.

For the same reason, educators and administrators understand that assessing students and evaluating the results must be ongoing and frequent. Scheduled assessments are important to the process, but teachers must also be prepared to re-assess students, even if informally, when they

sense students are either bored with the daily lesson or frustrated by material they are not prepared for. Teachers can improvise or re design instruction, using the

measurements of these intermittent formative assessments, to meet the needs of their students on a daily and weekly basis.

1.6 Summary

Measurement is the use of rules to assign a number to a specific observation of a variable. The rule of measuring length, for example, is to assign a number equal to the number of lengths of a standard ruler that fit exactly from one end of the desk to the other. Here, variable being measurement length and the observation is the length of a specific desk. In education measurement typically denotes how much of some cognitive or affective attribute is possessed by an individual or group of individuals. Measurement refers to the task of determining the amount of quantity in some determined units. As this is applicable in any field, determining student achievement as a quantitative measure is called educational measurement. Measurement scales differ in four levels of measurement, such as Nominal scale, Ordinal Interval Scale and Ratio scale. Whereas, assessment is the process of documentation usually in measurable terms, knowledge, skill, attitude and beliefs. Assessment is a process by which one attempts to measure the quantity and quality of learning and teaching using various assessment techniques, assignments, projects, objective type tests. It covers activities included in grading (formal and informal), examining, certifying and so on. Students' achievement on a particular course may be assessed. Assessment should be valid and reliable. Formal assessments are usually written documents such as tests; quizzes etc. informal assessment usually occurs in a more casual manner, including observation, inventories, participation, peer and self evaluation and discussion. Assessment can be either objective or subjective. Tests are systematic procedures for observing persons and describing them with either a numerical scale or a category system. Thus tests may give either qualitative or quantitative information. Two forms of mark referencing schemes are generally used; Norm Referenced Test and Criterion Referenced Test. An examination is a detailed inspection or analysis of an object or person. In academic context examinations are tests which aim to determine the ability of a student or a prospective practitioner. Examinations can be of different types, like essay type or written, objective type, oral type etc. "Value" is the inherent idea of evaluation. When we evaluate, what we are doing is engaging in some process that is designed to provide information that will help us make judgement about a given situation. It pertains to the estimation of quality. In the instructional context, this term is used widely to indicate the quality of learning. Instructional process has the purpose of

enabling each pupil to actualise one's capacities to the optimum possible extent. Evaluation can be classified into two groups; formative and summative. Assessment made during the instructional phase is formative in nature. On the other hand, summative evaluation is the terminal assessment of performance at the end of the instruction. Evaluation is a continuous process for improving the quality. It includes both academic and non academic subjects. Evaluation differs from measurement in a number of ways. Measurement is a quantitative determination of how much an individual has performed. Whereas, evaluation is a qualitative judgement of how good or how satisfactory an individual has performed. Measurement implies only a precise quantitative assessment of instructional outcomes. But evaluation is a continuous and comprehensive process which covers every aspect of an individual's achievement in the educational programme. Assessment has a narrower meaning than evaluation but broader meaning than measurement.

1.7 Self Assessment Questions

1. What is the significance of measurement in education?

2. What are the different levels of measurement?

3. Define evaluation.

4. Differentiate between formative and summative evaluation.

5. Discuss the scope of assessment in the modern education system.

6. Differentiate between test and evaluation.

7. What is the relation between measurement and assessment?

8. Differentiate between measurement and evaluation.

9. Differentiate between evaluation and examination.

10. State any two relevance of measurement and evaluation.

1.8 References Books

- Aggarwal, J.C. (2001): Essentials of Examination System: Evaluation, Test and Measurement, Vikash Publications, Delhi, page no.18-43, 154, 125-144, 218-244

- Dash, B.N. , Dash, Nibedita and Rath, M.P.(2003): A Textbook of Educa-

tional Psychology, Dominant Publishers, Delhi, India, page no.482-489

- Lovely Professional University (2012): Educational Management and Evaluation DEDU504, USI Publications, New Delhi, page no. 1-36

- Rout, Santosh Kumar (2009): Educational Measurement, Evaluation and statistics, Navyug Publishers and Distributors, New Delhi, ISBN 978-81-906940-8-7, page no. 1-21

- Sarkar, Amal kanti (2010): Measurement and Evaluation in Education, Kolkata, India, Rita Book House, page no 19-26, 31-32

- Singh, A.K. (2002): Tests, Measurements and Research Methods in Behavioural Science, Patna, Bihar, India, Bharati Bhawan Publishers and Distributors, page no 7-9

web references:

- https://keydifferences.com/difference-between-assessment-and-evaluation.html retrieved on 23.06.2020

- https://online.stu.edu/articles/education/educational-measurement-assessment-ev aluation.aspx retrieved on 23.06.2020

- https://peopleforeducation.ca/mwm-sharing-the-thinking/assessment-measureme nt-insight/ retrieved on 22.06.2020

- https://www.edutopia.org/assessment-guide-importance retrieved on 21.06.2020 https://www.reading.ac.uk/engageinassessment/why-is-assessment-important/eia-why-is-assessment-important.aspx retrieved on 21.06.2020

- https://cehs01.unl.edu/aalbano/intromeasurement/mainch2.html, retrieved on 15.07.2020

Unit-2 ❏ Tools & Techniques of Educational Evaluation

Structure

2.1 Objective

On completion of this unit the students will be able—

- to describe the various tools and techniques applied in educational assessment

- to define teacher made tests

- to classify teachers made tests on the basis of purpose, mode and nature of the tests.

- to identify the features, advantages and disadvantages of different types of tests.

- to discuss features of different techniques of Evaluation

- to compare the different techniques of evaluation and apply them in suitable situation

2.2 Introduction

Various tools and techniques are applied in the classroom situation to assess students' progress. All the four basic techniques of assessment, i.e. testing, observation, inquiry and analysis, are employed in different situations to collect the relevant evidence about pupils' growth and development in both cognitive and non-cognitive areas. Two types of Achievement tests are applied by teachers; Teacher made tests and Standardised Tests. Depending on the purpose and mode or nature of the tests teacher made tests can be classified into several types. Essay type, short Answer type and Objective type tests will be discussed in this unit. Moreover, different techniques of educational evaluation such as Viva Voce, Interview, Group Discussion and Observation will also be discussed in detail.

2.3 Tools and Techniques in Educational Assessment

Evaluation and assessment are intimately associated with the teaching-learning process. Continuous and comprehensive assessment (CCA) emphasizes on two fold objectives. These are continuity in assessment and assessment of all aspects of learning. Thus the term 'continuous' refers to assessment on an intermittent basis rather than a one-time event. When the assessment exercises are conducted in short intervals on regular basis, the assessment tends to become continuous. In other words, it can be said that if the time interval between two consecutive assessment events can be lessened or minimized then the assessment will become continuous. In order to make the assessment process continuous, the assessment activities must be spread over the whole academic year. It means regularity of assessment, frequent unit testing; diagnosis of the learning difficulty of the learners, using corrective measures, providing feedback to the learners regarding their progress etc will have to happen maximally. The second term 'comprehensive' means assessment of both scholastic and co scholastic aspects of student's development. Since all the abilities of the learners development cannot be assessed through written and oral activities, there is a need to employ variety of tools

and techniques (both testing and non-testing techniques) for the assessment of all the aspects of learners' development.

There is a wide choice of methods or tools and techniques which can be used by the teacher to assess different dimensions of a child's learning in different subjects. Apart from the traditional paper pencil test and oral tests, the teacher can use other modes of observation, self reporting, assignments, projects, portfolio, checklists, rating scales, anecdotal records, interview, questionnaire, projective, sociometric techniques etc. Use of multiple tools are required to enable the teacher to assess the learners in a more comprehensive and objective manner. The various tools include achievement test, psychological test, written test, oral test, diagnostic test.

- **Achievement test:** An achievement test is a test of developed skill or knowledge. Achievement is a term used to specify learning outcomes. Gronlund defined an achievement test as "system procedure for determining the amount a student has learned through instruction". Achievement test scores are often used in an educational system to determine what level of instruction for which a student is prepared. Achievement test may be of two types:

  A. **Teacher made test:** A test made by the teacher to assess comprehensively the extent and degree of students' progress with reference to specific classroom activities commonly known as teacher made test.

  B. **Standardized test:** Standardised tests are designed in such a way that the questions, conditions for administering, scoring procedures and interpretations are consistent and are administered and scored in a predetermined standard manner.

- **Psychological test:** Psychological tests are survey tools that gauge certain traits, aptitude, habits or needs. In our educational system most commonly used psychological tests are intelligence test, personality test, interest inventories, aptitude test etc. these tests assess and evaluate information that an individual gives to the examiner.

- **Written test:** Written tests contain questions requiring the students to respond in writing. This is the most popular test because of its ease of administration. It can be classified into essay type, short answer type and objective type.

- **Oral test:** In oral tests, the examiners ask questions, the answer of which the

students have to give orally. Oral tests are more flexible than other types of tests for the examiner can build upon the answers of the students. The students also get a chance to explain their answer and clarify points to the examiner.

- **Diagnostic tests:** A test designed to determine the specific learning needs of an individual, or a test used to diagnose, analyse or identify specific areas of weakness and strength, to determine the nature of weakness or deficiencies of an individual is known as diagnostic test. There are diagnostic achievement tests which are used for skills like reading, mathematics; spelling etc. questioning is a type of verbal diagnostic test before the lesson.

Besides these tools various techniques for conducting assessment and evaluation exercises have been developed and are in use among educational practitioners. Some of the most commonly used techniques of evaluation are as follows:

- **Observing Children's learning Activities:** Observing the children during the course of teaching is a technique of assessment for learning or formative assessment. Through observation information can be gathered about children in natural settings. Some observations can be done in course of teaching from the behaviours like how the child is answering the questions asked to him/her, how he is describing the facts, how he is responding to the answers of other children, what type of questions he is asking the teacher, how he is presenting the group reports, how he is participating in the discussions etc. The teacher could record the number of questions asked by distinct students during the activity. There will be valuable information indicating whether learning was taking place. Through observation various aspects of the child's personality development can be assessed. It can be used to assess individuals and groups. Through observation, the teacher can get evidence of the child's performance based 'on-the-spot' record. Over time, detailed observations of behaviour as well as interests, challenges-patterns or trends emerge which allow teachers to create a comprehensive picture of the child. So observations are to be made by the teacher over a period of time, across different activities and settings. When the child is engaged in a number of tasks or activities, it will be easier on the part of the teacher to observe the child and assess him. So teaching based on the lecture method cannot create scope for the child to do much work. Thus, the child may be engaged in a lot of activities either individually or in a group which will help the teacher to assess the child's

learning as well as enable the child to identify his own weaknesses and rectify them. For example, if the class is divided into small groups and told that they are going to use measurement to find some secret objects inside the school campus. Each group may be provided the name of a set of objects including the secret objects. The measurement of the secret objects may also be given to the children such as- the length of the object is 2 m 15 cm and its width is 1m 10 cm etc. Each group must measure objects provided to them until they identify the objects described by the measurement. While the child is busy in the activity teacher can observe him/her from a very close quarter without disturbing their attention on the activity. This would provide information on the children's style of learning and their learning difficulties. The teacher can observe the level of participation of students in the activities in the classroom. Many students are afraid of learning and do not like to get involved in group activities. Detecting their reservations, teachers can take appropriate actions to improve their participation in classroom activities. Observation provides a better opportunity to detect the degree of involvement of each child in the activities, several personality characteristics, the strengths and weaknesses of students on the topics of concern. Observing students in class, teachers may gain insight into several aspects of their learning of concepts like interest for activity, aesthetic sense, symptoms of anxiety and phobia, typical errors committed, alternative methods applied for solution of problems, specific points of difficulty etc. Checklist, rating scales, score cards etc are some of the examples of devices generally used for observation technique.

● **Self reporting technique:** Self reporting technique may include use of anecdotal record, interview, questionnaire, sociogram etc.

● **Interview:** An interview may be defined as a face to face verbal exchange. It refers to seeing and talking with an individual for some time with some definite purpose. Personal interview enables an interviewer to know an individual student. At the time of interview through friendly conversation and some skillful questions the interviewer tries to explore ideas, felt needs and problems of the students.

● **Anecdotal records:** An anecdotal record is a written record kept in a positive tone of a student's progress based on milestones particular to that student's social, emotional, physical, aesthetic and cognitive development. The teacher

observes and then records a student's action and work throughout the while the activities are occurring. Recording is informal and typically based on notes or a checklist with space for writing comments over a span of period at frequent intervals.

- **Sociogram:** Sociograms are the charts or tools used to find the sociometry of a social space, especially in the classroom environment they are a useful tool. It is also known as a friendship chart, as it is a diagram that allows the teacher to analyse the social makeup of the class. A sociogram is constructed after students answer a series of questions that illustrate the student's preferences about classmates. A sociogram is an important tool in the social discipline model that allows a teacher to provide misbehaving students opportunities for social acceptance. As a result the misbehavior will no longer occur.

- **Question bank:** We need a question bank in each class of each school for the purpose of collecting questions from various sources, arranging and keeping (storing) them properly and using them as and when required. Preparing objective based test items of different types is quite important on the part of a teacher and at the same time, it is also not an easy task. But with a stock of quite a large number of items (questions) at hand, it will have little difficulty in using appropriate test students at different times. The question bank is useful for the teachers in bringing reform in the traditional evaluation system. The National Curriculum Framework (2005) rightly observed that, the present evaluation system can be described as "one-exam fits-all", as one question paper is employed to all students during the examination. This is because the teacher has no other options but to use some questions which are available with him. But, if the teacher has a variety of questions in the question bank then he can prepare different question papers and use them for different learners as per their requirement. The other purposes of question bank are:

  - Question bank is useful to prepare a test for instant testing of the learners

  - Objective based test items in a question bank are helpful for the teachers to evaluate the learning progress with respect to learning objectives.

  - The learners can prepare themselves for the questions available in the question bank.

- The learners can also self-evaluate themselves by using the question bank.

- Questions not only help in assessment of learning, but also aid in classroom transactions for helping the students to learn better. Therefore, a variety of items on different learning outcomes should be available to both teachers and students in the classroom. Question banks in the classroom serve this purpose effectively.

- Every school should prepare question banks on their own. This brings ownership of the materials by the teachers and students of that school. While preparing the question bank and their proper use the following points should be taken into consideration.

- Both oral and written items on each chapter should be prepared.

- In each chapter, test items from knowledge, understanding, application and skill based objectives may be developed. Besides, project activities and practical activities should be there.

- After preparation and collection of questions those should be edited by the experts. Teachers from different schools or a cluster may sit and discuss each item and finalize them.

- It is better to write one or two questions on an item card - a postcard size paper, instead of a register. Cards have several advantages in developing, sorting, using and storing questions. Sometimes different cards containing questions may be supplied to different students in the class to engage everyone in learning.

- Different colour cards may be used for questions on different objectives and subjects. It will help the teachers to select and use the questions as per the purpose. For example-red colour cards may be used for knowledge type of items while blue and yellow colour cards may be used for comprehension and application objectives. The teacher may use different colours of cards for different purposes, like- different colours of cards may be used for extended response type of items (essay type of items), restricted response type of items (short answer type and objective type) and open ended items.

- **Project method:** Project method is a tool as well as a technique of educational assessment that is carried in a natural setting. Project creates

scope for learning concepts in a real life situation. This involves the application of knowledge. Projects are undertaken over a period of time and generally involve collection and analysis of data. Those provide opportunities to explore, work with one's hands by undertaking projects, the learners observe any situation or phenomena, collect data, analyse, organize and interpret data and draw generalizations. Project works provide opportunity to work in groups and real life situations. Projects help the learners to learn in an integrated approach. Here the assessment becomes an integral part of the routine classroom activities and the teaching learning process. The teacher has to observe the behaviour of the child during the execution of the project, his interest towards the work, process of collection, recording, interpreting the data. Accordingly the teacher can assist the learner and helps in improving the learning.

◆ **Portfolio:** Only the pencil-paper tests cannot assess all dimensions of learner's development, rather other modes of assessment are quite useful both for assessment of learner's progress as well as to ascertain the needs of the learner for further learning. Portfolio is one of the tools which can be used in assessment for learning. The students are encouraged to write or collect essays, poems, stories, collect paper cutting and give their own remarks on the articles/ issues, narrative or descriptive piece communicating a significant experience, riddles, mathematical puzzles, teaching-learning materials etc involving the concept. The students are also engaged in creating and collecting different products to complete the task. After a certain period all the students display their creations and describe them in detail in front of the teacher. "Portfolio is a purposeful collection of students' work that exhibits the students' efforts, progress, or achievement in a given area. This collection must include students' participation in selection of portfolio content, the criteria of selection, the criteria of judging merit, and the evidence of student self-reflection."(Reckase,1995). Thus, a portfolio provides an opportunity to the child to express his or her feelings and for the teacher to understand what is happening to the child outside the class. It is a collection of children's work over a period of time. It could be a day to day work or selection of the child's work. It provides a cumulative record. In the process emerges a picture of how a skill or knowledge develops. It encourages the child to show and tell

what they know and think about what they have included in their portfolio. The child becomes an active participant in learning and assessment.

◆ **Exhibition:** We can also use exhibitions to share and assess student's learning. Exhibition creates opportunities for the learners to show their talent in different subjects outside the formal classroom activities. Such type of activity not only fosters awareness among the learners but helps in skill building, developing positive attitudes among the learners. In the exhibition, the students learn certain concepts using concrete objects and verify many facts and properties using models, measurements and other activities. Thus, Spenser and Angus (1998) point out that student exhibitions involve complex cognitive skills as they must "collaboratively synthesize and evaluate information, and effectively communicate their ideas to others." Before organizing an exhibition, the teacher should discuss the time and venue of the exhibition. The children should be intimated well in advance of the exhibition so that they will get sufficient time to share among themselves, teachers and their guardians regarding their materials to be demonstrated in the exhibition. The children may prepare different models, materials, charts and interesting facts, puzzles etc. in the exhibition. The parents may also be given scope to participate. Different types of activities may be organized during the exhibition like demonstration of models, charts by the students and teachers. Photo exhibition(photograph of academicians reflecting their contribution) ,Teaching-learning material(TLM) preparation, popular talk on different concepts by the teacher or invited experts, demonstration of reference books in mathematics, activities for parents to show their talents and participate in different activities, different recreational activities for the students. This clearly indicates that exhibition can become helpful in formative assessment as well as in assessing the learner's ability to apply the acquired knowledge in different situations. Observing the nature of learners' participation during different activities of the exhibition, the teacher can assess the learners' understanding on a particular concept, his/her attitude towards learning. Further the teacher has to plan for further learning of the individual learner on the basis of the assessment findings. Moreover, it creates scope for learning from each other in an informal situation as well as peer assessment.

- **Quiz:** The dictionary meaning of the term quiz is 'series of questions testing people's general knowledge especially as a form of entertainment'. Normally in the quiz programme oral questions were asked to the participants and those were responded orally. But sometimes the respondents were allowed to use paper-pencil to get the answer. The question may be asked to an individual participant or to a group of participants (preferably 2 or 3). While conducting quiz the following points may be kept in mind:

- After teaching some concepts the quiz programme may be arranged, this will enable the learners to practice the concept and apply those in different situations.

- Questions based on real life situations may be asked to the learners

- Different rounds of answering like-answering using paper pencil, answering without paper-pencil, answering with clue and quick answer round may be conducted.

- Questions based on audio-visual support may be used to arouse interest among the learners.

- Questions prepared by the learners during the course of the teaching-learning process may also be used during quiz programmes.

- All the children may be allowed to participate in the quiz programme.

  Quiz programme enables the teacher to understand the student's progress in learning. It helps the teacher to know the learners interest towards learning. The teacher will observe how the students choose the answer and how they respond to the questions.

- **Games:** Game is a spontaneous activity for the children in which they participate naturally without any fear. Normally academic subjects are associated with fear and failure. But participation of the child in different types of games and interesting puzzles removes the fear from the child. Such games and puzzles enable the child to understand the concept thoroughly.

2.4 Teacher made Test in Educational Evaluation: Essay type, Short Answer type and Objective type

Teacher made tests are an important part of the teaching learning process. How far the instructional objectives, set by the teacher, are achieved by the student can be known by testing them. As we know, before starting the teaching-learning process, objectives are to be formulated and by employing an appropriate teaching learning process, the teacher and students jointly try to achieve the predetermined set of objectives. The objective based test item is supposed to measure a specific objective of instruction (learning outcomes). Such types of items describe the learners' achievement more accurately. The items are based on the specific objectives of a particular concept. Objective type of item is suitable for assessment of all the subjects. It is the nature of learning outcomes (specific objective) that determines the type of item to be used for assessment. For example, let us consider the following situations of assessing the objectives of mathematics learning. Use of open-ended items is now considered more effective in assessment for learning than the objective types. It is to be observed that the objective type of test item has a definite and unique answer that helps scoring the response easily and objectively. Such types of items are described as closed ended items. But there are test items which allow a variety of correct responses and elicit different kinds of students thinking. Such types of items are known as open ended items. This enables teachers with valuable information regarding individual students'way of thinking and way of solving problems and also helpful in enhancing students' creativity, reasoning and communicating skill. For the purpose of testing teacher follow different types of tests. These are as follows:

2.4.1 Essay type test

The essay type items are used to assess skills in organising and summarising the information and explaining the events of places. There is freedom in response in case of essay type tests. Essay questions are used primarily to measure those learning outcomes that cannot be readily measured by objective test items. Essay items assess students' understanding and ability to organise and apply information in a content area such as Literature and Language, Social Sciences like Geography, History, Psychology, Education, Economics, Political Science, and Sociology etc. the essay items also assess students' writing skill, knowledge of Grammar, knowledge of structuring a sentence and paragraph etc. The essay type tests are less valid and less reliable in comparison to the objective type tests. But if the essay type tests are prepared

consciously and scored with the help of scoring key, it will also achieve validity and reliability to a great extent. It is very difficult to standardise essay type tests.

Types: Essay type questions are classified into two types:

A. Restricted-Response essay questions: when dealing with the restricted response essay type questions, the students need to be very particular in their responding. These types of questions generally limit the content and response. The questions are followed with certain word limits. The restricted response essay type items are more structured and they are also most useful for measuring learning outcomes requiring the interpretation and application of data. For example, explain two most important effects of rotation of earth. The restricted response essay question also has its own limitations. It restricts the students' original way of expression, thinking. Organising and integration.

B. Extended- Response essay questions: This type of items need the students' response according to their freedom of organisation, integration, thinking and judgement. This type of freedom helps them to show their skill of analysing a problem, logically think over it, and finally describe it in their own words. This extended response essay items facilitates creativity, convergent and divergent thinking within the students. For example, explain the present day online education in India.

Advantages: The merits or advantage of essay type questions are as follows:

1. It measures complex learning outcomes that cannot be measured by other type of items.

2. It gives more importance on the integration and application of thinking and problem solving skills of the students.

3. It is easier to construct essay types of questions.

4. It evaluates the style of expression and the language knowledge of the students. It facilitates extensive study, convergent thinking and divergent thinking.

5. It develops the judgement ability of the students.

Disadvantages: Following demerits or disadvantages are also evident in essay type tests.

1. This type of question is less valid and less reliable in comparison to objective type of items.

2. Essay type items are subjective.

3. It encourages guess making tendencies within the students.

4. It facilitates cramming and rote memorisation.

5. It makes the total teaching learning situation a dictation work.

6. It is more time consuming at the time of scoring.

7. Examiners' physical health and mental health condition affects the scoring of the essay type items.

2.4.2 Short Answer Type Test

In the teaching learning situation, besides essay type and objective type questions we use the short answer type question to measure and evaluate students' performance. In this case the answer of students should be limited within a little limit of words. As it is difficult to ask essay type questions to the student over the whole content, the learning outcomes expected from the short answer type items are easier. Except for the problem solving outcomes measured in Mathematics and Science, the short answer item is used almost exclusively to measure the recall of memorized information. For example, write two effects of the industrial revolution. Although it is very difficult to measure the complex learning outcomes through short answer type tests. These are less time consuming in construction, administration and scoring.

Advantages:

The advantages of short answer type questions are listed below:

1. It is easy to construct.

2. It is easy to score.

3. These items are mostly used to measure the recall of memorized information.

4. It reduces the possibility that the students will obtain the correct answer by guessing.

5. Possibility of content coverage is higher

6. It is more valid as well as reliable than essay type questions.

7. The scope of short answer type items is more important than essay type items.

8. The discriminating power of short answer type items are more in comparison to the essay type items.

9. The administration of short answer type questions is easy in comparison to essay type questions.

Disadvantages:

The limitations of short answer type questions are listed below:

1. Short answer type items cannot measure the attitude and skills of the students.

2. The language skill and writing skill of students cannot be evaluated by short answer type items.

3. It gives more importance upon collection of information rather than understanding.

4. Short answer type items are not suitable for measuring complex learning outcomes.

5. This type of questions cannot be framed for all the objectives; cognitive, affective and psychomotor abilities.

6. It lacks in measuring the comprehension ability of students.

2.4.3 Objective Type Test

Objective type items are suitable for measuring a wide variety of relatively simple learning outcomes. Objective type items are more valid and more reliable than essay type and short answer type items. In case of objective type items the students are supposed to answer with a word, a digit, or a sentence. These are easy to construct, easy to administer and easy to score. The subjectivity in the form of personal preference of teachers or examiners does not affect the evaluation of objective type of tests. For example, who was the first prime minister of independent India? There are different forms of objective type items, the way of administration, the way of responding and the way of scoring is different in case of all the types of items. These are as follows:

a. Simple recall type question

b. Sentence completion type question

c. Classified question

d. True and false type question

   e.   Multiple choice type question

   f.   Matching type question

   g.   Analogy type questions

Advantages:

The merits or advantages of such type of items are:

1. These are more valid and reliable in comparison to other types of tests.

2. The usefulness of the test is higher

3. The discrimination index of this type of test is higher.

4. The major functions of these tests are measuring achievement, diagnosis, prognosis, research, counselling and guidance.

5. Scoring is simple and easy, mastery over subject for the evaluator is not necessary.

6. Items are standardised by developing norms.

7. Language and handwriting of the students do not have influence on scores.

8. Personal preference of teachers does not affect the scores.

9. It has high content coverage because a large number of items is included in this list.

10. In this type of test, students are supposed to study thoroughly all the contents.

Disadvantages:

The demerits of objective type tests are as follows:

1. Probability of guessing is higher

2. This type of test cannot be used effectively for affective and psychomotor learning objectives.

3. The students feel restless, when answering the question because in a short period of time a large number of questions are to be answered.

4. The language ability cannot be known through this test.

5. It gives importance to rote memory and craming.

6. Understanding aspects are neglected in this type of test.

7. The adoption of unfair means by students sometimes affects the evaluation system.

8. The instruction should be prepared for each type of question cautiously.

2.5 Techniques of Educational Evaluation

Viva voce, Interview, Group Discussion and Observation:

2.5.1 Viva voce

The oral examination, also called oral test or viva voce, is a practice in many schools and disciplines in which 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 to pass the test. Literally, "viva voce" means by or with the living voice - i.e., by word of mouth as opposed to writing. Normally viva voce is taken either by internal or external examiner where no one else is present in the examination hall except the examiner and examinee.

Purpose of the test: The purpose of the viva test is to:

- Demonstrate that the work or project is done by the examinee.

- Test the verbal communication skill, presence of mind and self confidence of the examinee.

- confirm that the examinee understands what he or she has written and can defend it verbally

- Investigate the originality of the project wok.

- Viva voce also reduces time and chances of adopting unfair means, chances of guessing habits.

- Beside the content knowledge of the examinee it also tests his or her command over the vocabulary and language development.

2.5.2 Interview

Interview is an important technique to elicit personal information from the subject. It is nothing but the face to face verbal exchange between interviewer and the interviewee. One-on-one interviews can be used effectively to assess students, especially when the

teacher wants to control the negotiating environment to test for certain skills and to assess students using an array of methods that do not overly bias grading towards students with strong writing skills. Furthermore, if assessment is also considered as a teaching moment, an interview of this type can provide an important simultaneous opportunity to explore what the students have learned, from the interview itself as well as from their ongoing real-time and real-life negotiations associated with the course. Interviews are normally tape recorded. If analysis, rather than just impression is required, then transcripts have to be produced. The transcripts are normally analysed by searching for responses or themes which commonly occur. Quotations from the transcripts can be used to illuminate or illustrate findings reported in reports and papers.

Purpose of Interview:

- For finding out the status of development of one's personality the personality assessment interview is conducted.

- For obtaining some information about the home, environment and school situation of the student the diagnostic interview is necessary.

- For determining the fitness of a person for admission, for a job, for scholarship etc assessment or evaluation interview is conducted.

- For helping the counselee or the student in gaining insight into the problem and assisting him or her while solving that problem the counseling interview is important.

Interviews are of two types: the structured interview and the unstructured interview. In a structured interview the interviewer is free to ask the candidates any question on any subject relevant to the situation in an unstructured interview. A systematic predetermined approach is adopted and all the subjects are asked similar questions in structured interviews. This type of interview has evolved to reduce the bias and unreliability of the interviewer. Usually it is a list of questions prepared beforehand and all the subjects are to answer these preplanned questions.

In addition to this, Boyd Fuller of National University of Singapore argues for using interviews for assessment. In a series of experiments, he finds that three distinct kinds of interview support assessment of different skills, different methods of testing those skills, and different lessons drawn from the experience by the student. One of the three, Fuller concludes, largely failed his goals for it; but the other two are more

promising. In common, the three approaches seek to assess students' skills in and/or understanding of negotiation through oral interaction rather than written product. They all differ from graded inter-student simulated negotiations in: 1) the students' performance is not dependent on their peers' actions; and, correspondingly, 2) the teacher can influence how the negotiation proceeds, in order to test certain skills. The first two methods were applied in the basic negotiation course; the third method, coaching interviews, has been applied only in an advanced negotiation course.

Mixed Scenario Interview Assessments: This interview assessment presented one or two brief negotiation scenarios, in which the teacher would ask the student to negotiate with him for five minutes. The other parts of the interview presented students with different scenarios and asked them to present their analysis of the situation, what they might do about it and why. During the exercise, no debriefing was provided, so that the answers were not being shared with other students. These interviews did provide some sense of how easily students could apply concepts and skills. It was clear, for example, that some students were quite familiar with the concepts and could use them to diagnose situations presented to them on the spot. Similarly, some students could apply the negotiation skills even when put on the spot while others struggled to do the same. Furthermore, there was some differentiation between what students achieved in their written assignments and what they demonstrated during the interview. Since one of the main purposes of the interview method is to provide people with weaker writing skills with a different medium for demonstrating their understanding. This outcome supported use of the tool. On the other hand, the students' negotiation skills were not truly tested. While the interview did allow students with better oral than written fluency to show their understanding, it still seemed to favour those who can think quickly. The negotiation scenarios were too short for them to move beyond a simple opening phase of a negotiation, and did not provide a chance for "warming up," including relationship building, the negotiation of communication dynamics, and so on.

Simulated Negotiation Interviews: Unsatisfied with the efficacy of the mixed scenario interview assessment, Boyd Fuller (2013) tried focusing the interview around a single negotiation scenario. The rationale here was: 1) to focus the interview assessment around skills rather than analysis; 2) to give some students the opportunity to "warm up," with their cognition that not all can start negotiations at the same speed; 3) to test a wider array of skills; and 4) to give students the opportunity to show their skills in more depth during a more difficult and complex negotiation. To

meet these objectives, instead of mixing the assessment of practice and analysis skills in one interview, he decided to focus on the practice skills, and assess students' analytical abilities through other assessment tools such as by assigning a quiz, or asking students to write a strategy memo or an analytical paper. The outcome of this interview assessment was satisfactory for him. The length of the interviews was sufficient to measure how the students were meeting the grading criteria. The focus on one negotiation as compared to multiple scenarios allowed people who were clearly slow starters to build momentum and show their skills over the duration of the negotiation. The set of skills that was tested turned out to be measurable by this scenario; they were distinctly different from what one sees in papers, examinations. It clearly distinguished some students from others. The debriefing after these interviews was also valuable. He could use his own direct observations of how they negotiated to provide comments.

Coaching Interview Assessments: This advanced course is in part designed around forcing students to apply their negotiation thinking and practice in intensive practice over the four years with fewer students, usually around a dozen. The coaching interview has two primary purposes. The first purpose is to get students to reflect on and design experiments to improve one or two aspects of their negotiation and facilitation repertoire of skills and behaviors. The second purpose is, of course, to assess those same skills. The first of the three coaching interviews is not assessed. One third of the assessment occurs during the second interview and two thirds during the third interview. This allows the students to develop their skills as reflective practitioners as the assessment increases. Each of the three interview sessions lasts approximately for forty five minutes. During the first interview, students were asked to identify a skill that they would like to improve and then a specific experiment or strategy that they will attempt as part of improving that skill or behavior. In the second interview, they were asked to analyze what happened during their experiment, why those reactions or outcomes occurred, and what they learned from that experience about why things can happen and what they can do better. The second half of the second interview then focuses on designing a new strategy or experiment, either for improving the same skill or another one that they have subsequently identified as something they want to work on. The third interview follows the format of the second, with the last five minutes focusing on terminating the class and coaching relationship while (again) encouraging the students to continue to be reflective negotiation practitioners. For this course, with its smaller numbers and greater focus

on skill and reflective practice, this assessment and teaching tool is an essential component. As a learning tool, these coaching interviews are the most effective. While not all students make good use of them, those students who do commit to applying the strategies they develop report and demonstrate much greater negotiation skill afterwards. Significant improvements have been visible over the three sessions, as well as during the class negotiation sessions and in real-life negotiation scenarios where students conduct their experiments elsewhere. As an assessment tool, they are useful for assessing pedagogical goals of helping students become reflective practitioners. It is clear by the third interview which students have begun to analyze their negotiation experiences on their own and which students are more capable at it

2.5.3 Group Discussion

"Group" is a collection of individuals who have regular contact and frequent interaction and who work together to achieve a common set of goals. "Discussion" is the process whereby two or more people exchange information or ideas in a face-to-face situation to achieve a goal. The goal, or end product, may be increased knowledge, agreement leading to action, disagreement leading to competition or resolution or perhaps only a clearing of the air or a continuation of the status-quo.``Group Discussion", popularly labeled as GD, is a popular methodology used by educational institutes and many organizations (company, institute, business school, etc.) these days to gauge whether the candidate has certain personality traits such as interpersonal communication skills, confidence in public speaking, team spirit, leadership abilities, social behaviour and problem-solving skills. GDs form an important part of the short-listing process for recruitment or admission in a company or institution. In this methodology, there are usually 7-12 participants in a group. The group of candidates is given a topic or a situation which could be either factual, abstract or case based, and typically given some time to think and make notes about the same. After this, the group of candidates is asked to discuss the topic among them for a specific duration ranging between 10-40 minutes. While the group discusses the pertaining issue at hand, the Moderators or Panelists silently observe each candidate on various predetermined parameters. The Panelists assign scores to every candidate based on his or her individual performance as well as how he performed within the group.

GD is based on teamwork, incorporating views of different team members to reach a common goal. So, a group discussion refers to a communicative situation that allows its participants to share their views and opinions with other participants. It is

a systematic exchange of information, views and opinions about a topic, problem, issue or situation among the members of a group who share some common objectives. Over the recent years, Group Discussion became a popular method of assessing students' soft skills. The contenders who are shortlisted on basis of written examination have qualified with their intelligence quotient, i.e., aptitude and knowledge. However, since the significance of emotional quotient arose, new tools such as GD were devised to gauge candidates' social and interpersonal skills. Organizations conduct GDs to find out whether you possess the critical qualities or skills to contribute effectively to the goal accomplishment process.

Purpose of Group Discussion:

- It helps to assess how a participant performs under different situations in a group.

- It helps to judge how one conceptualizes and establishes his ideas through the discussion.

- It helps in analysing the students' attitude towards fellow members through one's communication and interpersonal skills, listening ability, humility and tolerance to others ideas.

- It helps in shedding light on candidate's leadership and managerial skills, problem-solving aptitude, creative thinking and knowledge on diverse topics.

- It helps evaluate whether a candidate is the right fit for the organization.

It is basically a situation test wherein a sample of a candidate's group worthiness and potential as a worker comes out quite explicitly. it is helpful for achieving group goals. The examiner can evaluate both the personality and group skills of the candidate participating in a group discussion.

2.5.4 Observations

Observation can be invaluable when conducting an evaluation or it can also be of limited value if it is not done well. Those who study human behaviour indicate that there is often a gap between what people say they do and what they actually do. The observed behaviour is often a more dependable indicator than what is self-reported. Observation is used in a variety of ways in evaluation. Often it is a gateway method that leads to others methods. Careful observation is distinctive in three important ways: the person doing the observation is (1) trained, (2) prepared, and (3) systematic.

Training helps the evaluator know what to observe, when to conduct the observation, and how to document the observations. The evaluator must be prepared mentally and physically for the rigors and demands of observation. And the evaluator systematically captures the observation data often using multiple methods such as field notes, checklists, audio memos, etc. Observation can be time consuming and difficult. Here are four factors that limit our ability to observe:

1. **Fatigue -** Observation is tedious and tiring work. Being attentive takes energy and the observer needs to develop strategies for remaining alert for potentially lengthy periods of time.

2. **Emotional stress -** Sometimes the observation itself causes stress and other times the person doing the observation is stressed about other things. Stress tends to erode our ability to observe.

3. **Disruptions and distractions -** Observation takes time. Therefore, disruptions and distractions should be expected in the course.

4. **Time -** The length of the observation is specified in the evaluation plan and is based on a thoughtful assessment of what is being observed, the expected occurrence of meaningful events, and the available resources. In general it is better to have a longer period of observation because it increases the odds of observing activities or events that occur less frequently. A teacher needs observation techniques to improve teaching and learning, as it has an important role in collecting data and evidence about the teaching process and student learning. Followings are some of the techniques which can be used for the said purpose.

Student Assessment: Tests, examinations and continuous assessment can provide valuable data for observation. For teaching courses, we have to set up a method of student assessment and students have to be assessed. We should, however, be clear about the nature of the information we can obtain from examination results or assessment grades. Comparison of one set of results with another often has limited validity as assignments, examinations, markers and marking schemes are rarely held constant. In addition most assessment is norm referenced rather than criterion referenced. We also need to be very clear as to what is being assessed. Examination grades may bear little relationship to specific qualities we could be investigating.

Interviews: Interviews can provide even more opportunity for respondents to

raise their own issues and concerns, but are correspondingly more time-consuming and can raise difficulties in the collation and interpretation of information. The format can be on a spectrum from completely open discussion to tightly structured questions. Semi-structured interviews have a small schedule of questions to point the interviewee towards an area of interest to the researcher, but then allow interviewees to raise any items they like within the general topic area. Since interviews give an opportunity for students to raise their own agenda they are useful when issues are open, or at an exploratory stage. A small number of interviews can be useful to define issues for subsequent more tightly structured questionnaires.

Closed Ended Questionnaires: Closed ended questionnaires are ones which constrain the responses to a limited number chosen by the researcher; essentially it is a multiple choice format. Usually respondents are asked the extent to which they agree or disagree with a given statement. Responses are recorded on a Likert scale which ranges from 'definitely agree' to 'definitely disagree'. Questions should be carefully constructed so the meaning is clear and unambiguous. It is a good idea to trial the questionnaire on a limited number of students before giving it to a whole group. Closed questionnaires are easy to process and evaluate and can give clear answers to specific questions. However, the questions are defined by the researcher or teacher so could completely miss the concerns of the respondents. Therefore we can include some open-ended questions to give respondents a chance to raise other issues of concern. Most institutions now have some form of standard teaching evaluation questionnaire available. These may be of some help in evaluating a project but in most cases the questions will not be sufficiently specific to the particular type of innovation which has been introduced. What might be more helpful are the data banks of optional or additional questions which are available. These can be used to pick or suggest questions which might be included in a more tailor-made questionnaire. Traditionally, collecting questionnaires, survey data is done by using copies of paper questionnaires and answer sheets. With the availability of web technology, there is now the option of collecting survey data online.

Open Ended Questionnaires: Open questionnaires have a series of specific questions but leave space for respondents to answer as they see fit. Teachers are therefore more likely to find out the views of students but replies are more difficult to analyse and collate. The usual procedure is to search for categories of common responses. It is not necessary to have separate questionnaires for open and closed items. The most successful questionnaires often have both open and closed items.

Diary: Every teacher, involved in an action learning project should keep a diary or journal in which they record their initial reflections on the topic of concern, the plans that were made, a record of actions which were taken, observation of the effects of the actions, impressions and personal opinions about the action taken and reactions to them, results obtained from other observation techniques, references for, and notes on, any relevant literature or supporting documents which are discovered. By sharing observations and reflections it is possible to fine-tune the innovation. Sympathetic but critical discussion can also heighten awareness and contribute to changing perspectives.

Supporting Documents: Teacher should keep copies of any documents which are relevant to the courses that teacher is examining. These can include documents for the course development and accreditation process, minutes of course committees, the course syllabus, handouts to students, copies of tests and examinations, lists of test results and student grades.

Student Learning Inventories: Student learning inventories are examples of empirically derived measuring instruments. There are many number inventories which measure a wide range of characteristics. Student learning inventories have been highlighted because they examine the quality of learning. In particular they look at the categories of deep and surface learning. The inventories can be used to compare groups of students, examine approaches before and after changes to teaching methods, and to examine correlations with other variables.

Diagnosis of Student Conceptions: A good basis for improving teaching is to diagnose students' understanding of key concepts in a course. The usual method of diagnosing student conceptions is to ask a question which applies the concept to an every-day situation: one which cannot be answered by reproduction or by substitution into formulae. Answers are drawn from the students in interviews or in written form. The students' answers can usually be classified into a small number (usually two to five) of conceptions or misconceptions about the phenomenon. As with the analysis of interview data care needs to be taken when deriving classifications.

Interaction Schedules: Interaction schedules are methods for analysing and recording what takes place during a class. A common approach is to note down at regular intervals (say every minute) who are talking, and to categorise what they were saying or doing. An alternative to time sampling is event sampling in which behaviour is noted every time a particular event occurs. Examples of categories could be; tutor asking question, tutor giving explanation, tutor giving instruction, student answering

question or student asking question. The analysis can be made by an observer at the class or can be made subsequently from a tape or video recording. There are other approaches to recording and analysing happenings in a classroom situation. McKernan (1991) discusses an extensive range of techniques, gives examples of each and considers how the data gathered should be analysed. Flanders Interaction Analysis category is another method of evaluation of students' behavior in the classroom which is as follows:

	Behaviour type	Category	Verbal behaviour
Teacher Talk	Indirect/ response	1	Accept feeling
		2	Praises/encourages
		3	Accepts /uses ideas
		4	Asks questions
	Direct/ initiation	5	Lecturing
		6	Give directions
		7	Criticising or justifying authority
Pupil Talk	Indirect/response	8	Pupil talk response
	Direct/ initiation	9	Pupil talk initiation
Silence/confusion		10	Silence or confusion

Tape Recording: Making tape recordings is a way of collecting a complete, accurate and detailed record of discussions in class, conversations in interviews or arguments and decisions at meetings. It is easy to obtain the recording; we simply take along cassettes and a portable recorder, and switch it on. However, the presence of a tape recorder can inhibit discussion or influence people's behaviour. There are a number of ethical issues which need to be addressed over the use of tape recordings.

The group being taped should establish the purpose of making the recording and the way in which the tapes will be used. If any quotations are made in subsequent reports it is customary to maintain the anonymity of the source. If we need to do a detailed analysis of the conversations then it will be necessary to produce a transcript. This is a time-consuming and painstaking process, so limit the use of tape recordings to situations where it is really necessary.

2.6 Summary

The teacher needs to assess the students' progress in learning and to help him for further learning on the basis of the assessment results. There are different tools and techniques available for educational assessment. The teacher may use different types of test items as useful tools for this purpose. Teacher made tests which are constructed by the teachers for use within their classroom are quite useful for this purpose. The progress and achievement of learners can be known by assessment, as well as the effectiveness of the teaching-learning process, materials can also be ascertained by assessment. Assessment should not be a one time event, rather be a continuous one. Use of unit tests, giving assignments, observing the child during the teaching-learning process and preparation of TLM, project work and portfolio makes the assessment process continuous. Comprehensive assessment means assessment of both scholastic and co-scholastic areas of the curriculum. To assess both the areas the teacher has to use a variety of tools and techniques. The objective based test items are supposed to measure a specific objective of instruction (learning outcomes). Such types of items describe the learner's achievement more accurately. The teacher should prepare and use knowledge, comprehension, application and skill based questions. Question bank can help the teacher to conduct a test instantly and also enables the learners for self study and self evaluation. Participation of the learners in activities like project, portfolio, exhibition, quiz, games helps the learner to learn in an informal way. Such activities help the teacher to assess the learner in an informal situation. There are also standardized tests, psychological tests, written tests, oral tests, diagnostic tests.

Written tests may be of different types; essay type, short answer type and objective type. Usually an essay test refers to any written test that requires an examinee to write several paragraphs, asks the students to write long answers to one question. This gives the students training in self expression, use of proper word, organizing subject matter, and arguing logically. Short answer type questions require very short answers generally one or two sentences. While, objective test is a test containing questions

requiring the student to work or select a correct answer.

For educational evaluation we depend on techniques like viva voce, interview, group discussion and observation. The teacher has to observe the learners during activity and assess to what extent the learner has acquired the knowledge in a particular concept. Classroom observation is another form of ongoing assessment. Most teachers can "read" their students; observing when they are bored, frustrated, excited, motivated, etc. As a teacher picks up these cues, she or he can adjust the instruction accordingly. It is also beneficial for teachers to make observational notes (referred to as anecdotal notes). These notes serve to document and describe student learning relative to concept development, reading, social interaction, communication skills, etc. Other than this use of diary, journal, tape recorder, questionnaire, students' learning inventories, interaction schedules are also used by teachers for observation purposes. Through those activities the assessment process can be made child friendly. The principles of construction of teacher made tests will be discussed in the next subunit.

2.7 Self Assessment Questions

1. Name the tools commonly used by teachers for educational evaluation.

2. Classify essay type tests with examples.

3. State any four utilities of question banks in mathematics. Write two more utilities which have not been stated here.

4. Enlist three Social Science projects you want to give learners at secondary level.

5. Prepare a list of different materials that can be put together while preparing a portfolio.

6. Write some of the usefulness of an exhibition for the learners?

7. Write down the disadvantages of short answer type tests.

8. State the purpose of viva voce for educational evaluation.

9. Why is group discussion popular?

10. State the importance of anecdotal records as an observation technique of evaluation.

2.8 References Books

- Aggarwal, J.C. (2001): Essentials of Examination System: Evaluation, Test and Measurement, Vikash Publications, Delhi, page no. 70-72, 261-262

- Dash, B.N. , Dash, Nibedita and Rath, M.P.(2003): A Textbook of Educational Psychology, Dominant Publishers, Delhi, India, page no.504-509

- Lovely Professional University (2012): Educational Management and Evaluation DEDU504, USI Publications, New Delhi, page no.274-281, 323-351

- Rout, Santosh Kumar (2009): Educational Measurement, Evaluation and statistics, Navyug Publishers and Distributors, New Delhi, ISBN 978-81-906940-8-7, page no. 22-33, 35-47

- Sarkar, Amal kanti (2010): Measurement and Evaluation in Education, Kolkata, India, Rita Book House, page no 48-60, 61-66, 111-114, 214-216

- Singh, A.K. (2002): Tests, Measurements and Research Methods in Behavioural Science, Patna, Bihar, India, Bharati Bhawan Publishers and Distributors, page no 193-210

- Singha, H.S. (1999): Modern Educational Testing, New Delhi, India, Sterling Publication Private limited, ISBN812-07-217-13, page no 96- 99, 111-114, 128-139

Web References:

- http://ignoustudent2010.blogspot.com/p/bed-1st-year-es-333.html#:~:text=Assessmen t%20refers%20to%20the%20collection, for%20the%20purpose%20of%20describing %2C retrieved on 22.06.2020

- http://mooc.nios.ac.in/mooc/pluginfile.php?file=/11455/course/summary/ UNIT%2011%20T OOLS%20AND%20TECHNIQUES%20OF.pdf, page 19-43 retrieved on 24.06.2020

- https://gdpi.hitbullseye.com/Group-Discussion.php retrieved on 11.06.2020

- https://www.yourarticlelibrary.com/statistics-2/teacher-made-test-meaning-features-and-uses- statistics/92607 retrieved on 24.06.2020

- https://www2.le.ac.uk/departments/doctoralcollege/training/eresources/study-guides/viva/abo ut retrieved on 10.06.2020

Unit 3 ❑ Achievement Test Construction

Structure

3.1 Objective

On completion of this unit students will be able to—

- Define testing

- State the characteristics of a good test

- Explain the factors of objectivity, reliability, properties of validity and aspects of usability, classify norms.

- Identify the problems of teacher made test

- Define the principles and steps of test construction

3.2 Introduction

A test is a particular type of assessment that typically consists of a set of questions to be answered that are administered under reasonably comparable conditions for all students. A test is advice or systematic procedure for assessing a sample of behaviour by asking a set of questions in a uniform manner. The scientific phase of testing and measuring movement starts with the work of E.L.Thorndike. Thorndike constructed and published a pioneer psychological test. Then, the followers like Stone published tests measuring achievement in Elementary Arithmetic in 1908 and Courts published it in 1909. The testing movement seems to have reached its zenith in the twentieth century. The characteristics of a good test will be discussed in this unit. Along with that the problems of the teacher made test and principles of construction of achievement test will also be discussed.

3.3 Characteristics of a Good Test: Objectivity, Validity, Reliability, Usability and Norms

A test is a form of assessment. Psychological and educational tests are a standardized procedure to measure quantitatively and qualitatively one or more than one aspect or trait by means of a sample of verbal or non verbal behaviour. A test consists of a standard set of questions, to be answered or a task to be performed. Items of a test are placed in increasing order of difficulty and its procedure of administration is standardized to ensure maximum objectivity. In the teaching learning situation, psychological and educational tests are used to know about the level of achievement of the students. A Dictionary of Education (1981) defined as "a test is a compact task or series of tasks designed to ascertain the merit or quantity of something. Educational tests constitute a series of items for which a score is obtained. Depending on how they are constructed, they can serve a purpose." A good test must have the following five characteristics.

3.3.1 Objectivity

A test must have the trait of objectivity in terms of both item and scoring. It means it must be free from the subjective elements so that there can be complete interpersonal agreement among the experts regarding the item and scoring of the test.

Objectivity of items refers to the phrasing of items in such a manner that they are interpreted in exactly the same way by all those who take the test. For that, items

must have uniformity of order of presentation. It should be placed either in ascending or descending order. On the other hand, objectivity of scoring means that the scoring method of the test should be standard one to maintain complete uniformity when the test is scored by different experts at different times. A test is said to be an objective if the same scores of results are obtained by administering the test to a particular group or students on different occasions and scored either by the same scorer or different scorer. A test that is objective measures without reference to outside influences. For example, an objective test of personality will return the same answers regardless of whether the person completing the test uses a pen or pencil. Irrelevant, unrelated factors do not influence the test results if a test is objective.

In objective type tests the resulting scores are in no way influenced by the mental condition, judgment, personal bias or feelings of the scorer. Items are prepared in such a way that there will be only one correct answer of the scorer. Whoever may administer or score a test on a group of students the test score will always tend to be the same. This particular characteristic is said to be the objectivity of the test. For example, name the present capital city of West Bengal. The obvious and one and only correct answer is Kolkata. The score will always be the same under any circumstances.

Therefore, it is essential to preserve and maintain the objectivity of the test. Sometimes it becomes difficult to retain this characteristic. Following are the causes of low objectivity of a test

- Lack of comprehensiveness and too much generalisation reduce the objectivity

- Ambiguity in question framing leads to confusion and misinterpretation which may reduce objectivity of the item.

- Sometimes a student who gives the best answer at the beginning of the test is likely to get more marks in the subsequent questions and vice versa.

- Sometimes a particular student gets marks not only on the basis of his or her writing ability but in comparison to ability of the previous students.

- Lack of uniformity in scoring by the examiners in absence of a scoring key can be detrimental for objectivity.

- Lack of consistency in judgment even among competent examiners may lead to subjectivity.

- Subjectivity may also occur due to application of language, quality of handwriting of the examinee.

● Sometimes the examiners are influenced by the volume and length of the answer rather than depth of the content.

To minimize these subjective elements it is required to reform the conventional essay type questions and to introduce new types of objective tests. The term essay implies a written response that may consist of many sentences to several pages. The students are allowed to enjoy freedom with respect to the content, construction, wording, length and organization. Essay type questions, therefore, definitely have some unique characteristics to judge several qualities of students that otherwise are difficult to evaluate. These are the ability to select relevant facts from a body of acquired knowledge, establish relationships between various aspects of knowledge, analysis, synthesis, assimilation, expression, interpretation, organization, critical thinking and so on. In spite of so many advantages essay type questions are discarded due to its lack of objectivity in both item and scoring process. However, some precautions can be taken to reduce the subjectivity of a test.

a. **Construction of question:** The scope and range of wording should be clear. It is needed to define and restrict the area to be covered by the question. The questions must be related with the instructional objectives of the units in order to achieve maximum content validity. The value points in a question should be intimated to students. Optional questions should not be provided because it reduces the comparability and also makes it difficult to score papers on a common basis.

b. **Scoring system:** An answer key should be provided by the paper setter with which all answers can be compared for evaluation. For each question the factors to be considered in evaluating the response should be given. The students' identity should not be disclosed in order to avoid subjectivity.

3.3.2 Validity

Validity is another characteristic of a scientific instrument. The term "validity" means truth. Thus, validity refers to the degree to which a test measures what it claims to measure. Validity is not the self correlation of tests like reliability; rather it is correlation with some outside independent criteria.

Concept: Validity has been defined in different ways. Anastasi (1968) has defined "the validity of a test concerns what the test measures and how well it does so." Lindquist (1951) has defined validity of a test "as the accuracy with which it measures that which is intended to measure or as the degree to which it approaches infallibility

in measuring what it purports to measure." These two definitions reveal the fact that for determining the validity of a test the test must be compared with some ideal independent measures of criteria. Thus, correlation coefficient computed between the test and the ideal measures of criteria is known as the validity coefficient.

As we have already noticed that Validity reflects the degree to which an evaluation device approaches infallibility in measuring in what it intends to measure. It means to determine how valid a tool is one should compare the reality of what it does measure with some ideal conception of what it ought to measure. The measurement should be consistent. Thus, validity has two aspects: Reliability and Relevance. Reliability is discussed earlier. It appears that reliability is a necessary, though not a sufficient, condition for validity. An evaluation tool cannot be valid unless it is reliable, but it may be reliable still not valid. The latter may happen if the tool lacks relevance. Relevance concerns the closeness of the agreement between what the device measures and the function it is used to measure. Generally we need a criterion to define what the evaluation instrument should measure. The validity of the tool can be estimated by the correlation between the raw scores from it and the true criterion scores. The relevance of the test is an estimate of the correlation between the true scores from the evaluation device and the true criterion scores. It is interesting to note that no device is valid for all purposes in all situations and for all groups of students. In fact, validity is always related to the purpose. A device which is valid for one purpose may not be equally valid for any other purpose. Validity is not a problem in physical measurement. The relevance is not questionable because the features and traits are well defined. So, they are concerned with reliability of measurement. But in behavioural science like Education and Psychology, relevance becomes tremendously important as the functions to be measured can seldom be well defined and measurement has to be mainly indirect.

In a broader sense, validity is concerned with generalization. When a test is a valid one its conclusion can be generalized in relation to the general population. From the above discussion it can be said that Validity has three important properties. These are:

- Validity is a relative term. A test is valid for a particular purpose. For example, a test of mathematical ability will be worthless for measuring the knowledge of History.

- Validity is not a fixed property of the test. Validation is a changing and unending process. The old content of the test becomes less meaningful with

the discovery of new concepts and formulation of new meanings. Therefore, they need to be modified radically in the light of the new meaning.

● Validity, like reliability, is a matter of degree and not an all-or-none property. A test meant for measuring a particular trait or ability cannot be said either perfectly valid or not valid at all.

3.3.3 Reliability

Reliability refers to the consistency of a measure. A test is considered reliable only if we get the same result repeatedly. If a test is applied to an individual or to a group of students on two or more occasions and the results obtained in all occasions do not differ at all or differ in a very little degree it can be said that the test is a reliable one.

Concept: Reliability in statistics, Education and psychometrics is the overall consistency of a measure. A measure is said to have a high reliability if it produces similar results under consistent conditions. It is the characteristic of a set of test scores that relates to the amount of random error from the measurement process that might be embedded in the scores. Scores that are highly reliable are accurate, reproducible, and consistent from one testing occasion to another. That is, if the testing processes were repeated with a group of test takers, essentially the same results would be obtained. Various kinds of reliability coefficients, with values ranging between 0.00 (much error) and 1.00 (no error), are usually used to indicate the amount of error in the scores. For example, measurements of people's height and weight are often extremely reliable. According to Anastasi (1968) reliability refers to the "consistency of scores obtained by the same individuals when re-examined with test on different occasions or with different sets of equivalent items, or under other variable examining conditions". Consistency in results obtained in a single administration is the index of internal consistency of the test. Whereas, consistency in results obtained upon testing and retesting is an index of temporal consistency. Reliability thus includes both internal consistency as well as temporal consistency. A test can be called sound only when it is reliable because reliability indicates the extent to which the scores obtained in the test are free from such internal defects of standardization that are likely to produce errors of measurement. In other words, the test is free from variable errors. Therefore, a serviceable degree of reliability is another essential criterion of a good test. There are several general classes of reliability estimates:

● Inter-rater reliability assesses the degree of agreement between two or more raters in their appraisals.

- Test-retest reliability assesses the degree to which test scores are consistent from one test administration to the next. Measurements are gathered from a single rater who uses the same methods or instruments and the same testing conditions. This includes intra-rater reliability.

- Inter-method reliability assesses the degree to which test scores are consistent when there is a variation in the methods or instruments used. This allows inter-rater reliability to be ruled out. When dealing with forms, it may be termed parallel-forms reliability.

- Internal consistency reliability assesses the consistency of results across items within a test.

Reliability in its simplest sense refers to the precision or accuracy of the measurement or score. A well constructed scientific instrument should yield accurate results both at present as well as over time. reliability is actually a type of consistency. a test is said to be consistent over a considerable period of time when all the examinees retain their same relative ranks of two seperate testing with the same test.

Reliability is never the property of the test itself. Rather, it is the property of a test when it is administered to the examinees. In other words, it is the property of test scores. Thus, Reliability refers to the results obtained with an evaluation instrument and not to the instrument itself.

The correlation coefficient of internal stability is called coefficient of stability. Similarly, the correlation coefficient indicating internal consistency is called coefficient of internal consistency or alpha coefficient. Any statistical measure of reliability must indicate bothe the coefficient of stability as well as the alpha coefficient. That is why reliability refers to self correlation of the test.

Reliability of a device is the degree to which the device and its scores reflect true or non-error variance. Or it can be said that it reflects the degree to which the scores are free from chance or random errors. To make the devices more reliable we should avoid the occurrence of chance error in the scores. This nature of reliability also subsumes objectivity because the latter is due to one type of chance errors at the time of passing judgement. That is why objectivity of measurement is studied with reliability.

Reliability and validity are the two dimensions of test efficiency. Reliability is a matter of stability of test scores whereas; validity is the correlation of the test with certain outside independent criteria. Therefore, Reliability is a necessary but not a

sufficient condition for validity, because tests possessing poor reliability will yield low validity. Again, a test may be reliable but not valid. It may yield a consistent score but the score need not be representing what exactly we want to measure. Reliability merely provides the consistency that makes validity possible.

Factors of low Reliability: The Reliability of test scores is influenced by a large number of factors. These factors can be categorized into three heads: Extrinsic, Intrinsic and method used for determining reliability.

1. **Extrinsic factors:** Extrinsic factors are those which lie outside the test itself but tend to make the test reliable or unreliable. These are as follows:

   a. **Group variability:** When the group of examinees being tested is homogeneous in ability, the reliability of the test scores is likely to be lowered. The effect of variability on reliability can be examined by seeing what happens when variability is zero. In that case, individuals receive the same score; standard deviation and Z score become zero.

   b. **Guessing and chance error:** Guessing in a test is an important source of unreliability. In two alternative response options there is a 50 percent chance of answering the items correctly on the basis of the guess. In multiple choice items the chances of getting the answering correct purely by guessing are reduced. Guessing has two important effects upon the total test score. First, it tends to raise the total score making reliability coefficient high. Second, guessing contributes to the measurement error since the examinees differ in exercising their luck over guessing the correct answer.

   c. **Environmental conditions:** It is preferable to maintain a uniform testing environment. Infrastructure facilities like light, sound and other comforts should be arranged equally and uniformly to all the examinees. Otherwise, it will tend to lower the reliability of the test scores.

   d. **Momentary fluctuations:** Momentary fluctuations in the examinees influence the test score. Hence, they tend to affect reliability. A broken pencil, momentary distraction by the sudden sound of some microphone, anxiety regarding the non completion of test, mistake in giving answer, feeling fatigue or other emotional and physical problem are some of the factors that explain momentary fluctuation in the examinee, leading to lower reliability of test.

2. Intrinsic factors: It refers to those factors which lie within the test itself and influence the reliability of the test. These are as follows:

a. **Length of the test:** The length of the test or size of the sample is directly proportional with the reliability coefficient. Generally, shorter tests lead to lower reliability due to increasing sampling error. Lengthening the test or averaging the total test scores obtained from several repetitions of the same test tends to increase the reliability.

b. **Homogeneity of items:** Homogeneity of items is an important factor in reliability. The concept of homogeneity of items includes two things.-item reliability and homogeneity of function or trait measured from one item to another. When the test is a heterogeneous one the reliability is zero or very low.

c. **Range of the total score:** If the obtained total scores on the test are very close to each other or if there is lesser variability among them the reliability of the test is lowered. The standard deviation of the total score is low, the reliability is also low.

d. **Difficulty value of items:** In general items having indexes of difficulty at 0.5v or close to it, yield higher reliability than items of extreme indexes of difficulty. When items are too easy or too difficult the test yields very poor reliability, as these items do not contribute to the reliability.

e. **Discrimination value:** when items do not discriminate well between superior and inferior, or when items have poor discrimination values the item total correlation is affected, this leads to low reliability of the test.

f. **Scorer reliability:** It, also called reader reliability, is an important factor which affects the reliability of the test. It means how closely two or more scores agree in scoring or rating the same set of responses. If they do not agree, the reliability is likely to be lowered.

g. **Ambiguity of language:** Vague language coupled with unstructured nature of tasks presented to the students with no or improperly planned directions for scoring become the root cause of personal errors and subjectivity. The degree of seriousness of the problem obviously varies from subject to subject. For example, it is definitely less serious in Mathematics than in Social Science.

3.3.4 Usability/ Practicability

Practicability is another important characteristic of a good tool. It stands on a different area from the technical considerations discussed so far, but it has its own significance. Practicability imposes conditions on reliability and validity. In other words, reliability and validity depends on practicability. As we go on increasing the length of the test the reliability of the test goes on increasing. Although we may not be able to increase the length of the test beyond a certain level due to practical limitations. Thus the evaluation tools and techniques are likely to be ineffective if it lacks practicability. Some of the aspects of practicability which are to be taken into consideration at the time of construction of the test are economy, administrability, scorability, interpretability, utility and acceptability. These are discussed in detail in the following paragraphs.

- **Economy:** Economy related to time, cost and effort should be kept in mind during construction of the test. The device should undoubtedly be prepared by the teachers. A difficult tool will immediately invite resistance from the teachers for its use. It should not take away too much of teaching time so that the instruction itself suffers. As far as printing and other charges are concerned the device should be well within the easy reach of the budget of the institution.

- **Administrability:** The tools and techniques of evaluation should be easy to administer. They should not require expertly trained personnel to administer the test. It should be teacher friendly. So that the classroom teachers are able to handle the test efficiently. Moreover, the device should be such that it can be administered within the framework of the institution's timetable. It should have clear and unambiguous directions both for examiners and pupils. It should be in conformity with the format.

- **Scorability:** Objectivity and ease are the two aspects of scorability. Objectivity of scoring is an important part of reliability and therefore, should be taken into consideration while ensuring reliability as already discussed earlier. Steps will have to be taken to make easy scoring in the evaluation tool. This is mainly concerned with degree of expertness and time required for scoring. The degree of expertness is determined by the form of questions. Form of question is controlled by the instructional objectives, based on which the tool is constructed. It also depends on the purposes of evaluation. Therefore, the constructor has to determine the difficulty level of the tasks keeping in mind the purpose, ability of the pupils and instructional objectives of the test. Some steps can be carefully taken to reduce the time for scoring.

- **Interpretability:** Ease of interpretation of the score is another crucial factor that must be considered at the time of construction of the test. A good device possesses interpretability only when its scores can be given meaning in objective terms by the use of derived scores, norms, ratings or other methods. This is an essential feature of standardization of the test.

- **Utility:** An evaluation tool will have utility if it gives rich dividends to the users for example students, teachers, institutions etc. If the scores can be employed effectively for the guidance of pupils, improvement of instruction, improvement of educational programme, management, administration and betterment of the institution.

- **Acceptability:** Acceptability of the evaluation device among students, teachers, administrators, policy makers and the public at large is a very important consideration under practicability of a test. Evaluation is visualized as a cooperative process. The test may have an excellent technical and academic base, but if the teachers, students and other concerned stakeholders express resistance to accept a certain evaluation device, it certainly loses its practicability. For example, internal assessment procedures, although theoretically technically and academically very sound, are not always acceptable to students, teachers and the public in general. Acceptability is mainly a function of the environment. The use of certain types of evaluation devices or certain forms of question in them should be governed by the prevalent socio economic condition and reactions. For example, if general atmosphere of malpractice prevails, multiple choice question tests should be avoided as they are highly amenable to code signaling.

3.3.5 Norms

A test must also be guided by certain norms. Norms refer to the average performance of a representative sample on a given test. Norms increase the usability of a test. Two important definitions are discussed here. According to Frank S. Freeman, norm is the average or standard score in a particular test made by a specified population. Thorndike and Hagen defined norm as the average performance on a particular test made by a standardization sample. Norms help in interpretation of scores. The term norm conveys several meanings. Some important meanings are enumerated below:

- It is a statistical procedure to minimise the interpretive error in a test score. It is a device of transforming raw scores to standard scores.

- It is meant to determine the position of its score

- Norms are average score or values determined by actual measurement in a group of persons who are representative of a specified population.

- Norms reflects development under condition that may be or often less than optimal

- Norm is the standard level against which the value of certain activity can be compared.

The performance or score of an individual is compared against a determined level.

There are four common types of norms-age norm, grade norm, and percentile and standard score norms. A test constructor prepares any one of these norms depending upon the purpose and use for his test.

Age norms are based on average scores earned by pupils at different ages. They are interpreted in terms of age equivalent. Test performance is expressed in age level. Age norms are used in mental ability tests (mental age), achievement test in arithmetic (arithmetic age). Age norm represents test performance in unequal units. The school year is divided into 12 months. Age units do not have uniform meaning due to variation in growth pattern in different age levels. Age norms can be more meaningful in elementary level, where growth of children's ability is more regular and continuous, than at the secondary level of education. Age norms do not always represent the child's mental ability as the child's mental age does not always confer its chronological age.

Grade norms are widely used in standardised achievement tests as the tests are employed within academic settings. These educational tests are based on average scores earned by pupils in each grade and interpreted in terms of grade equivalence. Grade equivalence is expressed in two numbers. The first is year and second is month. The school year is divided into 10 months assuming little or no changes occur during vacation. Therefore, grade equivalence for grade 4 ranges from 4.0 to 4.9. During interpretation, we should always remember that grade equivalent indicates the average performance of the pupil in each grade level. For every grade level, 50 percent of the pupils in the standardization groups are above and 50 per cents are below. Therefore, we should not interpret this particular grade norm as something all our students should attain.

The percentile rank of a score is the percentage of scores in its frequency

distribution that are equal to or lower than it. For example, a test score that is greater than or equal to 75% of the scores of people taking the test can be said to be at the 75th percentile, where 75 is the percentile rank. Percentile norms are more convenient to use than age norm or grade norm as they are easy to understand. A percentile rank indicates the percentage of individuals who fall below a particular score. The only problem with this norm is percentile units are unequal and therefore, not easily comparable.

Standard scores are used in norm-referenced assessment to compare one student's performance on a test to the performance of other students of her age. Standard scores estimate whether a student's scores are above average, average, or below average compared to peers. Standard scores indicate the number of standard deviation units a raw score falls above or below the mean that can be converted to percentile..

In a distribution, deviations of the scores from its mean expressed in sigma (symbol of sigma to be inserted) are called standard scores. Observed values above the mean have positive standard scores, while values below the mean have negative standard scores. The standard score is a dimensionless quantity obtained by subtracting the population mean from an individual raw score and then dividing the difference by the population standard deviation. This conversion process is called standardizing or normalizing. Standard scores are also called z-values, z-scores, normal scores, and standardized variables. They are most frequently used to compare an observation to a standard normal deviate, though they can be defined without assumptions of normality.

Computing a z-score requires knowing the mean and standard deviation of the complete population to which a data point belongs. The standard score of a raw score x is

Where:

● is the mean of the population.

● is the standard deviation of the population.

The absolute value of z represents the distance between the raw score and the population mean in units of the standard deviation. z is negative when the raw score is below the mean, positive when above.

3.4 Problems of Evaluation through Teacher made Test

Meaning of Teacher Made Test: Carefully constructed teacher-made tests and

standardised tests are similar in many ways. Both are constructed on the basis of a carefully planned table of specifications, both have the same type of test items, and both provide clear directions to the students. Still the two differ. They differ in the quality of test items, the reliability of test measures, the procedures for administering and scoring and the interpretation of scores. No doubt, standardised tests are good and better in quality, more reliable and valid. But a classroom teacher cannot always depend on standardised tests. These may not suit his local needs, may not be readily available, may be costly, and may have different objectives. In order to fulfill the immediate requirements, the teacher has to prepare his own tests which are usually objective in nature.

Teacher-made tests are normally prepared and administered for testing classroom achievement of students, evaluating the method of teaching adopted by the teacher and other curricular programmes of the school. Teacher-made test is one of the most valuable instruments in the hands of the teacher to solve his purpose. It is designed to solve the problem or requirements of the class for which it is prepared. It is prepared to measure the outcomes and content of the local curriculum. It is very much flexible so that it can be adapted to any procedure and material. It does not require any sophisticated technique for preparation.

Taylor has highly recommended the use of these teacher-made objective type tests, which do not require all the four steps of standardised tests nor need the rigorous processes of standardisation. Only the first two steps planning and preparation are sufficient for their construction. Features of Teacher-Made Tests are as follows:

1. The items of the tests are arranged in order of difficulty.

2. These are prepared by the teachers which can be used for both prognosis and diagnosis purposes.

3. The test covers the whole content area and includes a large number of items.

4. The preparation of the items conforms to the blueprint.

5. Test construction is not a single man's business; rather it is a co-operative endeavour.

6. A teacher-made test does not cover all the steps of a standardised test.

7. Teacher-made tests may also be employed as a tool for formative evaluation.

8. Preparation and administration of these tests are economical.

9. The test is developed by the teacher to ascertain the student's achievement and proficiency in a given subject.

10. Teacher-made tests are least used for research purposes.

11. They do not have norms whereas providing norms is quite essential for standardised tests.

Problems or disadvantages through teacher made tests are as follows.

1. **Ambiguity:** Teacher made tests are often ambiguous and unclear.

2. **Length of the test:** Tests are either too short or too lengthy.

3. **Duration:** Tests are usually hurriedly conducted.

4. **Supervision:** Supervision is not proper. There is a lot of scope for copying

5. **Evaluation**: Answer books are not marked carefully.

6. **Coverage of content**: Teacher-Made Tests Designed to Measure Achievement of a Particular Unit of Work, therefore content area is lesser compared to standardized test.

7. **Purpose:** Teacher-made tests are aimed to be used in a classroom or local situation only. Teacher-made test is only valid for that situation as it is specifically prepared for that situation.

8. **Reliability:** In teacher-made tests reliability is unknown, typically having moderate to low reliability. Tests are developed usually by one teacher with little or no outside help without the skill of professional writers, reviewers and editors of test items. Although reliability can be high if items are carefully constructed.

9. **Technical qualities:** Teacher made tests are technically inferior. Teacher-made tests may be carelessly planned and executed that may lead to disastrous consequences. The reliability of teacher made tests is not ensured. Here test items are not expertly written, properly pretested, rigorously analyzed and scientifically retired. Therefore, the procedure in administration and scoring of the teacher-made tests are not dependable.

10. **Administration and scoring:** In teacher-made tests uniform procedure of administration and scoring may be possible and usually flexible and not standardized, leading to unscientific results.

11. **Learning outcomes and content measured:** Teacher-made tests are used to evaluate the outcomes and content of only what has been taught in the classroom. Therefore the scope is narrow.

12. **Interpretation of results:** The teacher-made tests are prepared keeping in view a particular class which is composed of 40 to 50 students maximum. The results of such students are compared with reference to that particular group. As no norm is accompanied this test lacks the ability to judge the position of the individual students in relation to all students of his own age or grade.

13. **Practical utility:** One important principle of good evaluation is that it should be well integrated with instruction and should be continuous. The students are required to be evaluated periodically and if possible after covering each unit. In this connection, teacher-made tests are useful but use items that have rarely been tried out, analysed or revised before becoming part of tests.

14. **School specific:** Based on content and objectives specific to the teacher's own class or school, lacks universal recognition, lacking generalization. These are limited usually to the class or a single school as a reference group, and does not provide norms for various groups that are broadly representative of performance throughout the country.

15. **Others:** Tests are cursory and conducted as rituals only.

3.5 General Principles of Achievement Test Construction

An achievement test has a great significance in all types of instructional progress of the individual. A classroom teacher depends upon the achievement test for measuring the progress of his students in his subject area. Several educational and vocational decisions about students are taken on their performances in the achievement test. It is, therefore, necessary that the teachers should be well acquainted with the meaning and characteristics of achievement tests. Thorndike and Hagen (1969) observed, "The type of ability test that describes what a person has learned to do is called achievement test". Gronlund (1977) defines an achievement test as "a systematic procedure for determining the amount of student has learned through instruction".

The following principles should be taken into consideration for achievement test construction.

- **Mastery of knowledge:** It will be general and comprehensive enough to test thoroughly the pupil's mastery of the desired skills or grasp of a subject matter. It will make it impossible for a pupil to get a good mark just because he got a few common questions and will not fail a good pupil just because of a few uncommon questions.

- **Grading based on ability:** It should reliably grade the pupil into at least six to ten or more different classifications according to their ability.

- **Objectivity and practicability:** It will be as objective as practicable. Some questions which are of discussion nature can be given advantage but there should also be some work which compels the examinee to answer that can be quite objectively scored. Even in the essay type examination, the objectivity can be greatly increased with study and care.

- **Size and duration:** Other things being equal, the test which examines most intensively or extensively in the least time and with the least figures on the part of the pupil is the best.

- **Enhancing creativity:** The test should encourage creative and self reliant work and discourage mere mechanical rote memory work.

- **Increase motivation**: The test should encourage the pupil to put forth his best effort.

- **Self appraisal:** The test should convincingly reveal to the pupil his deficiencies and encourage him to remove them.

- **Objectivity:** The test should examine exactly what it pretends to examine and not handicap pupils for deficiencies in unrelated subjects. The more a test tries to examine in subjects extraneous to the avowed subject, the less efficiently it examines its own.

- **Cultural performance:** The standards of a test should be based on cultural performance and not upon mere opinion of what a standard should be or on an arbitrary percentage work. The more a test makes it possible to compare the performance of a certain group of pupils to that of other pupils the more revealing it is.

- **Importance of habits:** One should keep in mind that no test is infallible. The daily work of the pupil and an observation of his habits of work are equally important in appraising the work and ability of a pupil.

- **Interpretation of score:** Nevertheless one must remember that poor results on the part of a group of pupils indicate a faulty test or one that is not graded to the ability of the pupil or deals with matters the children have not been taught.

- **Norms to be calculated:** Taking into consideration the performance of the group norms for the group should be calculated. Different types of norms are calculated like grade norms, sex norms, local norms, standard score norms and percentile norm.

- **Validity determination:** To determine content validity proper analysis of content and objectives must be made. By the help of item analysis each item should be validated. The validity of the test is indicated by the coefficient of correlation between two sets of scores. Higher the correlation co-efficient higher is the validity.

- **Reliability determination:** To determine reliability we can take the help of methods like test-retest, parallel form of equivalent, split-half and Kuder-Richardson method.

- **Preparation of the manual:** At the final stage a booklet is prepared, containing all details of the scope of the test, instructions in connection with administration, details of different norms, validity, reliability etc.

A good test must be valid, reliable and objective in nature. Thus, test construction remains largely an art rather than a science for which the teacher has to follow a series of steps. These are as follows:

A. Planning the test:

The teacher will have to understand the objectives to be assessed, the examinees to be examined and their test taking behavior. Therefore, careful planning is needed for better test construction. To obtain better result of test construction the following steps should be taken into consideration:

- **Determining the purpose of the test:** While planning the test the teacher has to determine whether the test is meant for placement or formative or diagnostic or summative evaluation.

- **Indicating the conditions which the test has to serve:** It is very much important to pay attention to the conditions under which it will be administered, how much time is needed for the test, the facilities available to repeat the test, age and experience of the individual to be tested.

- **Defining the objective to be measured:** In the entire planning process the most important step is to define the important objectives that are to be appraised. The objectives must be clearly listed, classified in terms of pupil behavior. Each objective should also be given weightage according to its relative importance.

- **Outlining the content area to be covered:** The test should reflect proportionate emphasis on the different content area or units to which the objectives are to be achieved. The weightage should be based on the relative importance and time devoted to each content area.

- **Selecting the appropriate item to be included:** While testing the teacher has to make certain major decisions regarding the type of test items to be used. The decisions will be made on the basis of objectives of the test, availability of time and nature of test.

- **Deciding the difficulty level of the test:** Every test item has a certain difficulty level. Unless it is decided at the right moment it will be very difficult to construct the test. Moreover, in order to discriminate between the high achievers and the lower one the teacher should also set questions such as, very easy, moderately difficult, difficult etc.

- **Preparing the table of specification or Blue print:** Blue-print is a chart which specifies the content deemed important by the teacher and indicates what the student is supposed to be able to do with that content. Thus blueprint may be two dimensional, three dimensional or four dimensional chart containing objectives, content area, forms of question and level of difficulties.

B. Preparing the test:

Like planning the test maker has to perform some important activities in preparing the test and shall follow the following steps:

- **Recording of test items:** If the test items are recorded properly they can be profitably utilized writing the test items simplifies the task of reviewing, selecting and arranging the test items in final test forms. It also facilitates the elimination of defective items. Flexibility is an important part of planning test items.

- **Review of the test items:** The tests are not free from errors even if constructed very carefully. There is a possibility of introducing ambiguous items while

increasing the difficulty of the test items. Two specific ways have been suggested by the experts to get rid of these technical defects. These are: (a) reviewing the items after they have been set aside for a few days. (b) Asking a fellow teacher to review and criticize the items.

- **Arrangement of the test item in the test:** Tests are mostly used in classrooms. Therefore, satisfactory arrangement of test items is very essential. The arrangement of the test items should be made item wise. To facilitate the pupils to retain the idea easily and it also facilitates the test maker for the sake of scoring. Test items can proceed from simple to complex, which may include true-false, then matching items, then short answer, multiple choice, interpretive exercise and lastly essay questions.

- **Preparation of directions for the test:** The vital point is preparation of direction for the test. Whenever the test items are prepared they should be accompanied with necessary directions to facilitate the learners to answer effectively. Therefore, written directions in the test items must be given.

- **Reproduction of the test:** When the test materials are ready for production it is important that the test items should be spaced and arranged in such a manner that they can be easily read, answered and scored. It is not possible to cram a large number of test items. Therefore, the number of test items should be printed or written by hand or memographed or typed. Introductory materials should be placed on a facing page or separate sheet with all of the items referring to it on a single page.

- **Test scoring procedures:** The test maker should create a proper environment for this purpose. The psychological, physical and time factors influence test results. Scoring is nothing but a matter of comparing the columns of answers on each paper used by the student. When we need to record the answer of the people on the test paper itself we need a scoring key by making the correct answer on a blank copy of the tests.

C. Trying-out the test:

The third phase in test construction is trying out the test. The trying out is of two types. These are as follows:

a. **Preliminary try-out:** The purpose of preliminary try out is to ascertain the effectiveness of the test items. In this the following points should be specially paid attention to:

- **Selection of sample:** The sample for administration can be randomly or purposefully selected.

- **Administration of the test:** All pupils must be given a fair chance to demonstrate their achievement of the learning outcomes to be measured in a favourable atmosphere.

- **Scoring:** It should be done in accordance with the key prepared for the test, using the simplest method.

- **Item analysis:** The process followed for ascertaining the effectiveness test item is known as test analysis. Generally the effectiveness suitability of items is ascertained on the basis of the difficulty level and discriminating power of the items. The formula used for this purpose are

  1. $Di = \dfrac{U + L}{2N}$ 2. $Dp = \dfrac{U - L}{N}$

  [Where, Di= Difficulty Index, Dp= Discriminating Power, U=number of correct responses in upper group, L= the number of correct responses in lower group, N= Total number of cases in each group]

- **Selection of items:** On the basis of the difficulty index and discriminating power considered at the same time items will be rated. Generally the difficulty index of a good item is considered to lie between 0.4 and 0.6.

- **Preparation of the final test:** After finalizing the items to be included in accordance with the blue print, these may be arranged in the order of difficulty and the instruction and other details added to get the final draft of the test.

b. **Try-out proper: This is an important test which includes the following steps:**

- **Selection of sample:** A convenient sample that represents the population is selected for the purpose. The size of the sample depends upon the size of the population

- **Administrations:** The test should be administered in accordance with the principle of test administration.

- **Scoring:** The scoring of the script strictly in accordance with the scheme of valuation prepared in advance by the paper setter or the team who finalized the test. This can be done with the help of the scoring key.

D. Evaluating the test:

It is the final step in construction. The test can be discussed in two different ways:

    A. Analysis from a diagnostic point of view which helps the teacher for re-teaching and remedial teaching.

    B. Inferences can be drawn about the effectiveness of the tests and also about its function.

3.6 Summary

Tests are systematic procedures for observing persons and describing them with either a numerical scale or a category system. Thus tests may give either qualitative or quantitative information. Further by its denotation, the term test refers to any tool or mechanism through which assessment can be made. A good measurement instrument possesses five major characteristics, such as, objectivity, validity, reliability, usability and norms. A test is said to be an objective if the same scores or results are obtained by administering the test to a particular group of students on different occasions and scored either by the same scorer or different scorer. Tests should be made free from errors. A test is said to be valid when it measures that it intends to measure. There are three types of validity; content validity, construct validity, criterion related validity. Criterion validity may be classified into concurrent and predictive validity. A test also may suffer from other forms of variable error or chance error. Reliability refers to the consistency of a measure. A test is considered reliable if we get the same result repeatedly. That is, if a test is applied to an individual or to a group of students on two or more occasions and the results obtained in all occasions do not differ at all or differ to a very little degree, we say that the test is a reliable one or the test is free from variable errors. A test has utility if it provides the test condition that would facilitate realization of the purpose for which it is meant. Finally norm is a device of transforming new scores into standard scores in a group. The norms of any educational test represent the average test performance of the standardized group or sample selected from a specific population. In this an individual score is compared with the standardized sample as a reference group.

Teacher made tests are frequently the basis of evaluating the progress of students in the class. These tests are of great value to the teachers, the students and their parents and administrators and educators. Although teachers made tests have some limitations. These are often ambiguous and unclear; either too short or too long, does

not cover the entire content, cursory, improper supervision, careless marking etc. A classroom teacher depends mostly on achievement tests for measuring the progress of students. An achievement test possesses some general principles and steps of Planning, preparing, trying out and evaluating the test.

3.7 Self Assessment Questions

1.   What are the characteristics of a good test?

2.   Name the forms of reliability.

3.   What are the factors of reliability?

4.   Define validity.

5.   What are the properties of validity?

6.   Discuss the different types of norms.

7.   What are the different aspects of usability of a test?

8.   Identify the problems of teacher-made tests in educational evaluation.

9.   State the major principles of achievement test construction.

10. Discuss the steps of achievement test construction.

3.8 References Books

•   Aggarwal, J.C. (2001): Basic Ideas in Educational Psychology, Shipra Publications, Delhi, page no.189-198

•   Aggarwal, J.C. (2001): Essentials of Examination System: Evaluation, Test and Measurement, Vikash Publications, Delhi, page no.268-272, 273-279

•   Biggs, J.B. & Collis, K.F.(1982): Evaluating the quality of learning: the SOLO taxonomy (structure of the observed learning outcome), New York: Academic Press.

•   Mangal, S.K. (2010): Statistics in Psychology and Education, New Delhi, PHI Learning Private limited, ISBN-978-81-203-2088-8, page no 71

•   Dash, B.N. , Dash, Nibedita and Rath, M.P.(2003): A Textbook of Educational Psychology, Dominant Publishers, Delhi, India, page no.416-525

•   Lovely Professional University (2012): Educational Management and Evaluation

DEDU504, USI Publications, New Delhi, page no.37-45, 63-89, 290-311

- Oriondo, Leonora Loyola (1984): Evaluating educational outcomes (Test, Measurement and Evaluation), Manila, Phillipines, Eleanor M. Dallo-Antonio Rex book store, ISBN 971-23-0330-6, pages 164 to 167, 177

- Siddiqui, Majibul Hasan (2009): Educational Evaluation, A.B.H. Publishing Corporation, New Delhi, page no.196-201

- Singh, A.K. (2002): Tests, Measurements and Research Methods in Behavioural Science, Patna, Bihar, India, Bharati Bhawan Publishers and Distributors, page no 17-18, 55-87

- Singha, H.S. (1999): Modern Educational Testing, New Delhi, India, Sterling Publication Private limited, ISBN812-07-217-13, page no 82- 94

- Thammishetty, Manichander (2016): Educational Evaluation, Laxmi Book Publication, Solapur, Maharashtra, india,ISBN132-93-1229-5, page no. 53, 54

Webliography:

- A Primer on Psychology a new approach to an ancient question. https://funpsychology.wordpress.com/psychological-testing/norms/retrieved on 02.11.2017

- https://en.wikipedia.org/wiki/Percentile_rank retrieved on 31.10.17

- https://en.wikipedia.org/wiki/Reliability_(statistics) retrieved on 31.10.17

- https://en.wikipedia.org/wiki/Standard_score retrieved on 31.10.17

- https://www.sagepub.com/sites/default/files/upm-binaries/11695_G06.pdf retrieved on 10.06.2020

Module - II

Testing & Evaluation

Unit 4 ❑ Types of Evaluation

Structure:

4.1 Objective

After going through the unit the students will be able to

● Define educational evaluation

● discuss the characteristics and significance of formative and summative evaluation

● differentiate between formative and summative evaluation

● state the characteristics and significance of Continuous Comprehensive Evaluation

● point out the concept and significance of peer participatory evaluation in education

4.2 Introduction

Evaluation refers to the assessment of students' progress towards set objectives. It is a systematic determination of a subject's merit, worth and significance, using criteria that are governed by a set of standards. It can assist an organisation, programme, project or any other intervention or initiative to assess any aim, realisable concept or

proposal, or any alternative, to help in decision making or to ascertain the degree of achievement or value with respect to the aims and objectives and results of any such action that has been completed. The primary purpose of evaluation is to enable reflection and assist in the identification of future change. Evaluation is often used to characterise and apprise subjects of interest in a wide range of human enterprises. It is long term and done at the end of a period of time. Contemporary education recognises evaluation as an integral part of the instructional process. Modern educational process has given equal importance to planning, directing and evaluating instruction. Even the most skilful teacher has to evaluate the actual results of his instructional procedure in terms of pupil achievement to carry his task to its ultimate conclusion. Therefore, the evaluation process, its types, methods will be discussed in this unit. In today's policy environment, testing has become a critical component of education reform. Policy makers and education administrators often view test scores as a measure of educational quality and use test scores to hold schools accountable for teacher performance. Continuous and comprehensive assessment offers a methodology for measuring pupil performance and using those findings to improve the success of pupils. CCE is a classroom strategy implemented by teachers to ascertain the knowledge, understanding, and skills attained by pupils. This process is much more than an examination of pupil achievement. It is also a powerful diagnostic tool that enables pupils to understand the areas in which they are having difficulty and to concentrate their efforts in those areas and also allows teachers to monitor the impact of their lessons on pupil understanding. The concept, characteristics and significance will be discussed in this unit. Lastly, Peer or participatory evaluation in education will be discussed in this unit. Participatory evaluation, as we shall see, is not simply a matter of asking stakeholders to take part. In education peer or participatory evaluation is very important today. The concept and significance will be discussed in detail.

4.3 Formative and Summative Evaluation: Characteristics and Significance

Evaluation is a continuous process. It is concerned with more than the academic achievement of students. It tends to develop the individual in terms of desirable behavioural changes in relation to feeling, thinking and action. Effective learning must bring about some behavioural changes in students. These desired changes are set during planning of instructional objectives. Educators were interested previously in measuring the achievement of students based on students' scores. Today, educational

measurement is considered as an important aspect of evaluation. Educational evaluation is the evaluation process of characterizing and appraising some aspects of an educational process.

Evaluation is inherently a theoretically informed approach (whether explicitly or not), and consequently any particular definition of evaluation would have been tailored to its context - the theory, needs, purpose, and methodology of the evaluation process itself. Keeping in mind this, evaluation has been defined as: A systematic, rigorous, and meticulous application of scientific methods to assess the design, implementation, improvement, or outcomes of a programme. It is a resource-intensive process, frequently requiring resources, such as, evaluates expertise, labour, time, and a sizable budget.

According to St Leger and Wordsworth-Bell "The critical assessment, in as objective a manner as possible, of the degree to which a service or its component parts fulfils stated goals". The focus of this definition is on attaining objective knowledge, and scientifically or quantitatively measuring predetermined and external concepts.

According to Stufflebeam and others (1971) evaluation is providing information for decision making. Stuffelebeam defines evaluation is "A study designed to assist some audience to assess an object's merit and worth". In this definition the focus is on facts as well as value laden judgements of the programs outcomes and worth.

Today we do not only test but also evaluate for appraisal of the effects of the learning experiences. Evaluation designates a process of appraisal which involves the acceptance of specific values and the use of a variety of instruments of observation, including measurement, as the basis of value judgement. From the functional point of view it involves identification and formulation of a range of major objectives of a curriculum, their definition in terms of students' behaviour and construction of valid, reliable and practical instruments. These instruments or tools are used for observing the specific phases of students' behaviour such as knowledge, skill, attitude, appreciation, personal and social adaptability, interest and work habits. While the child is acquiring knowledge and skill, learning in attitude, appreciation and interest are also taking place. This approach indicates a shift from a narrow conception of subject matter outcomes to a broader, conception of growth and development of individuals. Evaluation refers to the assessment of students' progress toward stated objectives, efficiency of our teaching as well as effectiveness of the curriculum. Therefore, evaluation has three dimensions, which are represented diagrammatically:
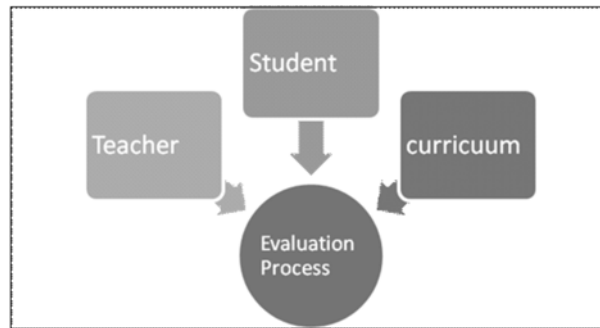
Fig1: Dimensions of Evaluation
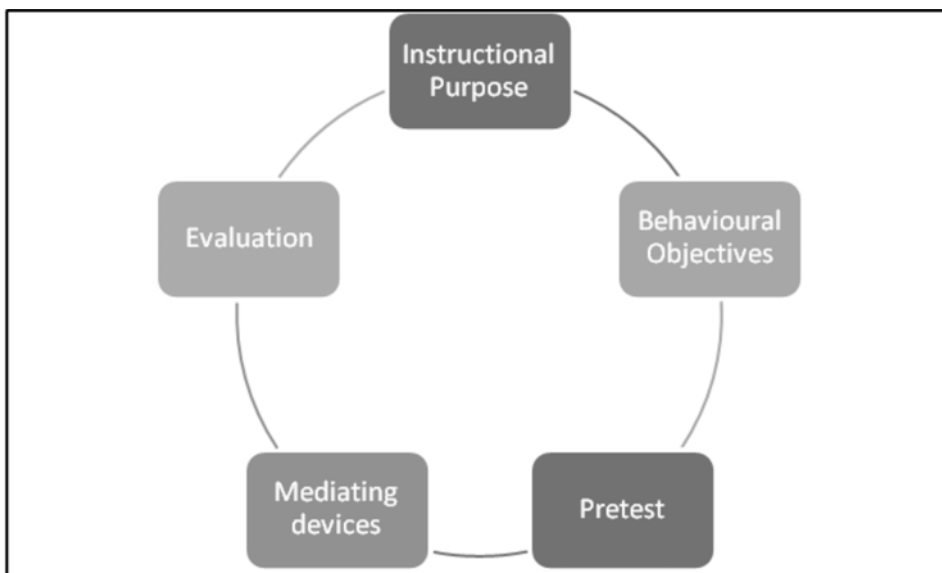
Fig 2: Cyclic process of evaluation

The nature of evaluation as a continuous process is illustrated in the above diagram. Evaluation plays an important role in the field of education. We evaluate student's skills and abilities in order to select students for a special programme and to diagnose a student's weakness and strength. We also take the help of evaluation to determine the effectiveness of instruction, usefulness of a special programme or a new curricular, and mastery of certain learning objectives. 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. Honer,

M. Seriven (1967) coined the terms of summative and formative evaluation. Evaluation also helps the teachers to report student progress to parents. The report is also used for guidance by counselors in career counseling. Evaluation helps the researcher to investigate the effectiveness of teaching methods used by classroom teachers. Evaluation can also be diagnostic in nature. Depending on the purpose of evaluation a classroom teacher usually uses four types of evaluation; placement, formative, summative and diagnostic. The two major evaluation types are discussed here.

4.3.1 Formative Evaluation

Formative evaluations evaluate a programme during development in order to make early improvements that help to refine or improve programme. Different scholars have defined formative evaluation. Such as, According to A.J.Nitko (1983) "formative evaluation is concerned with judgement made during the design and or development of a programme which are directed towards modifying, forming or otherwise improving the programme before it is completed."

Gronlund (1985) defined formative evaluation as "formative evaluation is used to monitor learning progress during instruction and to provide continuous feedback to both pupil and teacher concerning learning successes and failures. Feedback to pupils reinforces successful learning and identifies the learning errors that need correction. Feedback to the teacher provides information for modifying instruction and prescribing group and individual remedial work".

R.L. Ebel and D.A. Frisbie (1986) opined that "formative evaluation is conducted to monitor the instructional process, to determine whether learning is taking place as planned."

Gilbert Sax (1989) stated that "formative evaluation takes place during instruction by letting the teacher or evaluator know if students are meeting instructional objectives, if the programme is on time and if there are ways that the programme might be improved. Formative evaluation helps current students to learn more effectively"

W. Wiersma and S.G.Jurs have written that "formative evaluation occurs over a period of time and monitors students' progress."

Characteristics: After going through definitions by different scholars some characteristic features of formative evaluation as follows:

1. Formative evaluation is done during an instructional programme.

2. The instructional programme should aim at attainment of certain objectives also during the implementation of the programme.

3. Formative evaluation is cause seeking.

4. It relatively focuses on molecular analysis.

5. It is interested in the broader experiences of the programme users.

6. Its design is exploratory and flexible in nature.

7. It often ignores the local effects of a particular programme.

8. It seeks to identify influential variables.

9. Formative evaluation is used to monitor learning and modifying the programme if needed before its completion.

10. It requires detailed analysis of instructional material for mapping the hierarchical structure of the learning tasks and actual teaching of the course for a certain period of time.

Significance: Following are the significance of formative evaluation for the classroom teacher.

- Formative Evaluations provide the information on improving a product or a process.

- Formative evaluations serve a monitoring function rather than focusing solely on measurable programme outcomes or evaluation findings

- Formative evaluation is used for current students.

- This type of evaluation provides continuous information that can be used to modify the programme to improve its effectiveness and efficiency. In a classroom situation, the purpose of formative evaluation is to provide feedback to the teacher and students about progress in unit tests, measures of interest and attitude.

- Interviews or conferences with students and parents during the programme can provide important indications for improvement and betterment of the programme by redirecting it.

- These are mostly teacher made tests. At the end of the teaching of a chapter the teacher takes a test or quiz. Then the teacher reviews the test result to determine which material requires further discussion.

4.3.2 Summative Evaluation

Summative evaluation provides information on programme effectiveness normally conducted after the completion of the programme design. Following definitions are provided by different scholars.

According to A.J.Nitko (1983) "summative evaluation describes judgments about the merits of an already completed programme."

Gronlund (1985) stated that summative evaluation as "summative evaluation typically comes at the end of a course or unit of instruction. It is designed to determine the extent to which the instructional objectives have been achieved and is used primarily for assigning course grades or certifying pupil mastery of the intended learning outcomes."

R.L. Ebel and D.A. Frisbie (1986) defined that "summative evaluation is conducted at the end of an instructional segment to determine if learning is sufficiently complete to warrant moving the learner to the next segment of instruction."

Gilbert Sax (1989) stated that "a summative evaluation can provide evidence that the programme is satisfactory and should be continued for next year students or that student learning and learning attitudes are so negative that a new programme is needed."

W. Wiersma and S.G.Gurs have opined that "summative evaluation is done at the conclusion of instruction and measures the extent to which students have attained desired outcomes".

Characteristics: After going through definitions by different scholars the following characteristics of the summative evaluation are identified:

- Summative evaluation takes place at the end or completion of a particular programme whose duration may vary from semester to whole year.

- Summative evaluations should check whether there has been learning or not. If the learning is found to take place the quantity and quality of that learning in relation to predetermined objectives should be kept in mind.

- Summative evaluations provide feedback to the classroom teacher for the success or failure of the programme or of the student.

- There must be some instructional programme prior to summative evaluation. The instructional programme should be set for attainment of some objectives.

- Summative evaluations use well defined evaluation designs.

- It also focuses on analysis.

- It provides descriptive analysis.

- It takes into account the local factors.

- It is non-reactive and unobtrusive as far as possible.

- It is concerned with a broad range of issues.

- Reliable and valid instruments are used here.

Significance :

- Summative Evaluations provide information of short-term effectiveness or long-term impact to deciding the adoption of a product or process.

- Summative evaluation usually occurs after instruction. It is designed to determine how well the instructional objectives are met. Formal classroom tests, such as unit tests or final examinations are the most frequently used tools used in this evaluation.

- The term summative means summing up all the available information regarding a programme at its terminal point. Such information can be a valuable way of assessing the effectiveness of the whole programme.

- It also provides correction if the programme continues. The techniques used here are teacher made, evaluation product such as achievement test, research, report, themes, drawing etc. it also depends on rating of different types of performances such as oral report, laboratory experiments etc.

- It is scientific as reliable and valid instruments are used here.

- Summative evaluations provide feedback to the classroom teacher for the success or failure of the programme or of the student. This helps the teacher to evaluate his or her standard of teaching. Based on this evaluation teacher can change plans, teaching methods, tools and techniques subsequently.

4.3.3 Difference between Formative and Summative evaluations

Initially the terms were applied for evaluation of curricular work only. M. Seriven in his book Evaluation Thesaurus (1980) differentiated between these terms. Formative evaluations are conducted during development or improvement of a programme or

product. It is conducted for internal or in-house staff and normally done by an internal or combination of internal and external evaluators. Whereas, summative evaluation is conducted after completion of a programme or a course of study. It is done for the benefit of some external audience or decision maker by an internal or external evaluator or a combination of both. Following are the main differences between these two types of tests:

● The tests differ in purpose, nature and timing.

● Summative evaluation is the terminal assessment of performance at the end of the instruction, whereas formative evaluation is the assessment made during the instructional phase to inform the teacher about progress in learning and taking necessary steps.

● The use of record and profile of achievement are frequently used in formative evaluation, lesser used in summative evaluations.

● The formative evaluation is termed as process and the summative evaluation is termed as product evaluation. Summative evaluations determine the extent to which the examinee has mastered the knowledge and skill associated with the course, but Formative evaluations deals with the process by which the examinee achieved these outcomes.

● In formative evaluation both students and teachers are being evaluated. Whereas, in summative evaluations, assessment is done to test the learning outcomes of the students against a set of pre-determined criteria without revealing the route to the teacher which the students followed in reaching the point.

● Formative evaluations use a variety of instruments which are either locally developed or standardized. It relies on observation and informal data collection devices. In contrast to this, summative evaluations tend to use well defined evaluation devices. These are non-reactive, comparative and concerned with a broad range of issues. These are publicly accepted, reliable and valid instruments.

● In formative evaluation scoring is based on criterion referenced approach, but in summative evaluation scoring is normally based on norm referenced approach. Criterion referenced test can also be applied in summative evaluation.

● The method used in formative evaluation of reporting score is the individual

pattern of pass-fail scores in each task in hierarchy, whereas in summative evaluation attainment is reported in terms of total score.

● In formative evaluation immediate and continuous feedback is provided to the students. Thus it forces and reinforces learning mastery by providing data that can direct remedial teaching. Summative evaluations are in real sense 'final' tests of students' achievement typically covering large blocks of instructional materials.

Table-1: Broad differences between Formative and Summative evaluations

criteria	Formative	Summative
Purpose	To monitor progress of students by collecting feedback	To check final status of students
Content focus	Detailed, narrow scope	General, broader scope
Methods used	Daily assignments, teachers observation	Tests, projects
Tools used	Teacher made test	Standardized test
Frequency	daily	Weekly, quarterly, monthly, half yearly, annual term examination

4.4 Continuous and Comprehensive Evaluation (CCE): Characteristics and Significance

Evaluation experts have strongly felt that the traditional procedures and practices in the examination are more harmful to our students. It has given rise to a number of malpractices including frequent use of unfair means, misuse of result, huge mental pressure etc. At present there is an imperative need to make evaluation an integral part of the total teaching learning process. This means that evaluation should be a continuous process and there is also a need to make it comprehensive. Hence the Continuous Comprehensive Evaluation has evolved.

Continuous comprehensive evaluation or CCE is a classroom strategy implemented by teachers to ascertain the knowledge, understanding, and skills attained by pupils throughout the year. Teachers administer assessments in a variety of ways over time

to allow them to observe multiple tasks and to collect information about what pupils know, understand, and can do. These assessments are curriculum-based tasks previously taught in class. CCE occurs frequently during the school year and is part of regular teacher-pupil interactions. Pupils receive feedback from teachers based on their performance that allows them to focus on topics they have not yet mastered. Teachers learn which students need review and remediation and which pupils are ready to move on to more complex work. Thus, the results of the assessments help to ensure that all pupils make learning progress throughout the school cycle thereby increasing their academic achievement.

The addition of continuous and comprehensive evaluation in the instructional and testing process is intended to achieve two major purposes: to improve both the validity and reliability of the results of pupils' performance on tests and exercises, and secondly to help the pupil to develop effective learning and work habits. The present CCE system is essentially based on frequent test taking and does not really serve the two critical purposes of continuous comprehensive evaluation. Classroom tests are based on assessment of lower level abilities and rote memorization, here assessments are based on low level thinking skills i.e., "Knowledge" and "Comprehension". As a result pupils complete their education but still they are unable to analyze and apply their knowledge to solve problems. This education is considered ineffective as it is unable to transform the pupil from the stage of "knowledge recipient" to the status of "knowledge producer and problem solver". The central purpose of CCE is to help the pupil to become a better learner and producer by encouraging pupils to improve their knowledge and skills through learning, test taking and project undertaking in the critical and important objectives of the school curriculum.

Characteristics of Continuous and Comprehensive Evaluation: The features of continuous comprehensive evaluation are:

1. **Integral part of teaching learning:** Continuous Comprehensive evaluation is an integral part of the teaching learning process. More valid and reliable evidence about scholastic and non scholastic growth of learners can be collected. CCE can help to maintain desired performance standards through quality control measures at regular intervals.

2. **It is a feedback device:** CCE is also used as a feedback device to improve teaching and learning. On the basis of CCE more sound judgments and timely decisions about teaching and learning are provided.

3. **Longer time for collecting assessment information:** To obtain accurate and reliable assessment data on a pupil, the assessments are spread over a longer time, allowing the pupil to take tests and other assignments at different times throughout the course. The average of the scores for the various assessments is a more reliable indicator of the pupil's performance in the subject than the score the pupil obtains in a one-shot examination.

4. **Use of different testing tools and techniques:** By extending the time span for collecting assessment information throughout the duration of a course, different forms of testing and different assessment situations including acquisition and demonstration of practical skills can be introduced in this process. Practical skills such as the skills and competencies involved in conducting interviews, writing and presenting reports, presenting and analyzing data in graphical forms, and production of three-dimensional objects in a variety of subjects could be encouraged in schools to provide a more comprehensive and more valid assessment of pupil's ability.

5. **Inclusion of high complex thinking skills in the testing programme:** By extending the period for collecting assessment data, forms of knowledge and competencies that cannot be easily assessed under strictly timed conditions can also be assessed. High level thinking skills involving analytical thinking and problem solving skills and other competencies that require extended time for learning and for test response can then be added to the continuous assessment programme. The addition has the effect of helping pupils to acquire the habit of using high level thinking skills in a variety of situations rather than using pure memorization and other low level thinking processes.

6. **Teacher assistance and guidance:** Another feature of the continuous and comprehensive evaluation process is to foster cooperation between the pupil and teacher especially in the area of pupils' class projects. The process requires the teacher to provide assistance in the form of advice on various aspects of pupils' projects. The pupil learns to consult the teacher, classmates and other sources on aspects of his/her project work, while maintaining his/her position as the leader in the project undertaking. This is the normal work procedure in the adult world where production is essentially based on cooperation and not on timed test situations.

7. **Appreciative Inquiry:** CCE uses the principles of appreciative inquiry as a

basis for the development of the activities in each unit. Appreciative inquiry is a method of figuring out how an organization or group can best undergo positive change. With appreciative inquiry, the starting point of change is acknowledging the strengths and skills one already has. The process of change begins by identifying assets (strengths) rather than problems. CCE activities ask teachers and other educators to identify their strength. The next stage of appreciative inquiry which is also incorporated into many of the activities is the dreaming or aspiration stage. In this part of appreciative inquiry, those involved in the change are asked to think about how they might see their classroom, their teaching, or school in the future. The third stage of appreciative inquiry is designing the future. This is where teachers describe what they will do in order to make the changes they want to bring out. In this way CCE can provide a road map or action plan for accomplishing the goals they have described.

8. **Adult Learning:** As each unit focuses on activities, many of the activities include working together with colleagues. A series of discussion questions and actions that will help teachers to analyze their own teaching context in light of the information provided are included. Teachers, head of the institution, district supervisors, and parents have important perspectives about their school's particular situation. These adults offer a rich source of ideas and talents that may be used to help their schools for improvement.

9. **Experiential Learning:** Adults learn through experience and reflection on that experience. CCE asks teachers to try new techniques in their classrooms (experience) and then discuss what happened (reflection) with colleagues. The many activities with experiences will lead them to new ways of teaching and thinking about teaching. Teachers are encouraged to carry out the activities and answer the questions for each activity. In this way teachers will gain a deeper understanding of the teaching and learning strategies and ideas.

10. **Understanding Reasons for Change:** It is important that teachers have opportunities to analyze reasons for change. Most teachers resist being told what to do. For teachers, theory and reasons for implementing new strategies and concepts are an important part of making a decision to try new techniques. Teachers need to understand why a particular innovation is beneficial to the learners.

11. **Responsibilities for applying new learning:** Ultimately, teachers themselves decide whether or not to apply their new learning to the classroom. They may need to adapt their new skills and knowledge to fit their particular situation. Teachers are the agents of their own change and the students as well through CCE.

Types of Continuous and Comprehensive Evaluation:

Continuous and Comprehensive Evaluation is more likely to be formative, process-oriented, informal, internal, learner involved, and/or self- referenced in nature. It can take the form of daily work (e.g. essays, quizzes, presentation and participation in class), projects/term papers and practical work (e.g. laboratory work, fieldwork, clinical procedures, drawing practice).

- **Daily work:** Experience indicates that students learn faster and have less confusion when provided with more feedback. This is why daily work seems to be the most sustained and extensive means of testing, and in providing relatively prompt feedback, it serves to reinforce or correct learned responses. It assists in pacing learning. If a course is broken into units, each assessed on completion, students have regular feedback of what they have mastered (and what they need additional work on). As the most extensive means of assessment, it has a sustained impact on and improves the quality of student learning. As it is fairly labour intensive for both students and the teacher, decisions will have to be taken regarding an optimum level of task. Too many assignments may become overburdened for all of them. It is probably better to have regular but smaller assignments than large scale but infrequent ones. Vigilance is needed to ensure the prohibition of plagiarism and other forms of cheating. Students should be aware of the punishment against such offences. As a counter check, some of the assignments are to be arranged to be done in the class.

- **Projects or term papers:** This provides freedom to the students as well as a measure of the students' ability to identify, define, collect, select and use data or information, undertake independent study and enquiry, plan and follow through a fairly large-scaled piece of work, integrate theory and practice, work with others, when the exercise involves teamwork. It is difficult to ensure parity in the projects. Careful coordination in the setting of topics is important. Project work may lead to over-involvement at the

expense of other aspects of the course. Teachers or supervisors should therefore guide students in defining tasks and in being selective with regards to content. It may be difficult to maintain consistency and fairness in marking across a range of projects/papers. To mitigate this, project objectives should be clearly identified and some broad criteria for assessment are set. Wherever possible, two examiners should be used. Students may receive unequal supervision and this may affect the quality of the completed project. Agreement should be made among teachers or supervisors to standardise the degree of direction and guidance to be given to students. Criteria for evaluation should be available to students in advance as these provide direction. It may be difficult to assign individual marks to students working in a group project, but this may be resolved by orally examining each student through viva voce to establish their knowledge of the subject and the degree of his/her contribution. Alternatively, students should have work on projects individually.

- **Practical work:** It is important as it provides the opportunity for applying the learning by students. It provides a measure of ability to relate theory to practice, students' techniques, procedures and practical skills. Too much emphasis may be given to the written report without sufficient consideration to the actual process and performance. To avoid this, instructors should scrutinise practical work closely and question students about their organisation of the work and their findings. The theoretical and practical aspects of the course may not be sufficiently linked. Obviously, it is important to communicate frequently with those involved in managing different parts of the course.

Tools and Techniques of CCE: CCE involves the following tools and techniques for evaluating scholastic and non scholastic aspects of students' development.

A. Evaluation of scholastic aspects:

Both testing and non testing techniques can be used for evaluation of scholastic aspects of students' development. In the case of testing, following types of tools are usable:

I. Readiness test

II. Diagnostic test

III. Term test

IV. Annual test

In the case of non-test devices the following are usable for judging students' scholastic attainment.

I.   Performance on assignments

II. Performance on orals

III. Performance on practical work

IV. Students' products

B. **Evaluation of non-scholastic aspects: The following non-test devices are used for students' non scholastic aspects of development.**

  I.   Observation schedule for practical work, participation in games etc.

  II.   Rating scales for social qualities, interest, attitude etc.

  III. Checklist for health habits, work habits, study habits etc.

  IV.   Questionnaire for opinion of experts, teachers etc.

  V.   Interview schedule to conduct an interview.

  VI.   Opinionnaire to seek face to face opinion and views.

  VII. Sociogram to identify social relationships among peers.

  VIII. School records to record evidence about students' attainments, progress cards, anecdotal records, etc.

In addition to this, standardized tests like personality tests, interest inventories, attitude scales can also be used for assessing students' development in non-scholastic areas.

Significance of Continuous Comprehensive Evaluation:

- The CCE process is much more than an examination of pupil achievement. It is also a powerful diagnostic tool that enables pupils to understand the areas in which they are having difficulty and to concentrate their efforts in those areas.

- Continuous and comprehensive evaluation also allows teachers to monitor the impact of their lessons on pupil understanding. Teachers can modify their

pedagogical strategies to include the construction of remediation activities for pupils who are not working at the expected grade level and the creation of enrichment activities for pupils who are working at or above the expected grade level. Hence, the continuous and comprehensive evaluation process supports a cycle of self-evaluation and pupil-specific activities by both pupils and teachers.

● Frequent interactions between pupils and teachers imply that teachers know the strengths and weaknesses of their learners. These exchanges foster a pupil-teacher relationship based on individual interactions. Pupils learn that the teacher values their achievements and that their assessment outcomes have an impact on the instruction that they receive. Teachers can provide individual attention to the students.

● In CCE as teachers assess the curriculum as implemented in the classroom, it also allows teachers to evaluate the effectiveness of their teaching strategies relative to the curriculum, and to change those strategies as dictated by the needs of their pupils.

● In addition, continuous and comprehensive evaluation provides information on achievement of particular levels of skills, understanding, and knowledge rather than achievement of certain marks or scores. Thus, CCE enables pupils to monitor their achievement of grade level goals and to visualize their progress towards those goals before it is too late to achieve them.

Advantages of Continuous and Comprehensive Evaluation:

● One of the expected advantages of continuous and comprehensive evaluation lies in its being guidance oriented. Since it will involve data gathering over a long period of time, it will yield more accurate data reaching the teachers early enough to modify instruction. This could play a vital role in diagnosing and remediating areas of learners' weaknesses if properly anchored in what occurs in classrooms.

● CCE is an approach that would capture the full range of learners' performance. Teachers and administrators would thus be able to assess learners' progress and would have time to correct the problems.

● Another advantage of continuous and comprehensive evaluation is that it places teachers at the centre of all performance-assessment activities. It

encourages more teacher participation in the overall assessment or grading of his/her learners. Teachers must be given opportunities to select and review assessments so that they become involved and knowledgeable in the process. Through this approach, teachers would be able to integrate assessment and assessment results into instructional practice. Teachers will be expected to incorporate assessment into the larger learning framework and possibly to provide evidence regarding how assessment information is used to inform and guide instruction for individual learners. According to Lewis (1997), with continuous assessment teachers must embed the assessment in their instructions, score the assessments and discuss standards for good learners' work with colleagues, parents and learners.

● One of the important aspects of continuous comprehensive evaluation is the availability of valid and reliable tests which could be used in all schools. There is a need to construct these tests following established procedures and practices. To make the results comparable across all the schools, teachers need to be equipped with skills of test construction and administration. This could be done through teacher training institutions so that teachers are equipped with such skills as part of their training and certification.

● Both scholastic and non-scholastic abilities are assessed through CCE. Apart from the skills of test construction measuring cognitive aspects of learning, teachers should also be able to measure the learners' affective attributes such as attitudes, motives, interests, values and other personality characteristics. Such characteristics could be as important as others associated with intelligence. They could assist the teachers and administrators in understanding the learners better, both in the process of education and in the practical affairs of everyday life. They could help us answer questions such as why learners do not do well in schools in spite of being perceived to have high academic abilities. They also provide clues about the interest patterns of learners which could be used in their placement into schools of higher learning and for employment purposes. It is believed that anti-social behaviours such as truancy, lying, cheating, stealing and poor attitude to work could be corrected by providing effective education in schools.

Limitations of Continuous and Comprehensive Evaluation:

● We should keep in mind for successful implementation of the CCE; teachers

need to give most tests, which means more reliance on marking. They need to observe the learners more keenly to assess their affective outcomes, and there will be more records to be kept on the learners. All these could mean more work to the teacher, more demand on his or her time and more responsibility on him or her. This means they must be professionally and attitudinally prepared for operating the system. If the teacher is not adequately prepared for operating the system, it may lead to a tendency to merely 'cook up' scores in the name of continuous assessment. Thus, teachers should be encouraged adequately trained to form favourable attitudes toward the practice. They should be made aware of the requirements of the system, its importance and how to implement it.

● Another problem with CCE is the issue of record keeping. Learners' records have to be adequately and meticulously kept over a long period of time. They should be properly stored and easily retrievable. A related issue is that of collation. Scores may have to be combined from different sources using various weights. Teachers will need basic arithmetical operations of addition and multiplication.

4.5 Peer (Participatory) Evaluation in Education: Concept and Significance

When most people think of evaluation, they think of something that happens at the end of a project that looks at the project after it is over and decides whether it was effective or not. But evaluation actually needs to be an integral part of any project from the beginning. Participatory evaluation involves all the stakeholders in a project that are directly affected by it or by carrying it out in contributing to the understanding of it, and in applying that understanding to the improvement of the work. Involving everyone affected may change the whole nature of a project from something done for a group of people or a community to a partnership between the beneficiaries and the project implementers. Beneficiaries participate making sure that their real needs and those of the community are recognized and addressed. The essence of participatory evaluation is that professional evaluators, project staff, project beneficiaries or participants, and other community members all become responsible and active to improve the community's quality of life.

Now this participatory principle can also be applied in educational evaluation.

Peer/ participatory or peer review in education provides a structured learning process for students to criticise and provide feedback to each other on their work. It helps students develop lifelong skills in assessing and providing feedback to others, and also equips them with skills to assess and improve their own work.?

Concept of peer evaluation in education: It is often observed that group work can be more successful when students are involved in developing the assessment process. This may include establishing their own assessment criteria through consultation with teaching staff. Alternatively students can be provided peer evaluation criteria. The main aims of peer evaluation are to:

- increase student responsibility and autonomy

- strive for a more advanced and deeper understanding of the subject matter, skills and processes

- lift the role and status of the student from passive learner to active learner and assessor that also encourages a deeper approach to learning

- involve students in critical reflection

- develop in students a better understanding of their own subjectivity and judgement.

This approach to planning and evaluation is not possible without mutual trust and respect. These have to develop over time, but that development is made more probable by starting out with an understanding of the local culture and customs. Respecting students and the knowledge and skills they have will go a long way towards promoting long-term trust and involvement. The other necessary aspect of any participatory process is appropriate training for everyone involved. Peer group members should be aware that evaluation takes place. The real purpose of an evaluation is not just to find out what happened, but to use the information to make the evaluation better. Peer evaluation can:

- Empower students to take responsibility for and manage their own learning.

- Enable students to learn to assess and give others constructive feedback to develop lifelong assessment skills.

- Enhance students' learning through knowledge diffusion and exchange of ideas.

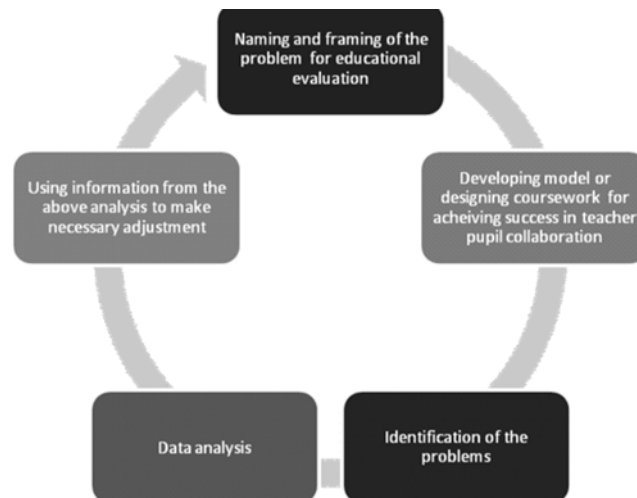- Motivate students to engage with course material more deeply.

Fig: A framework for participatory Environment in School

Considerations for using peer evaluation in evaluation: For successful execution of peer participatory evaluation the following preparations are necessary.

- Students should know the rationale for doing peer review. They should explain the expectations and benefits of engaging in a peer review process.

- Students should be considered having evaluated anonymous assignments for more objective feedback.

- Teachers should be prepared to give feedback on students' feedback to each other. Some examples of feedback of varying quality must be displayed and discussed which kind of feedback is useful and why.

- Students should be given clear directions and time limits for in-class peer review sessions and set defined deadlines for out-of-class peer review assignments.

- Teachers should listen to group feedback discussions and provide guidance and input when necessary.

- Student familiarity and ownership of criteria tend to enhance peer assessment validity, so students should be involved in a discussion of the criteria used.

- Teachers should identify assignments or activities for which students might benefit from peer feedback.

- Teachers must consider breaking a larger assignment into smaller pieces and incorporating peer assessment opportunities at each stage. For example, assignment outlines first draft, second draft, etc.

- Teachers should design guidelines with clearly defined tasks for the reviewer.

- Teachers should introduce rubrics through learning exercises to ensure students have the ability to apply the rubric effectively.

- Teachers should determine whether peer review activities will be conducted as in-class or out-of-class assignments.

- Teacher helps students learn to carry out peer assessment by modeling appropriate, constructive criticism and descriptive feedback through his or her own comments on student work and well-constructed rubrics.

- Teachers should incorporate small feedback groups where written comments on assignments can be explained and discussed with the receiver.

Significance of peer/participatory evaluation in education: Peer or participatory evaluation in education carries the following significances;

- **Perspective evaluation:** It gives us a better perspective on both the initial needs of the students' evaluation, and on its ultimate effects. If students are involved from the beginning in determining what needs to be evaluated and why they are much more likely to aim their work in the right direction, to correctly determine whether the evaluation is effective or not, and to understand how to change it to make it more so.

- **Generates reliable information:** It can find out information one would not get otherwise. When evaluation depends, at least in part, on information from a peer group, that information will often be more forthcoming if it is asked for by someone familiar. Peer group members interviewing their friends may get information that an outside person would not be offered.

- **Point out the loopholes:** It tells us what worked and what did not from the perspective of those most directly involved. Those implementing the evaluation and those who are directly affected by it are most capable of sorting out the effective from the ineffective. It also points out why something does or does not work properly. Students are often able to explain exactly why they did not

respond to a particular technique or approach, thus giving the teacher a better chance to adjust it properly.

● **Effective evaluation:** It brings out a more effective result. Students are much more apt to start out in the right direction, and to know when they need to change direction. The consequence is a project that addresses the appropriate issues in the appropriate way, and accomplishes what it sets out to do.

● **Increases students' responsibility:** It empowers students that also brings a sense of responsibility. Participatory evaluation gives those who are often not consulted, particularly the students, the chance to be full partners in determining the direction and effectiveness of a teaching learning.

● **Brings out sense of involvement:** It can provide a voice for those who are often not heard.   Students who seldom have and often do not think they have a right to the chance to speak for themselves, by involving them from the beginning in evaluation, teachers assure that their voices are heard, and they learn that they have the ability and the right to speak for them.

● **Training for real life practices:** It teaches skills that can be used in employment and other areas of life. In addition to the development of basic skills and specific research capabilities, participatory evaluation encourages critical thinking, collaboration, problem-solving, independent action; meeting deadlines and all other skills valued by employers, and useful in family life, education, civic participation, and other areas.

● **Fulfills self esteem needs:** It bolsters self-confidence and self-esteem in those who may have little of either. This category can include those students who may, because of circumstance, have been given little reason to believe in their own competence or value to society. The opportunity to engage in a meaningful and challenging activity, and to be treated as a colleague by professionals, can make a huge difference for folks who are seldom granted respect or given a chance to prove themselves.

● **Build up dynamic attitude:** It demonstrates to people ways in which they can take more control of their lives. Working with professionals and others to complete a complex task with real-world consequences can show people how they can take action to influence people and events.

● **Enhances leadership:** It encourages students' leadership of the academic

activity. If those involved feel the project is theirs, rather than something imposed on them by others, they will work hard both in implementing it, and in conducting a thorough and informative evaluation in order to improve it.

- **Nurturing creativity:** It can spark creativity in everyone involved. For those who have never been involved in anything similar, a participatory evaluation can be a revelation, opening doors to a whole new way of thinking and looking at the world. To those who have taken part in evaluation before, the opportunity to exchange ideas with people who may have new ways of looking at the familiar can lead to a fresh perspective on what may have seemed to be a settled issue.

- **Enhances collaborative working:** It encourages working collaboratively. For participatory evaluation to work well, it has to be viewed by everyone involved as collaboration, where each participant brings specific tools and skills to the effort, and everyone is valued for what he or she can contribute. Collaboration of this sort not only leads to many of the advantages described above, but also fosters a more collaborative spirit for the future as well, leading to other successful community projects.

- **Community involvement:** It fits into a larger participatory effort. In order to conduct a good evaluation, its planning should be part of the overall planning of the evaluation project. Furthermore, the participatory process generally matches well with the philosophy of community-based or grassroots groups or organizations outside the school.

Advantages: Peer or participatory evaluation has following advantages

- Agreed marking criteria means there can be little confusion about assignment outcomes and expectations.

- Encourages student involvement and responsibility.

- Encourages students to reflect on their role and contribution to the process of the group work.

- Focuses on the development of student's judgment skills.

- Students are involved in the process and are encouraged to take part ownership of this process.

- Provides more relevant and authentic feedback to students as it is generated by their peers.

- It is considered fair by some students, because each student is judged on their own contribution.

- When operating successfully can reduce a lecturer's marking load.

- Can help reduce the 'free rider' problem as students are aware that their contribution will be graded by their peers.

Disadvantages: Limitations or disadvantages are also there

- Additional briefing time can increase a lecturer's workload.

- The process has a degree of risk with respect to reliability of grades as peer pressure to apply elevated grades or friendships may influence the assessment, though this can be reduced if students can submit their assessments independent of the group.

- Students will have a tendency to award everyone the same mark.

- Students feel ill equipped to undertake the assessment.

- Students may be reluctant to make judgments regarding their peers.

- At the other extreme students may be discriminated against if students 'gang up' against one group member.

- Chances of bullying and a sense of rivalry may increase among peer groups.

Students may have little exposure to different forms of assessment and so may lack the necessary skills and judgements to effectively manage self and peer evaluation. There may also be a perception amongst students that the academic is 'shirking' their responsibilities by having students undertaking peer evaluation. In this situation students may be reminded of the Graduate Student Attributes. This also highlights the need to fully prepare and equip students for their own assessment and for the assessment of others. It is helpful to introduce students to the concepts and elements of assessment against specified criteria in the first weeks of class when teachers explain the unit of study outline. This requires taking time at the outset of the group activity or unit of study to discuss what is required, and to provide guidance on how to judge their own and others' contributions. Students will need to be assisted to develop criteria that match the learning outcomes with regards to the output and process of the group work. If assessment criteria for each element are set up and clearly communicated, the teacher's role will also change to one of facilitator.

4.6 Summary

Evaluation is a continuous process that is concerned with more than the academic achievement of students. There are two common purposes in educational evaluation which are, at times, in conflict with one another. Educational institutions usually require evaluation data to demonstrate effectiveness to funders and other stakeholders, and to provide a measure of performance for marketing purposes. Educational evaluation is also a professional activity that individual educators need to undertake if they intend to continuously review and enhance the learning they are endeavoring to facilitate. Educational evaluation can be broadly categorized into two types; Formative evaluation and Summative evaluation. Formative evaluation of a programme is designed to monitor the instructional process to determine whether the learning is actually taking place as planned earlier. This type of evaluation provides continuous information that can be used to modify the programme to improve its effectiveness and efficiency. In a classroom situation, the purpose of formative evaluation is to provide feedback to the teacher and students about progress in unit tests, measures of interest and attitude. Interviews or conferences with students and parents during the programme can provide important indications for improvement and betterment of the programme by redirecting it. These are mostly teacher made tests. At the end of the teaching of a chapter the teacher takes a test or quiz. Then the teacher reviews the test result to determine which material requires further discussion. Whereas, Summative evaluation usually occurs after instruction. It is designed to determine how well the instructional objectives are met. Formal classroom tests, such as unit tests or final examinations are the most frequently used tools used in this evaluation. The term summative means summing up all the available information regarding a programme at its terminal point. Such information can be a valuable way of assessing the effectiveness of the whole programme. It also provides correction if the programme continues. The techniques used here are teacher made, evaluation product such as achievement test, research, report, themes, drawing etc. it also depends on rating on different types of performances such as oral report, laboratory experiments etc.

Traditional evaluation system is criticized by modern educationists. As a substitute to traditional evaluation, Continuous Comprehensive Evaluation or CCE is very popular. In CCE teachers not only assess the curriculum as implemented in the classroom, it also allows teachers to evaluate the effectiveness of their teaching strategies relative to the curriculum, and to change those strategies as dictated by the needs of their

pupils. The addition of CCE in the instructional and testing process is intended to achieve two major purposes: to improve both the validity and reliability of the results of pupils' performance on tests and exercises, and secondly to help the pupil to develop effective learning and work habits. Classroom tests are based on assessment of lower level abilities and rote memorisation. Where assessments are based on low level thinking skills i.e., "Knowledge" and "Comprehension", pupils complete their education still unable to analyze and apply their knowledge to solve problems. The central purpose of CCE is to help the pupil to become a better learner.

Peer or participatory evaluation in education is a modern day concept. In peer or participatory evaluation students individually assess each other's contribution using a predetermined list of criteria. Grading is based on a predetermined process, but most commonly it is an average of the marks awarded by members of the group. The objective of such evaluation is making students aware, responsible, and cooperative and motivating in evaluation of the teaching learning process. Peer or participatory evaluation can be effective if some precautions are taken carefully before execution.

4.7 Self Assessment Questions

1. State the characteristics of formative evaluation

2. What is the significance of summative evaluation?

3. Differentiate between formative and summative evaluation.

4. Differentiate between process and product evaluation.

5. What are the types of Continuous Comprehensive evaluation?

6. Discuss the significance of Peer evaluation in education.

7. What are the advantages and disadvantages of Peer evaluation in education?

8. Name the tools and techniques used in CCE.

9. What are the limitations of CCE?

10. Discuss the considerations needed for preparation of peer or participatory evaluation in education.

4.8 References Books

- Aggarwal, J.C. (1997): Essentials of examination System: Evaluation, test and

measurement, Vikash Publishing House Private Limited, Delhi, page no. 190-195

- Dash, M. (2004): Education in India: Problems and Perspectives, Atlantic Publishers and Distributors, New Delhi, India, ISBN 81-7156-881-5, page no 176-186

- Sarkar, Amal kanti (2010): Measurement and Evaluation in Education, Kolkata, India, Rita Book House, page no 22-26

- Lovely Professional University (2012): Educational Management and Evaluation DEDU504, USI Publications, New Delhi, page no. 250-260, 352-362

Web Reference

- https://www.sydney.edu.au/education_social_work/groupwork/docs/ SelfPeerAssessm ent.pdf, retrieved on 11.07.2020

- https://ctb.ku.edu/en/table-of-contents/evaluate/evaluation/participatory-evaluation/ main#:~:text=Participatory%20evaluation%20involves%20all%20the,the%20improve ment%20of%20the%20work., retrieved on 20.07.2020

- https://teaching.cornell.edu/teaching-resources/assessment-evaluation/peer-assessmen t retrieved on 01.07.2020

- https://en.wikipedia.org/wiki/Educational_evaluation (visited on 17.11.2017)

- https://en.wikipedia.org/wiki/Evaluation#Methods_and_techniques (visited on 17.11.2017)

Unit 5 ❑ Types of Testing

 Structure

5.1   **Objectives**

5.2   **Introduction**

5.3   **Open Book Testing: Concept and Characteristics**

5.4   **Norm referenced Testing: Concept and Characteristics**

5.5   **Criterion referenced Testing: Concept and Characteristics**

5.6   **Summary**

5.7   **Self Assessment Questions**

5.8   **References**

5.1 Objectives

After going through the unit students will be able to;

● Elaborate the concept of Open book testing

● State the characteristics of Open Book testing in today's educational evaluation

● Define Norm Referenced Testing

● Discuss the characteristics of Norm Referenced Testing

● Explain the concept of Criterion Referenced Testing

● Elucidate the characteristics of Criterion Referenced Testing

● Differentiate between Norm Referenced Test and Criterion Referenced Test with examples

5.2 Introduction

Unlike traditional closed book examinations where students are required to rely solely on their recall of topics, open book tests allow students to consult reference materials (notes, textbook, etc.) in the course of completing the examination. But it is not any easier just because the test is an open book or any less challenging, than a closed book examination. In fact, the opposite is often true. While closed book tests place emphasis

on rote memorisation and recall, open book tests place focus on higher level learning and typically require students to evaluate, analyze, or synthesize information, rather than simply remember it. The strategies outlined below will help improve students' performance on open book tests in this unit. According to Robert Glaser's point of view, tests are of two types; Norm Referenced Test and Criterion Referenced Test. A test designed to provide a measure of performance that is interpretable in terms of an individual's relative standing in some known group is called Norm Referenced Test. On the other hand, a test designed to provide a measure of performance that is interpretable in terms of a clearly defined and delineated domain of learning task is called Criterion Referenced Tests. In this unit the characteristics and significance of both the tests will be discussed in detail.

5.3 Open Book Testing: Concept and Characteristics

An "open book testing" is an assessment method designed in a way that allows students to refer to class notes and summaries or a "memory aid", textbooks, or other approved material while answering questions. An open book test may also mean that students are provided with the examination questions prior to sitting the formal examination or are to complete as a 'take-home' examination.

Concept: Open-book tests allow one to take notes, texts or resource materials into an examination situation. They test the ability to find and apply information and knowledge, so are often used in subjects requiring direct reference to written materials, like law statutes, statistics or acts of parliament. Open-book examinations usually come in two forms: Traditional sit-down / limited-time tests, with varying degrees of access to resources and references. Take home open-book tests are conducted at home. Questions are handed out, answers are attempted without help from others, and the answer script is returned within a specified period of time usually the next day. In an open book examination, it is meaningless to ask questions like "Define the term education", since all that the student has to do is copy the relevant information from the textbook directly into the answer book. In a closed book examination, the student first copies the information from the textbook to his memory, and then copies it into the answer book. Open book test attempts to eliminate this intermediate stage of memorization. Given the availability of textbooks in the examination room, teachers will not ask questions that require the mere transfer of information from the textbook to the examination book. Needless to say, indirect problem solving questions that test the students' thinking skills can be used in closed book examinations as well.

The essential difference between closed book examinations and open book examinations is that the former can still be used to evaluate how much the students have memorised, while the latter cannot. If the purpose is not only testing memorised information closed book examinations should not be used. It is needless to say that memory has no value in education, but there are differences between passive and static memory, involved in rote learning, from active and dynamic memory, involved in accessing memory as part of creative and critical thinking. Memorization without understanding involves mechanical memorisation.

The materials one can take into open-book tests vary depending on the examination authority. Some restrict the type of materials, e.g. formula sheets, tables or a limited number of texts, while others may be totally unrestricted (any quantity of any material). Materials might be notes, readings, reference materials or textbooks, equipment like calculators; drafting tools etc. Materials used in take-home tests are usually unrestricted. Students are allowed to work with materials on their own without any help from others. But there are some misconceptions about open book testing. These are as follows;

1) **Open-book tests are a breeze:** Open-book tests are not an easy option. Answering the questions well requires more than just copying information straight from texts. For example, having access to a textbook can stop one from giving a wrong answer if the examinee cannot remember a fact or formula, but just getting that fact correct would not score good marks here. In open-book examinations, it is how one can locate, apply and use the information that is important. Open Book tests basically emphasise on the comprehension and application ability of the student.

2) **No requirement of study:** Probably the biggest misconception about open-book examinations is that there is no need to study. One should study just like any other examinations. Having books and notes to refer may help students not to memorise as much information, but still they need to be able to apply it effectively. This means they must fully understand and be familiar with the content and materials of course so they can find and use the appropriate information. In open-book tests, students need to quickly find the relevant information in the resources they have. Without study no one would be able to find out the information.

3) **Copying is allowed from the book:** Examinees cannot copy chunks of text

directly from textbooks or notes in any open book test. Copying without acknowledgement is called plagiarism. In open-book examinations, the resource materials are made available to students, so they are expected to do more than just reproduce them. They must be able to find, interpret and apply the information in sources to the examination questions. One usually needs to reference as well, just as one would for any other assignment.

4) **The more materials the better:** This is another misconception with students. Because one should not get carried away and overload with materials and resources in the examination, they should take only what they need. Stacks of books would not necessarily guarantee good performance, and there will not be time for extensive reading.

Types of Open Book tests: One may think of two kinds of open book examinations; the restricted type and the unrestricted type.

I. **Restricted Open Book Test:** In the restricted type of open book examinations, students are permitted to bring into the examination room one or more specific documents approved by the course instructor. Here students may be permitted to consult printed documents such as the logarithmic tables, dictionaries, or complete works of Shakespeare, but no handwritten material or printed documents which did not have prior approval. One may also need to make sure that the printed documents that students bring do not contain any scribbles on the margin. In this type of examination, the approved documents function more or less as appendices to the question paper itself. These examinations are not radically different from closed book examinations. They do not present any special problems, irrespective of the nature of the course.

II. **Unrestricted Open Book Test:** In the unrestricted type of open book examinations, students are free to bring whatever they like. Here questions are set in such a way that no prepared material will be of any use. So there are no restrictions on what the students can bring in an unrestricted open book examination. They may bring any books (with or without scribbles on the margin), lecture handouts of the course instructor, or their own handwritten notes. The use of such examinations presupposes certain teaching strategies and types of questions. In particular, it demands that the course focuses on a set of intellectual skills, rather than on the information content, and that no content based questions be asked in the examination. When used properly, it

will be pointless for students taking the unrestricted open book examinations to consult any material they have brought, because the questions will be designed in such a way the answers will not be found in the textbooks, handouts or class notes. The use of these examinations then acts as a symbolic gesture that makes the students realise the nature of the course and the examinations, and indulge them into a mode of studying that does not involve cramming.

Need of Open Book Testing:

- The main premise for open book tests is that teachers can set questions that require students to answer in more critical and analytical ways thus encouraging high-order thinking skills in their students; as compared to closed book or traditional examinations that tend to encourage rote learning and more superficial application of knowledge.

- Open-book tests are used for more than just rote-learning as it does not emphasise on memory level learning. At higher studies simply memorising and repeating information is not enough to score high. Higher education is supposed to equip students with intellectual abilities and skills. Open-book tests judge ability to quickly find relevant information and then to understand, analyse, apply knowledge and think critically.

- Open-book tests also evaluate memory; ability to find and use information for problem solving, and to deliver well-structured and well-presented arguments and solutions.

- Open-book test questions usually require applying knowledge, and they may be essay type questions or involve problem solving or delivering solutions. The style of question depends on the faculty or school setting the test.

Characteristics: Following characteristic features are observed in Open Book Testing.

- **Preparation is important:** Open book testing focuses on "higher level learning". It means that focus is shifted from rote memorization or recall to understanding and application. The purpose of the open book testing is to test students' understanding of a subject matter. Students will be asked to formulate, compare, analyze, evaluate, or synthesize information, rather than just recall it. Consequently, open book examinations are often more challenging than

other types of tests even with a plethora of reference materials at fingertips. Since open book tests can be more challenging than other types of test, students must adequately prepare.

- **Reference materials are not emphasized:** During an open book test, especially a timed examination, extensive notes, text books and other reference materials can quickly become a liability rather than an asset. Students do not have time to review notes, re-read text books, or discover new information during an open book examination. Heavy dependence on study material reduces the time.

- **Reference materials are well organized:** All reference materials brought to the examination centre should be carefully organized and that one should know ahead of time exactly where all information is available in answering questions. The main purpose is locating information from class lecture notes, books. In the case of open book mathematics and science tests, important data and formulas are listed separately for easy access during the test.

- **Major concepts are familiarized:** Open book tests usually cover more information and take longer to complete than traditional tests. Students will likely run out of time if they rely on books for every answer. Books and notes should only be used as reference resources for difficult questions or questions that require specific information.

- **Plagiarisation is prohibited:** Students are not expected to copy information directly from the book and write answers on an open book test. Teachers are concerned with what their students have learned, not whether they can locate information in a book. It is permissible to use quotes from the book, or lecture notes, to support viewpoint or analysis, but heavy reliance on them is not expected.

- **Effective time management:** The same time management rules that apply to general test taking also apply to taking open book tests. First, at the offset of the test one must review how many questions there are and determine how much time one can spend on each question. Then questions are answered according to preference of familiarity. Finally, concise, accurate, yet thorough answers are written.

Importance of Open Book Testing in Teaching: An "open book examination"

is one in which examinees are allowed to consult their class notes, textbooks, and other approved material while answering questions. This practice is common in law examinations, but in other subjects, it is mostly unheard of till date. Although the idea may sound puzzling to those who are used to conventional examinations, it is ideally suited to teaching programmes that especially aim at developing the skills of critical and creative thinking.

- **Teaching as Transmitting Information:** In order to appreciate the merits of open book testing, it is first of all necessary to understand the nature of teaching programmes. Traditionally teaching was "dissemination of knowledge". This approach to education treats the information content of a subject to be the most important and memory level teaching was emphasized. But in modern times the teacher's role is viewed as facilitating the transfer of information from the textbook to the students' minds. Today students are expected to understand this information, retain it, and retrieve it during the final examination. Based on the above approach, most conventional examinations test how much information the students have been able to store in their minds. In order to fulfill this expectation, students develop a habit of storing and cramming information where they can only memorise the information in class notes and textbooks, and transfer it to answer books during the examination. In this type of examination, success depends on the quantity of information memorised, and the efficiency with which it is reproduced. In this case a restricted type of open book examination is preferred.

- **Teaching as fostering Mental Development:** Constructivists claim that students construct their own knowledge. Therefore, an alternative view is that teaching should not be just transferring information from the library or textbooks to the students' minds. Rather, true teaching is teaching students how to learn. Today learning is being emphasized and eventually lesson plans are being replaced by learning designs. That is why teaching should equip students with the ability to acquire knowledge, to modify existing knowledge on the basis of new experience, to build new knowledge, and to apply available knowledge to solve problems and make wise decisions. Education is not only meant for obtaining a degree, but the lifelong process of mental development that does not terminate with any degree. If this view of education is accepted, then the main focus of teaching will be shifted on the skills of acquiring, modifying and creating knowledge, that is, on processing information, rather

than on the information content itself. In other words, the focus shifts from rote learning to the development of certain mental faculties. The teacher's function then is not summarising the information in the textbook but ensuring an environment that triggers the development of these creative and critical faculties. This can be done by activating learning through questions, exercises, projects, assignments, and so on, and sustaining and guiding it by providing comments, criticisms, and other forms of feedback. We may say that subjects taught in a high school or university should aim to develop the students' intellectual fitness, intellectual stamina, and intellectual flexibility. Conventional memory testing must give way to examinations that test the intellectual skills of the student. This is where open book examinations come in. questions based on information can be easily answered from books, but questions based on concepts need to be answered from their own critical appraisal skill. On the other hand, if the examination tests the skills of problem solving and critical thinking, then there is no harm in students consulting their text books and class notes. If students have to evaluate a conclusion that crucially involves their understanding of the concept, reproducing what the textbook says would be pointless. While open book examinations are not suited to testing rote learning, they can be used effectively to train students in the use of active memory.

● Impact on joyful Learning Strategies A more important reason for using open book examinations is that they have a tremendous impact on promoting the right mental sets in both learning and teaching. The most immediate result for students will be that they will stop "mugging" or cramming information or rote learning. Most students prepare for conventional examinations by mechanical memorisation of information in textbooks and class notes in order to reproduce it in examinanations. Open book examinations will effect a fundamental change in this attitude and study habit. If textbooks can be consulted in the examination rooms, no one would bother to memorise them. They will study not to memorise, instead, it should be understanding concepts, and using these concepts (along with available information) to practise the skills of modifying and building knowledge, thinking critically, and solving problems. In acquiring the right strategies of studying, nothing is as effective as the shocking realization that mugging is of no use in the examinations. Learning will be free of the burden of just storing information; education can

be a pleasurable activity, not a painful drudgery. What is learnt with pleasure is learnt more effectively, and retained better.

- Impact on Teaching Strategies The effects of open book examination is not only effective on learning but equally effective on teaching strategies. For this purpose the question pattern will be changed. Instead of using action verbs like writing an essay, explaining the term, defining something, we will have to frame questions carefully and intelligently to test the students' understanding, and the skills of applying that understanding. With the changing question pattern, strategies for preparing students to take those examinations will also have to change. It will no longer be enough to paraphrase or simplify the content of the text books in the classroom. Teachers will have to design tasks that will provide exercises for the appropriate mental skills required in each subject. Instead of the teacher talking and students taking down notes throughout the class, classes will have discussions, questions, and other active processes. In other words, teaching will no longer be the transfer of information from the teacher to student. Rather it will be the training of the mind in certain intellectual skills. Thus, open book examinations can restore the true meaning of the word education for both teachers and students. It will surely demand more time, energy and effort from both teachers and students to adapt themselves to the demands of open book examinations. But the changes will be inevitable. When combined with the mode of teaching that focuses on thinking skills, they will make education an exciting and enjoyable intellectual adventure and the beginning of a lifelong thirst for knowledge.

Precautions for Open Book Testing: As open book testing is not common in schools teachers and students need to prepare for this. The main issues that arise when making use of open book testing is that teachers may not know how to develop and devise effective examination questions that require students to apply their knowledge through analysis and critical thinking; and students may fail to properly prepare for an open book testing. Students may falsely assume that the testing will be easy, and they will be able to find all the answers in the textbook or on their memory aid. Following are the considerations when designing open book testing.

- Questions in open book tests need to be devised to assess the interpretation and application of knowledge, comprehension skills, and critical thinking skills rather than only knowledge, recall.

- Case-based examination questions are needed to be designed that require students to apply critical thinking, reasoning skills in response to a trigger scenario

- Clear and unambiguous questions are to be devised to limit student confusion and time spent interpreting the question so students can spend their time making use of their textbook or memory aid to effectively answer the questions

- Questions should be framed in such a way that require students to apply and make use of the information from their textbook or notes rather than simply requiring them to locate and re-write this information

- Questions should be designed keeping in mind the learning outcome. i.e. what skills and knowledge are being assessed from students.

- Questions should be problem-based scenarios or real-world cases, requiring students to apply their skills and knowledge to the given problem or scenario

- Information or background information on a given topic or area of study should be provided.

- Relevant qualitative or quantitative data must be presented first and then students should be asked interpretative and application questions.

- Content or topic questions should be structured in a way that tests for an ability to apply, analyse, evaluate, create, analyse, synthesise, interpret etc.

- When devising questions to probe student understanding, skills and knowledge, the Socratic questions and questions reflective of levels and stages of learning may be useful.

- Combating academic dishonesty is among the most challenging of the many obstacles that web-based technologies present in the recent past. Cheating becomes an appealing option when the response to a question is one that can be easily Googled. A student need not read a single chapter or attend any classes, if they know their smartphone will come to their rescue. An open-book test, with challenging application questions that relate directly to the course material, can help minimize the problem even with smartphone or laptop and internet connection.

With limited time students would not be able to blindly scavenge the course notes for the answer. They will recognize the need to prepare and have some familiarity with the material or they will simply run out of time.

5.4 Norm Referenced Testing: Concept and Characteristics

Generally we get a score after administering a test. On the basis of scores higher achievers are praised and lower achievers are identified for improvement. Marks obtained by a student in an achievement test are called raw scores. Raw scores do not indicate the merit of a student. From raw scores we cannot determine that children in school A are better in English than children in school B. For this comparison we need some broader, uniform, objective and stable standard of reference or unit of measurement in educational and psychological testing. The establishment of such reference points is important for both standardized tests and teacher made achievement tests. It is important for all sorts of tests, like intelligent tests, mathematical aptitude, personality tests etc. If the scores need value that is meaningful, significant, objective and stable we require comparing an individual pupil's performance with that of a group, either with a teacher made test or a standardized test, a group similar to his upon whom the test was developed. A score is not high or low, it is higher or lower than other scores. The better type of unit to express test results than raw score of a crude percentage is called norms. In other words, the medium or average performance on standardized tests of pupils of a teenage group or a placement or of a school group is termed as Norm. Such an average is obtained by administering the test to a large group of students who represent different geographical areas in a country comprising both boys and girls, rural and urban. Among these students there may be children from private and public schools. Such a group will provide a frame of reference for interpreting individual scores. Robert Glaser originally coined the terms Norm-Referenced Test and Criterion-Referenced Test.

Concept of Norm-Referenced Test (NRT): In the last four decades revolutionary changes have taken place in the field of teaching and learning processes. As individual instruction has gained importance tests should be designed in such a way that reflects the changes in what the individual has learned. In the case of Norm Referenced Test (NRT) one interprets learners' test performance according to the performance of others. A norm-referenced test is not measured against defined criteria. This type of test is relative to the student body undertaking the assessment. It is a very effective method of comparing students. The IQ test is the best known example of NRT. Many entrance tests are examples of NRT, permitting a fixed proportion of students to pass.

Meaning and definition of NRT: Raw scores are expressed in terms of different units, such as number of trials taken within a specified period to reach a criterion.

When we compare an examinee's test score with the core of a specific score of examinees on that test the method is called Norm Referenced Test. When raw scores are compared to the norms, it becomes scientific. Norms may be defined as the average performance on a particular test made by a standardization sample. A standardization sample is the true representative of the population. It takes the test to provide data for comparison and subsequent interpretation of the test scores. To have an adequate representation the sample must include cross sectional representation of different parts of the population with a view to compare the raw scores with performance of the standardization sample. When these are converted they are called derived scores.

Gronlund (1976) defined NRT as the tests are "designed to rank students in order of achievement, from high to low, so that decisions are based on relative achievements (e.g., selection, grouping, grading) can be made with greater confidence".

Bormuth (1970) stated that NRT is designed "to measure the growth in a student's attainment and to compare his level of attainment with levels reached by other students and norm groups".

This test is primarily used for comparing achievement of an examinee to that of a large representative group of examinees at the same grade level. The representative group is known as the 'norm group'. Norm group may comprise of examinees at the local, district, state or national level. Since the development of NRT is expensive and time consuming they are produced by commercial test publishers. The following derived scores are generally reported for these published tests:

- **Percentile Rank:** A number telling the percentage of examinees in a defined group scoring lower than the particular raw score.

- **Linear standard score:** A number indicating the location of a particular raw score in relation to the mean and standard deviation of a defined group.

- **Normalized Standard Score:** A number showing the location of the particular raw score in relation to a normal distribution defined in terms of a particular group.

- **Grade Equivalent Score:** A number telling the grade placement for which the particular raw score is the average for a defined group.

Chief characteristics of NRT: Following characteristic features are observed in NRT;

1. The basic purpose of NRT is to measure students' achievement in curriculum based skills.

2. Most achievement NRTs are multiple-choice tests. Some also include open-ended, short-answer questions. The questions on these tests mainly reflect the content of nationally-used textbooks, not the local curriculum. This means that students may be tested on things that local schools or state education department decided were not so important and therefore were not taught.

3. Generally it is reported in the Percentile Rank, Linear standard score, Normalized Standard Score, Grade Equivalent Score. Scores are usually reported as percentile ranks. The scores range from 1st percentile to 99th percentile, with the average student score set at the 50th percentile. If John has scored at the 23rd percentile, it means he scored higher than 23% of the test takers in the norming group. Scores also can be reported as "grade equivalents," "stanines," and "normal curve equivalents."

4. Norm-Referenced Tests (NRTs) compare a person's score against the scores of a group of people who have already taken the same test, called the "norming group." For example, someone has scored at the 63rd percentile.

5. NRTs are designed to "rank-order" test takers to compare students' scores. A commercial norm-referenced test does not compare all the students who take the test in a given year. Instead, test-makers select a sample from the target student population (say, ninth graders). The test is "normed" on this sample, which is supposed to fairly represent the entire target population (all ninth graders in the nation). Students' scores are then reported in relation to the scores of this "norming" group.

6. The result looks like a bell-shaped curve. Test makers make the test so that most students will score near the middle, and only a few will score low (the left side of the curve) or high (the right side of the curve).

7. One more question right or wrong can cause a big change in the student's score. In some cases, having one more correct answer can cause a student's reported percentile score to jump more than ten points. It is very important to know how much difference in the percentile rank would be caused by getting one or two more questions right.

8. Selection of questions is important because that sort people along the curve

than it is to make sure that the content covered by the test is adequate. The tests sometimes emphasize small and meaningless differences among test takers. Since the tests are made to sort students, most of the things everyone knows are not tested. Questions may be obscure or tricky, in order to help the test taker in ranking order.

9. Tests can bring about biased results. Some questions may favour one kind of student or another for reasons that have nothing to do with the subject area being tested. Non-school knowledge that is more commonly learned by middle or upper class children is often included in tests. To help make the bell curve, test makers usually eliminate questions that students with low overall scores might get right but those with high overall scores get wrong. Thus, most questions which favour minority groups are eliminated.

10. NRTs usually have to be completed in a time limit. Some students do not finish, even if they know the material. This can be particularly unfair to students whose first language is not English or who have learning disabilities. This "speediness" is one way test makers sort people out.

11. All tests have "measurement error." There are many possible causes of measurement error. Sub-scores on tests are even less precise. Scores for young children are much less reliable than for older students. Teaching to the test explains why scores usually go down when a new test is used.

12. It is prepared for a particular grade level.

13. It is administered after instruction.

14. It is used to form homogeneous or heterogeneous class groups.

15. It classifies achievement as above average, average and below average for a given grade.

Reliability of NRT: The following factors can increase the reliability.

- As we know that test length affects reliability, other things being equal, the reliability of the test can be increased by increasing its length.

- Homogeneous items in content increases reliability.

- Moderate difficulty value of items increase reliability over the extremely easy or difficult value

- Increased range of performance of the examinees being tested tends to increase reliability

Validity of NRT: The validity of NRT can be increased by:

- Constructing items of proper difficulty level

- Increasing the test length.

- Increasing the heterogeneity of the group which is being tested.

- Administering the test under proper conditions.

Interpretation of NRT:

- If the scores obtained by students deviate largely from expected scores by a regular classroom teacher, then the teacher should explore by instructional implication.

- If the range of scores obtained by the student in different subjects varies largely the teacher should try to locate the reason for it motivation, special interest and abilities or difficulties.

- If the test results do not match the expectation of the teacher for individual students the validity of the test is said to be threatened.

- Small differences in sub-test scores may be due to chance error also for which the standard error of measurement should be used.

- Information collected from various tests should be used to explain performance on other tests.

Uses of NRT: N. Vasantha Ram Kumar and K.N. Lalithamal (1990) stated that NRT is useful in the following test:

1. In aptitude testing for making differential prediction.

2. To get a reliable rank ordering of the pupils with respect to the achievement that is measured.

3. To identify the pupils who have mastered the essentials of the course more than others.

4. To select the best of the applicants for a particular programme.

5. To find out the effectiveness of a programme in comparison to other programmes.

Application of NRT:

- Many college entrance examinations and nationally used school tests use Norm-Referenced Tests. The SAT, Graduate Record Examination (GRE), and Wechsler Intelligence Scale for Children (WISC) compare individual student performance to the performance of a normative sample. Test takers cannot "fail" an NRT, as each test taker receives a score that compares the individual to others that have taken the test, usually given by a percentile. This is useful when there is a wide range of acceptable scores, and the goal is to find out who performs better.

- California Achievement Test, the Iowa Test of Basic Skills (Riverside), and the Metropolitan Achievement Test (Psychological Corporation) are formed using a national sample of students. Because norming a test is such an elaborate and expensive process, the norms are typically used by test publishers for years. NRTs compare a person's score against the scores of a group of people who have already taken the same examination, called the "norming group."

- IQ tests are Norm-Referenced Tests, because their goal is to see which test taker is more intelligent than the other test takers. The median IQ is set to 100, and all test takers are ranked up or down in comparison to that level.

- Theatre auditions and job interviews are examples of NRT, because their goal is to identify the best candidate compared to the other candidates, not to determine how many of the candidates meet a fixed list of standards.

Advantages and limitations: NRTs possess some advantages and limitations as well. The primary advantage of NRT is that they can provide information on an individual's performance on the test compared to others in the reference group. The limitations are as follows;

- A serious limitation of NRT is that the reference group may not represent the current population of interest. The Oregon Research Institute's International Personality Item Pool website has notes that, one should be very wary of using canned "norms" because it is not obvious that one could ever find a population of which one's present sample is a representative subset. Most "norms" are misleading, and therefore they should not be used. Far more defensible are local norms, which one develops oneself. For example, if one wants to give feedback to members of a class of students, one should relate

the score of each individual to the means and standard deviations derived from the class itself. To maximize informativeness, one can provide the students with the frequency distribution for each scale, based on these local norms, and the individuals can then find (and circle) their own scores on these relevant distributions.

● Test items that are answered correctly by most of the pupils are not included in these tests because of their inadequate contribution to response variance. They will be the items that deal with important concepts of course content.

● There is a lack of congruence between what the test measures and what is stressed in a local curriculum. Norm-referencing does not ensure that a test is valid (i.e. that it measures the construct it is intended to measure).

● Norm referencing promotes unhealthy competition and is injurious to self concepts of low scoring students.

● Another disadvantage of norm-referenced tests is that they cannot measure progress of the population as a whole, only where individuals fall within the whole. Rather, one must measure against a fixed goal, for instance, to measure the success of an educational reform programme that seeks to raise the achievement of all students.

● With a Norm-Referenced test, grade level was traditionally set at the level set by the middle 50 percent of scores. By contrast, the National Children's Reading Foundation believes that it is essential to assure that virtually all children read at or above grade level by third grade, a goal which cannot be achieved with a norm-referenced definition of grade level. Norms do not automatically imply a standard.

● A Norm-Referenced Test does not seek to enforce any expectation of what test takers should know or be able to do. It measures the test takers' current level by comparing the test takers to their peers. A rank-based system produces only data that tell which students perform at an average level, which students do better, and which students do worse. It does not identify which test takers are able to correctly perform the tasks at a level that would be acceptable for employment or further education.

5.5 Concept and characteristics of Criterion Referenced Test (CRT)

Criterion Referenced Test is a test designed to provide a measure of performance that is interpretable in terms of a clearly defined and delimited domain of learning tasks. The word criterion refers to a domain of behaviour. In CRT one is interested in referencing an examinee's test performance to a well defined domain of behaviour measuring an objective or skill. It is often, but not always, used to establish a person's competence.

Concept of Criterion Referenced Test: When we follow a second method of interpreting a test score is to establish an external standard or criterion and compare the examinee's test score with it this process is known as Criterion Referencing. In this test there is a fixed performance criterion against which the candidates are measured. When an examinee passes some predetermined number of items or answers them correctly, it is said that it is capable of the total performance demanded by the test. Glaser (1963) first used the term CRT to highlight the need for tests that can describe the position of a learner on a performance continuum, rather than the learner's rank within the group.

A Criterion Referenced Test is a style of test which uses test scores to generate a statement about the behaviour that can be expected of a person with that score. Most tests and quizzes that are written by school teachers can be considered as CRT. In this case, the objective is simply to see whether the student has learned the material. Criterion-referenced assessment can be contrasted with norm-referenced assessment. Criterion-referenced testing was a major focus of psychometric research in the 1970s.

Meaning and definition of CRT: We can define CRT as one in which the test performance is linked or related to some behavioural measures. An obvious question may come to our mind is that from where do the criteria come with which a test is referenced? According to Cox and Vargas, a major criterion for referencing a test is the training that increases the skill or proficiency. The test scores of results can be interpreted as an indication of increased skill or proficiency.

A common misunderstanding regarding the term is the meaning of criterion. Many, if not most, CRT involves a cut score, where the examinee passes if their score exceeds the cut score and fails if it does not (often called a mastery test). The criterion is not the cut score; the criterion is the domain of subject matter that the test is

designed to assess. For example, the criterion may be "Students should be able to correctly add two single-digit numbers," and the cut score may be that students should correctly answer a minimum of 80% of the questions to pass.

The criterion-referenced interpretation of a test score identifies the relationship to the subject matter. In the case of a mastery test, this does mean identifying whether the examinee has "mastered" a specified level of the subject matter by comparing their score to the cut score. However, not all criterion-referenced tests have a cut score, and the score can simply refer to a person's standing on the subject domain. Examples can be cited from assessment where there is no cut score, it simply is an assessment of the student's knowledge of high-school level subject matter. Because of this common misunderstanding, CRTs have also been called standards-based assessments by some education agencies, as students are assessed with regards to standards that define what they "should" know, as defined by the state.

Gronlund, N.E. (1985) stated that CRT is a "test designed to provide a measure of performance that is interpretable in terms of a clearly defined and delimited domain of learning tasks.

Ivon (1970) defines a CRT as "one consisting of items keyed to a set of behavior objectives."

Glaser and Nitko in the book 'Educational Measurement' (1971) stated that "a CRT is one that is deliberately constructed to yield measurements that are directly interpretable in terms of special performance standards".

From above definitions it is concluded that CRT is meant to measure the achievement of an examinee on a certain domain to find out his level of achievement in that domain. It has nothing to do with the achievement level of other examinees.

Chief characteristics of CRT: The characteristic features of CRT are as follows;

1. Its main objective is to measure students' achievement of curriculum based skills. Criterion-referenced tests place a primary focus on the content and what is being measured.

2. Content validity plays an important role in CRT development. Item statistics play less of a role in item selection though highly discriminating items are still greatly valued, and sometimes item statistics are used to select items that maximize the discriminating power of a test at the performance standards of interest on the test score scale. A good criterion-referenced test will permit

content-referenced interpretations and this means that the content domains to which scores are referenced must be very clearly defined.

3. Each type of test can serve the other main purpose (norm-referenced versus criterion-referenced interpretations), but this secondary use will never be optimal. For example, since criterion-referenced tests are not constructed to maximize score variability, their use in comparing candidates may be far from optimal if the test scores that are produced from the test administration are relatively similar. Because the purpose of a criterion-referenced test is quite different from that of a norm referenced test, it should not be surprising to find that the approaches used for reliability and validity assessment are different too.

4. In criterion-referenced tests, scores are often used to sort candidates into performance categories. Consistency of scores over parallel administrations becomes less central than consistency of classifications of candidates to performance categories over parallel administrations. Variation in candidate scores is not so important if candidates are still assigned to the same performance category. Therefore, it has been common to define reliability for a criterion-referenced test as the extent to which performance classifications are consistent over parallel-form administrations. For example, it might be determined that 80% of the candidates are classified in the same way by parallel forms of a criterion-referenced test administered with little or no instruction in between test administrations.

5. With criterion-referenced tests, the focus of validity investigations is on (1) the match between the content of the test items and the knowledge or skills that they are intended to measure, and (2) the match between the collection of test items and what they measure and the domain of content that the tests are expected to measure. The "alignment" of the content of the test to the domain of content that is to be assessed is called content validity evidence. This term is well known in testing practices.

6. Many criterion-referenced tests are constructed to assess higher-level thinking and writing skills, such as problem solving and critical reasoning. Demonstrating that the tasks in a test are actually assessing the intended higher-level skills is important, and this involves judgments and the collection of empirical evidence. So, construct validity evidence too becomes crucial in the process

of evaluating a criterion referenced test.

7. Probably the most difficult and controversial part of criterion-referenced testing is setting the performance standards, i.e., determining the points on the score scale for separating candidates into performance categories such as "successful" and "unsuccessful." The challenges are great because with criterion-referenced tests in education, it is common on state and national assessments to separate candidates into not just two performance categories, but more commonly, three, four, or even five performance categories. With four performance categories, these categories are often called failing, basic, proficient, and advanced.

8. CRT is prepared for a particular course or grade level.

9. It has a balanced representation of both goals and objectives.

10. It is used to evaluate the curriculum plan instruction progress and group students' interaction.

11. It can be administered before and after instruction

12. It is normally reported in terms of minimum scores for partial and total mastery of main skill areas, number of correct items, and percent of correct items

13. Derived scores are based on correct items and other factors in case of CRT

Reliability of CRT: Reliability can be increased by taking the following measures;

● Test length should be sufficient enough to find out test score reliability.

● The sample of examinees in finding out reliability should be adequate and representative.

● The reliability information should be provided in the test for each intended use of the test score

● The reliability information provided in the test should be appropriate for the use of the scores of the test.

Validity of CRT:

● The validity evidence should be adequate for the intended use of the test score.

- The test manual should provide an appropriate discussion on the factor affecting the validity of the scores.

Interpretation of CRT:

- Interpretation should be based on what the items actually measure

- For each type of interpretation there should be sufficient items. If these are less than 10 items, they should be combined with other items through lengthening content items.

- The test should contain both easy and difficult items. Otherwise it would be difficult to describe what low achievers could do.

Uses of CRT: CRTs can be used in the following ways;

- To discover the weakness, inadequacies in learners are learned and assist the weaker section of learners to reach the level of other students through a regular programme of remedial teaching or instruction.

- To identify the master learners and average and slow learners in a class for modification in teaching strategy.

- To find out the level of attainment of various objectives of instruction and re-plan accordingly.

- To find out the level at which a particular concept has been learnt.

- For better placement of concepts at different grade levels.

- To make instructional decisions for teaching a learner in an individually prescribed instruction programme.

Advantage of CRT: CRT has certain advantages. These are as follows:

- The most important advantage of CRT is assessment of mastery of Subject Matter. Criterion Referenced Tests are more suitable than Norm Referenced Tests for tracking the progress of students within a curriculum. Test items are designed in such a way that can match specific objectives. The scores on CRTs indicate how well the individual can correctly answer questions on the material being studied, while the scores on a NRT report how the student scored relative to other students in the group.

- Moreover, Criterion-Referenced Tests can be managed locally. Every teacher must assess students' progress. Criterion Referenced Tests can be developed

at the classroom level. If the standards are not met, teachers can specifically diagnose the deficiencies, weakness of the student. Scores for an individual student are also independent of how other students perform. In addition, test results can be quickly obtained to give students effective feedback on their performance. Although NRTs are most suitable for developing normative data across large groups, CRTs can produce some local norms.

Limitations of CRT: According to Chase (1974) CRT has following limitations:

- CRT only shows whether a learner has reached proficiency in a task area but it does not indicate the extent of quality of learner's ability level.

- Tasks included in the CRT may be highly influenced by a given teacher's interest or biases, leading to general validity problems.

- CRT includes partial areas for listing behavioural objectives and this may be a constructing element for teachers.

- CRT is important for only a small fraction of important educational achievements. On the contrary, for promotion and assessment of various skills that are important functions of school need NRT.

Similarities between NRT and CRT: NRT and CRT have the following similarities:

- Both the tests require validity and reliability

- Achievement domain is measured in both the tests.

- Both the tests demand relevant and representative sample of test items

- Same types of items can be used in both tests.

- Same rules are followed for writing items in both accepting the item of difficulty.

Difference between CRT and NRT:

Sl. no	NRT	CRT
1	NRT covers a large domain of learning tasks with just a few items measuring each specific task	CRT focuses on a delimited domain of learning tasks with a relatively large number of items measuring each specific task

2	It gives emphasis on discrimination among individuals	It stresses what examinees can do and what cannot do
3	Contains items of average difficulty	Contains both easy and difficult items
4	Interpretation of test needs a defined group	Interpretation needs defined as well as delimited achievement domain
5	In NRT student is tested after each unit of the new material presented	In CRT the student is tested after each unit for mastery of objectives
6	Students are assigned marks or grade to indicate its performance	Students are allowed to proceed to new material if mastery is obtained
7	A student is allowed to go to the next unit and presented new materials along with the whole class	A student is allowed to proceed to the next unit if mastery is obtained after remedial instruction.

In India work on CRT is yet in the stage of infancy. There seems to be more emphasis on NRT in the evaluation of students. In spite of the dissimilarities among NRT and CRT, the tests also possess some common elements. Both the tests need reliability, validity, measured achievement domain, similarity of items; same rules are followed for writing items excepting the items of difficulty. Moreover, samples of test items should be relevant and representative in both the tests. Therefore, we cannot develop a clear distinction between NRT and CRT. Therefore, these two types of test may be considered as complementary. The only difference between these two tests is that CRT is always based upon a predetermined cutoff score whereas a NRT is always based upon the performance of a normative group or standardization sample.

5.6 Summary

Open Book testing is common in Law examinations, but it is still uncommon in the general stream of education. In this type of test students are allowed to take study

materials, text books and other reference books and equipment at the examination hall. It does not mean students do not have to study or they can directly copy from the books. If the purpose of an examination is to test the information that students have memorised, open book examinations are inappropriate, since students can easily transfer the information in the textbook or lecture notes to the examination paper. Therefore, the question pattern and objectives are important in Open Book Tests that normally differ from traditional closed book tests. Informative questions are generally avoided to test the critical thinking ability of the students.

Glaser (1963) has classified tests into Norm referenced test and Criterion referenced tests categories. NRT determines a student's placement on a normal distribution curve. Students compete against each other on this type of assessment. The major reason for using a norm-referenced test is to classify students. NRTs are designed to highlight achievement differences between and among students to produce a dependable rank order of students across a continuum of achievement from high achievers to low achievers. Most achievement NRTs are multiple-choice tests. NRTs are designed to "rank-order" test takers, that is, to compare students' scores. To make comparing easier, test makers create examinations in which the results end up looking at least somewhat like a bell-shaped curve. Scores are usually reported as percentile ranks. As one more question right or wrong can cause a big change in the student's score, it is often more important to choose questions that sort people along the curve than it is to make sure that the content covered by the test is adequate. Tests can be biased. The items on the test are only a sample of the whole subject area. It is often easiest to use a norm-referenced test because they were created to rank test-takers. NRTs are a quick snapshot of some of the things most people expect students to learn. Any one test can only measure a limited part of a subject area or a limited range of important human abilities.

On the other hand, in Criterion Referenced Tests questions are written according to specific predetermined criteria. A student knows what standards are set for passing and only competes against him or herself while completing. A criterion-referenced test is designed to measure how well test takers have mastered a particular body of knowledge. The term "criterion-referenced test" is not part of the everyday vocabulary in schools, and yet, nearly all students take criterion-referenced tests on a routine basis. Criterion-referenced tests place a primary focus on the content and what is being measured, ensuring content validity where item statistics play less a role in item selection though highly discriminating items are still greatly valued, and sometimes

item statistics are used to select items that maximize the discriminating power of a test at the performance standards of interest on the test score scale. With criterion-referenced tests, scores are often used to sort candidates into performance categories. Many criterion-referenced tests are constructed to assess higher-level thinking and writing skills, such as problem solving and critical reasoning. Anyway, the most difficult and controversial part of criterion-referenced testing is setting the performance standards, i.e., determining the points on the score scale for separating candidates into performance categories such as "successful" and "unsuccessful." The challenges are great because with criterion-referenced tests in education, it is common on state and national assessments to separate candidates into not just two performance categories, but more commonly, three, four, or even five performance categories. Criterion-referenced tests are more suitable than norm-referenced tests for tracking the progress of students within a curriculum and diagnose the weakness in the classroom. Test items can be designed to match specific programme objectives. Criterion-referenced tests also have some built-in limitations. Both valid and reliable test construction requires fairly extensive and expensive time and effort. In addition, results cannot be generalized beyond the specific course or program. Criterion-referenced tests are used in many ways. Classroom teachers use them to monitor student performance in their day-to-day activities. States find them useful for evaluating student performance and generating educational accountability information at the classroom, school, district, and state levels.

5.7 Self Assessment Questions

1. What is Open Book testing?

2. State the importance of Open Book testing in teaching learning.

3. Mention a few precautionary measures to be adopted during Open Book testing.

4. What is meant by norm referenced test?

5. What are the features of norm referenced tests?

6. Give the disadvantages of norm referenced tests?

7. What is the need of a norm referenced test?

8. What do you mean by criterion referenced test?

9. Give the advantages of the criterion referenced test.

10. Explain the feature of the criterion referenced test.

11. What are the applications of criterion referenced test

5.8 References Books

- Aggarwal, J.C. (1997): Essentials of examination System: Evaluation, test and measurement, Vikash Publishing House Private Limited, Delhi, page no. 68-69, 280-291

- Dash, B.N., Dash, Nibedita and Rath, M.P. (2003): A Textbook of Educational Psychology, Delhi, Dominant Publishers, page no. 483 to 492

- Sarkar, Amal kanti (2010): Measurement and Evaluation in Education, Kolkata, India, Rita Book House, page no 28-30

- Singh, A.K. (2002): Tests, Measurements and Research Methods in Behavioural Science, Patna, Bihar, India, Bharati Bhawan Publishers and Distributors, page no 3 to 87

- Singha, H.S. (1999): Modern Educational Testing, New Delhi, India, Sterling Publication Private limited, ISBN812-07-217-13, page no 79 to 99

- Lovely Professional University (2012): Educational Management and Evaluation DEDU504, USI Publications, New Delhi, page no. 181-193

Web References

- Mohanan, K.P., "Open Book Examinations" retrieved from http://www.iiserpune.ac.in/~mohanan/educ/openbook.pdf retrieved on 02.07.2020

- Farell Matt and Maheu Shannon (2015): article published in Faculty Focus on May 18, 2015. © Magna Publications, https://www.facultyfocus.com/articles/educational-assessment/why-open-book-tests-d eserve-a-place-in-your-courses/, retrieved on 10.07.2020

- https://www.educationcorner.com/openbook-tests.html, retrieved on 05.07.2020

- https://en.wikipedia.org/wiki/Criterion-referenced_test(visited on 17.11.2017)

- https://en.wikipedia.org/wiki/Norm-referenced_test(visited on 17.11.2017)

- https://student.unsw.edu.au/open-book-and-take-home-exams, retrieved on 10. 07.2020

Unit 6 ❑ Modern Trends in Evaluation

Structure

6.1 Objectives

After going through the unit students will be able to

● Discuss the shifts from traditional scoring to normalized scoring

● State the concept and need of online assessment

● Elucidate the concept of percentile ranking with example

● Explain the concept of grading

● Point out the uses of credit based cumulative grade point

● Use of averaging in educational evaluation

● Mention the significance of transcription in the contemporary education system.

6.2 Introduction

Almost all the Commissions on education as also the National Policy on Education (1986), Programme of Action (1992) and National Curriculum Framework (2000) of NCERT have all stressed the need for a more continuous and comprehensive evaluation of students in order to pass more sound judgments about students' learning and

growth. Regular Testing assessment through pupils' lives in school is considered essential for remedial treatment of those who fall below the acceptable performance level. It is in this context that the role of Continuous and Comprehensive Evaluation (C.C.E.) is highlighted in appraising the whole child and recognising it as a complement of external examination by providing due weightage to C.C.E. in the final assessment of the total development of the child. This current trend in educational evaluation has also changed the traditional marking and scoring system. Traditional system of absolute scoring is replaced by normalized scoring particularly in national level competitive examinations. Online assessment has become order of the day. Therefore, modern techniques like percentile ranking, grading, credit based cumulative grade point, averaging, transcription are invariably used in educational evaluation. In this unit all these new techniques will be discussed.

6.3 Current trends in evaluation: Shifting from traditional absolute scoring system to normal scoring system:

Evaluation is based on scoring. Scoring is the procedure used for training teachers to score students' work reliably within and across schools, and the processes used to evaluate score reliability and comparability across teachers and schools. Test scores often produce ordinal scales with some amount of meaning in their intervals. The particular rules for assigning values within these scales depend on the type of scoring mechanisms used. The two most common scoring mechanisms are dichotomous and polytomous, which are used to create rating scales and composite scores.

Dichotomous scoring: Dichotomous scoring refers to the assignment of one of two possible values based on a person's performance or response to a test question. A simple example is the use of correct and incorrect answers to score a cognitive item response. These values are mutually exclusive, and describe the correctness of a response in the simplest terms possible, as completely incorrect or completely correct. Most cognitive tests involve at least some dichotomously scored items. Multiple-choice questions are usually scored dichotomously. Dichotomous scoring could involve different score values, besides correct and incorrect. The most common example is scoring that represents a response of either yes or no. When we intend to measure affective domains, such as attitude surveys and behavior checklists, often use this type of dichotomous scoring. Opinionnaires or expression inventories, for example, may present individuals with lists of statements that people with depression typically identify

strongly with. Individuals then respond to each statement by indicating whether or not the statements are characteristic of them.

Polytomous scoring: Polytomous scoring is more subjective in nature than the previous one. It simply refers to the assignment of three or more possible values for a given test question or item. In cognitive testing, a simple example is the use of rating scales to score written responses such as essays. In this case, score values may still describe the correctness of a response, but with di?ering levels of correctness, for example, incorrect, partially correct, and fully correct. Therefore, Polytomous scoring with cognitive tests can be less straightforward and less objective than dichotomous scoring, primarily because it usually requires the use of human raters with whom it is di?cult to maintain consistent meaning of assigned categories such as partially correct.

Problems of traditional scoring system: At present we assess students on their learning for three different purposes: (a) with reference to his own progress (self-referenced), (b) With reference to his peer group (norm-referenced), and (c) with reference to criteria set up in terms of Intended Learning Outcomes (criterion-referenced). In all these cases the purpose may be different but the students are always awarded numerical marks on an interval scale of 0-100 marks. However, artificially created zero represents 'nothingness' and 100 represents 'perfection' in achievement. Both are fallacious, since zero does not mean that the student does not know anything, and 100 does not mean that the student knows everything about the subject. This makes comparison of scores difficult from test to test as well as from subject to subject. Moreover, there are chances of many types of measurement errors due to intra-examiner's and inter examiner's subjectivity in marking, inadequacy in sampling of content and competencies implied in instructional objectives, poor quality of evaluation tools (question papers), arbitrary time limit and variations in testing situations. Some of these errors are identifiable but others are not. The magnitude of such errors according to research reports varies from 7 to 15%. The 101 (0-100) point numerical scale cannot serve to discriminate individuals on the basis of a single score, i.e. 50, 51, 52 etc. If the standard error of measurement is 10, then the true score of 50 is likely to be between 40 and 60 in 2 out of 3 cases. It is therefore difficult to measure precisely an individual on a 101 point scale, when measurement errors are known to exist from 7 to 15%.

However, from Kothari Commission (1964-1966) to National Curriculum

Framework (2005) have emphasized on the inadequacies of evaluation system specially the lack of full disclosure and transparency in grading and mark reporting. In any educational process reporting of the results is very important. The symbols and signals of such reporting are usually marks.

Preceding the recommendation of NPE, 1986 the marks were replaced by grading system. There are factors responsible for unreliability of traditional absolute scoring systems.

1. Different standards of excellence both among different teachers and the same teacher on different occasions.

2. Psychological factors, such as due to fatigue unable to differentiate between closely allied degrees of merit.

3. Systematic changes.

4. Influence of handwriting.

5. Marks may carry different meanings to different persons in different countries and subjects.

Absolute grading is a mode of grading where grades are given based on predetermined cutoff levels. Here, each point value is assigned a letter grade. Most schools adopt this system, where it is possible for all students to receive an A. In the U.S.A, an absolute grading system assigns an A for points 90-100, a B for points 80-89, a C for 70-79, a D for points 60-69 and an F for points 59 and below. Absolute grading refers to a marking system where instructors pre-specify performance standards. But absolute grading is also inflexible, meaning that if everyone gets below 59 percent, all students in a class fail and a teacher would not make any adjustments to grades. Absolute grading is based around fixed percentage scales that may be constraining for students who all receive a low grade, for instance. Unlike absolute grading, in relative grading students' marks fluctuate depending on how they did in comparison to others in class. Relative grading refers to a system of evaluation that allows educators to convert the outcomes of a student's test, project or assignment and adjust that final grade in relation to grades from other students in the course. Relative grading is similar to bell curving or grading on a curve, and considers the highest score as the baseline (A+), relatively adjusting all others compared to that score.

In the recent past there is a shift from traditional absolute scoring to normal scoring in public examinations in India. The concept of normalisation of marks was

introduced because the level of difficulty of government examinations conducted in various shifts or boards/councils is different from each other. For example, a student who has appeared in the first session or shift of the examination might have scored low marks. However, the same student would have scored more or even higher marks if he had appeared in any of the latter shifts or other boards of the same examination. To eradicate this discrepancy the examination conducting Governing boards of India have introduced the concept of normalisation of marks in examinations.

Normalisation of marks means increasing and or decreasing the marks obtained by students in different timing sessions to a certain number. By that as it may, students who have scored 30 marks in session 1 because of a hard level of test will get 60 marks. Similarly, students who have scored 90 marks will also be marked as 60 marks following the normalisation process as followed by CBSE. The normalised average could also vary if the minimum or maximum marks are increased or decreased. The process and concept of normalisation of marks would also vary depending on the examination board of India.

Normalisation as used in the Indian context is a process for ensuring that students are neither advantaged or disadvantaged by the difficulty of examinations that they do for the Boards. This process is used in other countries with similar issues as in India. Normalisation Process requires us to know the following Percentile Score of a Candidate in a Board or JEE (Main) will reflect what Percentage of Candidates have scored below that Candidate in that Board or JEE (Main) Examination. A Percentile score is the value of below which a certain Percent of observations fall. For example, the 20th Percentile is the value (or score) below which 20 Percent of the observations may be found. The Percentile of a Candidate will be calculated as follows $100 \times$ no of candidates in the group with aggregate marks less than the candidate/total no of candidates in the group Most people have different traditions about what grade constitutes an A, B etc. Working to this scale requires an instructor to make many more subjective judgments; every assignment, quiz, or other graded item must be designed beforehand to try to yield the desired level of numeric performance. This is an extremely difficult task, which is why so many instructors wind up applying arcane and often arbitrary "curves" afterwards. Instead one can normalize all scores so that, no matter how easy or hard the assignment, the class's scores get mapped into a comparable range. This is essentially the same technique that is used on the SATs, ACTs, GREs and other national standardized tests.

There are different techniques for normalizing scores, and the topic of how to do so properly belongs in a class on statistics. The best-known normalization formula used for examinations and other situations where the number of scores above and below the average is likely to be equal, is the "z-score":

Where x is the student's score, the class average, and sd is the class standard deviation (a measure of how widely spread the class scores have been).

The score normalization techniques aim, generally, to reduce the scores variabilities in order to facilitate the estimation of a unique speaker-independent threshold during the decision step. In terms of actual score, the impact of normalisation would lead to decrease in the marks of candidates who have scored higher in the examination. However, such candidates should not worry about their scores because they would have cleared their examination. They should just follow the next steps in their selection process. Similarly, students who think they might have scored less, but would have received higher marks should not be amazed. They should understand that their actual result varied because of normalisation of marks and following their next steps to the selection process.

Major effects due to Normalisation of marks: It is evident that normalisation of marks would have a drastic effect in government examinations. Here are some of major effects caused due to normalisation of marks:

- Students who think they have scored the minimum marks in their shift might receive higher scores.

- Students who think they might have scored high marks would actually get lower normalised marks.

- Selection of candidates would become a much faster and easier process for all governing boards of India.

- Students who have appeared in examinations with difficult and easier levels would now be considered on a single platform.

- There will be no disregard of marks based on different shifts.

Conclusively, it can be said that normalisation of marks would have a drastic impact on all the upcoming government examinations.

6.4 Online Assessment and Percentile Ranking

Online assessment: Assessment can be focused on the individual learner or all individuals together, like the whole class, an institution or specific programme. Formative assessment will give us an overview of students in the very beginning of the instruction. It gives us the opportunity to still have the chance to improve our instruction. While, summative will give us the outcome of the whole instruction. Online assessments are very popular in the 21st century. One of the best blessings of globalization is the use of Information Communication Technology (ICT) in education. Starting from teaching learning to evaluation of the students' online mode is very effective. In very simple terms, an online assessment may be defined as an evaluation of a person's abilities, behaviours and or characteristics when a test is conducted over the Internet by using available web technologies. An online assessment may set out with clear objectives such as:

- To test the knowledge or learning of a student

- To select suitable candidates from a huge pool of applicants.

- To identify the strengths and weaknesses of the student

- To identify specific personality and character traits

- To provide clues in educational, vocational or career counseling and identify the suitable training, job or career for the test taker.

Need of Online Assessment: Online assessments are becoming quite popular and are being used quite extensively in various sectors including education, government and corporate companies. The educational institute has taken up online assessments with quite a lot of enthusiasm and in a huge way specially. These online tests have some clear benefits over the traditional methods of assessment both for the institution and students.

Online assessment tools are scientifically designed to help in testing aptitude, cognitive ability, communication, and many other skills. These tools come with auto-evaluation, report generation, and even grading features that transform a slow assessment process into a fast-paced one.

- The tests also enable teachers to diagnose the weaknesses and strengths of the students. The test results can also be used by teachers and educational or vocational counselors to provide training required for upskilling or re skilling the ability.

● An online assessment provides the advantage of speed and accuracy when compared with a traditional assessment method. The robust online tools eliminate any chances of malpractice, streamline hiring, and guarantee a positive candidate experience.

Online assessment platforms: There are different types of online assessment. Online assessment is used in different working areas and all of them have their own goal to achieve. Educational assessment is the systematic process of documenting and using empirical data on the knowledge, skills, attitudes and beliefs. By taking the assessment, teachers try to improve student learning.

There are more than 20 different types of online assessment tests. A few of the most popular tests cover skill assessment, communication assessment, cognitive assessment, behavioral assessment, etc. The tests are conducted over the internet to measure skills and personality of the students. Compared to pen and paper methods of assessments, online or e-assessment methods are less labour-intensive and more secure. Online assessment platforms contain all the tools required to administer the examination, measure talent, assess domain-specific skills, create comprehensive test reports, and more. Most of the platforms come with a plug and play feature and can be used on multiple devices. The type of questions varies from one assessment to the other depending on the domain and the expertise required of the job role. The questions are presented as MCQs, MAQs, fill in the blanks, true or false, etc. Some tests offer a vast library of tests that feature questions created by subject-matter experts and psychometricians.

Use of Online assessment in education: The educational sector has been able to streamline their test preparation and examination processes with online assessments in following ways:

● **Entrance examinations:** Online assessment is very effective to conduct entrance examinations. Huge number of applicants can go through this test sitting at their home. It will save time, energy, and money both for the institution and the students. Based on the result of online screening selected candidates may appear for counseling or admission.

● **Semester examinations:** Semester examinations can also be conducted with the help of online assessment. It is convenient for both teachers and students. It saves a lot of time, energy. It is also eco friendly as it is almost paperless.

- **Guidance and counseling:** At an individual level, an online assessment can be used to assess a student and plan his/her career progression.

- **Campus recruitment:** Companies tend to conduct campus recruitment to hire fresh talent for entry-level positions. An online assessment can help them manage the huge numbers by acting as a first-level screening. At the end of this screening, the recruiters will have a manageable size of candidate pool with the right talent for them to select the job-fit candidates for open positions.

- **Walk-ins:** Similarly, when open positions are notified and a huge number of applicants apply, an online assessment can prove useful in quickly selecting the candidates with the right skill sets.

Advantages of online assessments: The use of online assessments saves institutions a lot of time and money. Often the assessments can be completed in less time, multiple candidates can complete the online assessment at the same time and there is no need for specialized (and expensive) personnel. Also, the test takers are able to take the assessment during class, or at home, using their own devices. They get to see their results and answers and get instant feedback about their chosen topic. That helps us, a researcher, recruiter, teacher or trainer, to learn more about our users and adapt to their needs, strengths and weaknesses.

Disadvantages of online assessments: Not much can be said about the disadvantages of online assessments, since the advantages outweigh them by far. But there might be some, for example, one needs to be computer literate (or able to use a computer well) in order to create and take an assessment. Technology is not always reliable; there might be connection or internet problems, energy breaks and other things like that. Also, there is a cost involved in online assessment software and affordability of data by students especially in countries like India.

Percentile Ranking: A percentile is a measure that tells us what percent of the total frequency scored at or below that measure. In other words, percentile is a point on a rank ordered scale, found by dividing a group of observations into parts in the order of magnitude from lowest to highest.

A percentile rank is defined as the proportion of scores in a distribution that a specific score is greater than or equal to. For example, if someone has received 90 on a mathematics test and this score was greater than or equal to the scores of 85 percent students taking the test then the percentile rank of the student would be 85

and he or she would be placed in the 85th percentile. Alternatively, percentile rank is sometimes defined as "the proportion of a distribution that a score is greater than". Same grade in two classes does not necessarily mean that one did equally well in both the classes. By the help of Percentile Ranking we can precisely measure the difference.

Features of Percentile Ranking: Percentile ranking has the following features:

- Percentile rank is a number between 0 and 100 indicating the percent of cases falling at or below that score.

- Percentile ranks are usually written to the nearest whole percent. For example, 70.5%=71%=71st percentile

- Scores are divided into 100 equally sized groups

- Scores are arranged in rank order from lowest to highest.

- There is no 0 percentile rank. The lowest score is the first percentile.

- There is no 100th percentile as well. The highest score is at the 99th percentile.

Methods used to find Percentile Ranking: Linear interpolation method is used to calculate the Percentile rank. The formula is as follows;

PR=[C+fp×(S-L/i)]×100/N

Where, C = sum of all scores below the classes in which the PR is present. Fp= Frequency in that class in which the percentile rank falls

S = The score whose PR has to be determined

L = exact lower limit of the class interval upon which PR lies

i = Class interval

N = total number of scores

Illustration: Let us calculate the Percentile rank of a student whose score is 55 in the following distribution

scores	90-99	80-89	70-79	60-69	50-59	40-49
f	2	10	15	16	3	4

Scores	Class limits	Frequency (f) Frequency (C)	Cumulative
90-99	89.5-99.5	2	50
80-89	79.5-89.5	10	48
70-79	69.5-79.5	15	38
60-69	59.5-69.5	16	23
50-59	49.5-59.5	3	7
40-49	39.5-49.5	4	4

Here,   C=4, fp= 3, S=55, L= 49.5, i=10, N=50

PR= [4+3×(55-49.5/10)]×100/50

= [4+3×0.55]×2

= 11.3

= 11

Therefore, the PR of the student is 11 whose score is 55.

6.5 Grading, Credit Based Cumulative Grade Point, Averaging and Transcription

Grading: Grading in education is the process of applying standardized measurements of varying levels of achievement in a course. 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). Grading is a means of reporting the result of evaluation. It is a better means than the traditional marking system which is very common in use to report results. We award individual students for their performance by giving marks. But marks do not indicate anything about the quality of the performance, rather it shows only quantity. Grades provide the quality of performance. Grading is done either using symbols (A, B, C, D, E) or using verbal description (very good, good, average, poor, failure). There are different scales which can be employed in the grading system.

Five point scale-A, B, C, D and E

Seven point scale-O, A, B, C, D, E and F

The five point scale provides wide variation within the grade points. Whereas, the seven point scale keeps the evaluation within reasonable limits of accuracy and 'spacing' within a grade point.

Objectives of grading: The grading system has the following objectives:

- To indicate the quality of performance of students in examination

- To ensure greater uniformity in the assessment of students' performance.

- To facilitate better comparability of students' performance in various subjects and among various universities.

- To facilitate the inter university migration of students

- To enable students to make the correct choice of subjects in accordance with their abilities.

- To remove the defects of marking system

Significance of grading: Grading system is meant to replace the traditional marking system. Grading system is superior due to following reasons:

- The traditional marking system employs 0-100 scale. Based on the performance of a student the examiner has to place the student on a suitable point in the 0-100. The logic behind placing a student securing 59 in second division or class and just 60 in first division or class is not justified. The grading system reduces 101 (0-100) point scale to a variable short range of five points or seven point scale.

- The traditional marking system suffers from lack of objectivity, reliability and validity. There lies 50 percent chance of error in marking. Grading helps in removing errors in marking.

- The lack of reliability clearly indicates that the cut off marks used for award for division or class and for determining pass or fail is completely arbitrary. Yet so much importance is given to the percentage of marks, class or division for the entire academic career of the student. Grading system also solves this problem.

- There is no uniformity in computing cut off points for declaring students pass or fail or placing them in divisions and classes. Different universities use different cut off points as minimum qualification for entrance examination.

This creates a lot of complications during migration from one university to others. On the contrary, grading system provides uniformity in terms of the scale adopted by them

- Marking system creates problems for inter subject comparison. For example a score of 60 is pretty good in literature, but the same score may not be considered very well in Mathematics. Grading system helps us to compare the performance of a student in different subjects.

Computation of CGPA: 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. 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 (CGPA) is a calculation of the average of all of a student's total earned points divided by the possible number of points. This grading system calculates for all of his or her complete education career. Grade point averages can be unweighted (where all classes with the same number of credits have equal influence on the GPA) or weighted (where some classes are given more influence than others).

Credit Based Cumulative Grade Point: Currently an important concern which is strongly mentioned in recent times by the University Grants Commission (UGC), the National Assessment and Accreditation Council (NAAC), is the need to develop a Choice-Based Credit System (CBCS) in tune with global trends and the adoption of a proper grading system for measuring performance of the learner. The major system engaged in Higher Education in the global scenario is operating a system of credits. The European Credit Transfer System (ECTS), the 'National Qualifications Framework' in Australia, the Pan-Canadian Protocol on the Transferability of University Credits, the Credit Accumulation and Transfer System (CATS) in the UK as well as the systems operating in the US, Japan, etc are already in a system of credit to measure the level of competency. In tune of the above scenario the Indian Universities have adopted a credit-based-grading- system for the purpose of assessment of the students, which will be acceptable to the global universities.

Concept of Credit System: The credit system is basically a system for simplifying the process of 'academic book keeping'. The definition of credits in higher education

systems may be based on different parameters, such as student workload, learning outcomes and contact hours. A student earns his diploma by accumulating a specified number of academic credits instead of attending for a specified number of years. Thus, a part time student may spread out his studies over several more years than normal, keeping beside his other activities. While a brilliant student may take a heavier load than normal and shorten his education by a semester or so. A university usually prescribes that a certain number of minimum credits must be obtained in the students' main department. The remaining credits may be distributed as unrestricted or may be restricted in accordance with any scheme designed by the university. Credit system has the effect of adding considerably to the flexibility allowed by the semester organization of the academic schedule. The semester credit hour is defined as a unit for expressing quantitatively the content of a course at the level of higher education.

The most widely accepted meaning of a credit is that a student earns one credit by attending one period of 50 minutes of lecture class instructions and doing 2 to 3 hours of home study, each week throughout one semester (15 to 18 weeks). Thus the terms credit and semester hours are often used interchangeably. The credit system allows greater flexibility in fitting course time to the particular subject taught. Most semester courses carry 2 to 4 credits each. Credits do not always require a particular number of class periods. Since one hour of theory class is expected to be supplemented by 2 to 3 hours of home study, it is considered equivalent to 3 hours of laboratory practical work. Similarly, 2 to 3 credits per week may be allowed for a thesis that involves no class attendance at all. A full time student making pace at a normal pace will complete 30 to 32 semester credit hours of coursework in an academic year of 9 months. To obtain a Bachelor's degree in a three year college, students must normally accumulate 90 to 94 semester hours. Another important feature of this system is greater reliance on home study.

The most important feature is perhaps the flexibility it permits to students intensively as well as extensively to cover a wide range of subjects, coping up with needs and abilities of the individual students. The selection of course is done with utmost care and in consultation with the adviser concerned. It is also ensured that nearly 2/3rds of course credits is from the subject in which the students intend to major (honours) at the end of the first 2 years of general education programme. The students may also allow offering some courses for which no credit may be given. Generally, a student is required to offer not less than 9 credit courses and not more than 16 credit courses in one semester.

Semester System and Choice Based Credit System: The Indian Higher Education Institutions have been moving from the conventional annual system to semester system which is much customized in nature. Currently many of the institutions have already introduced the choice based credit system. The semester system accelerates the teaching-learning process to enable vertical and horizontal mobility in learning. The credit based semester system provides flexibility in designing curriculum and assigning credits based on the course content and hours of teaching. The choice based credit system provides a 'cafeteria' type approach in which the students can take courses of their choice, learn at their own pace, undergo additional courses and acquire more than the required credits, and adopt an interdisciplinary approach to learning. In assessing the performance of the students in examinations, the usual approach is to award marks based on the examinations conducted at various stages (session, mid-term, end-semester etc.,) in a semester. Some of the higher education institutions convert these marks to letter grades based on absolute or relative grading system and award the grades. There is a marked variation across the colleges and universities in the number of grades, grade points, letter grades used, which creates difficulties in comparing students across the institutions. The UGC recommends the following system to be implemented in awarding the grades and CGPA under the credit based semester system.

Letter Grades and Grade Points: Two methods of grading; relative grading or absolute grading, have been in vogue for awarding grades in a course. The relative grading is based on the distribution (usually normal distribution) of marks obtained by all the students of the course and the grades are awarded based on a cut-off marks or percentile. Under the absolute grading, the marks are converted to grades based on pre-determined class intervals. To implement the following grading system, the colleges and universities can use any one of the above methods. The UGC recommends a 10-point grading system with the following letter grades as given below:

Letter Grade	Grade Points
O (Outstanding)	10
A+(Excellent)	9
A(Very Good)	8
B+(Good)	7
B(Above Average)	6
C(Average)	5

P (Pass)	4
F(Fail)	0
Ab (Absent)	0

(N.B.: A student obtaining Grade F shall be considered failed and will be required to reappear in the examination. The Universities can decide on the grade or percentage of marks required to pass in a course and also the CGPA required to qualify for a degree taking into consideration the recommendations of the statutory professional councils such as AICTE, MCI, BCI, NCTE etc.)

Computation of SGPA and CGPA: The UGC recommends the following procedure to compute the Semester Grade Point Average (SGPA) and Cumulative Grade Point Average (CGPA). The SGPA is the ratio of sum of the product of the number of credits with the grade points scored by a student in all the courses taken by him or her and the sum of the number of credits of all the courses undergone by a student, i.e

SGPA (Si) = ?(Ci x Gi) / ?Ci

where Ci is the number of credits of the ith course and Gi is the grade point scored by the student in the ith course.

The CGPA is also calculated in the same manner taking into account all the courses undergone by a student over all the semesters of a programme, i.e.

CGPA = ?(Ci x Si) / ? Ci

where Si is the SGPA of the ith semester and Ci is the total number of credits in that semester. The SGPA and CGPA are rounded off to 2 decimal points and reported in the transcripts.

Computation of SGPA and CGPA: Illustration for SGPA

Course	Credit	Grade letter	Grade point	Credit Point (Credit x Grade)
Course 1	3	A	8	3x8=24
Course 2	4	B+	7	7x4=28
Course 3	3	B	6	3x8=18
Course 4	3	O	10	3x10=30

Course 5	3	C	5	3x5=15
Course 6	4	B	4	4x4=16
	20			130

Thus, SGPA =139/20 =6.95

Illustration for CGPA

Semester 1	Semester 2	Semester 3	Semester 4	Semester 5	Semester 6
Credit:20	Credit: 22	Credit : 25	Credit:26	Credit: 26	Credit : 25
SGPA:6.9	SGPA:7.8	SGPA: 5.6	SGPA:6.0	SGPA:6.3	SGPA: 8.0

Thus, CGPA = 20 x 6.9 + 22 x 7.8 + 25 x 5.6 + 26 x 6.0 + 26 x 6.3 + 25 x 8.0/ 144 = 6.73

Advantages of the Credit-Based-Grading-System:

- The flexibility it permits students to cover a wide range of subjects is perhaps the most important advantage of the credit system in education. Credits do not always require any particular class period. Moreover, it offers more flexibility to the students allowing them to choose inter-disciplinary courses along with major/ Honours courses, which makes education more broad-based.

- The course can be adjusted keeping in mind the needs and abilities of the individual student.

- The selection of courses is done with utmost care and in consultation with the advisor who ensures that preparatory courses are taken in the first or second semester.

- It is beneficial for the part time students who can spread out his studies over several more years than normal according to his convenience. Especially brilliant students also may shorten his education by a semester based on his need and ability.

- It respects 'Student Autonomy' and individual differences. It represents a shift in focus from teaching based to learning education since the workload is based on the investment of time in learning.

- It records student's workload realistically. It calculates not only the time spent

by the students in lectures or seminars but also the time they need for individual learning and the preparation of examinations etc.

● Students may undertake as many credits as they can cope with without repeating all the courses (papers) in a given semester if they are unsuccessful in one or more courses (papers).

● It also facilitates students' mobility. Providing opportunity to transfer the credit earned at one institution to another. Provide more transparency and compatibility between different educational structures.

● Finally, an important advantage associated with the credit system is ample opportunity for home study. This system encourages self-learning.

Averaging: An average test score is the sum of all the scores on an assessment divided by the number of test-takers. For example, if three students took a test and received scores of 69, 87, and 92, these numbers would be added together and then divided by three to get an average of 82.6. Public schools rely on average, below average, or above-average test scores to identify how well a group of students learns. In some cases, students may be compared to their peers in the class, schools, district, county or state. With the rollout of the Common Core State Standards, which established a common set of academic guidelines for states across the country, students may more frequently be compared to their peers nationally. Other times, school officials distinguish average students from others to see who is on grade level or how well an individual child performs in school compared to classmates on nationally normed tests. Averaging is very important in special education, particularly useful in standardized evaluations and in tests designed by teachers. Educators determine the average by adding a set of numbers and dividing the sum by the total number of numerals used in calculating that sum, also known as the mean. Teachers and specialists can use averages to determine the "middle" group of test-takers.

Statistically, on a normal curve about 68 percent of any large group of students will score within the low average to high average range on most tests. The other 32 percent will be in either the above average or below average group. Teachers may use averages to monitor the rate at which the class is learning the material. Teachers also use averages to estimate where an individual student's scores place them in relation to the rest of the class. Educators may also use averages to measure how an individual students' abilities rate on tests used to diagnose learning disabilities.

Sometimes educators and analysts use other methods to identify an average score. Rather than the mean, they may refer to the median, or the 50th percentile, which represents the score in the exact middle of the list of numbers. If anyone has scored in the 60th percentile, then he or she has performed better than 60 percent of test-takers.

Very often mean and median are used interchangeably with the term "average," but mean and median can be very different numbers, so care must be taken when calculating. For example, six students scored 72, 75, 78, 82, 84, and 92 on a test. To calculate the average, the test scores are added together and divided the sum (483) by six. The average score would be 80.5. If we are trying to find the median of the same set of numbers, we would identify the exact middle score. Since there is an even amount of numbers and thus no exact middle, we average the two middle scores (78 and 82) to arrive at a median of 80. In this case, the mean and median are very close, but that is not always the case.

Transcription: In education, a transcript is a certified record (inventory) of a student throughout a course of study having full enrolment history at educational institute including all courses (or subjects) attempted, grades earned and awards conferred. A transcript is sometimes referred to as a marksheet, marklist or report card. In the United States, a transcript is also called Cumulative Record File (CRF), permanent record, or simply record. In the European ECTS system, transcripts are called Transcript of Records (ToR). Therefore, a transcript can be defined as documentation of a student's permanent academic record, which usually means all courses taken, all grades received, all honours received, and degrees conferred to a student. Transcripts can be official and unofficial. Official transcripts often bear a security watermark in the paper on which the transcript is printed and bear the signature of the counselor, principal, or registrar who certifies the document. When employers require an official transcript, it must be requested by the student or alumnus and sent directly from the high school or college or university registrar's office to the Human Resources Office. If an unofficial transcript is requested, it is acceptable to send a photocopy, faxed copy, or a downloaded copy of a transcript as long as the copy is legible and contains the required information.

Transcript evaluation: Since transcript evaluation is an assessment of academic performance, preparing this type of an evaluation requires an understanding of the message conveyed by grades. This information is required regardless of grading scale

type, whether letter, number, percentage, description, or any combination thereof. Moving forward, it becomes necessary to identify any perceivable performance categories and matching them with the categories contained in the US A-F scale accordingly. Stratification or categorization of different levels of performance is frequently included on the credentials themselves, while many resources on international education systems contain this information as well as recommendations for conversion to the US grading scale. The final step in the basic principles of credential evaluation is converting a grade point average (GPA). Transferring post-secondary credit from international academic credentials requires the same type of detailed credential evaluation as for an initial admission decision. To be effective, it requires the accurate conversion and quantification of foreign academic coursework to the indigenous system used by the institution. At the post-secondary level, transfer credit is typically granted according to institutional policy for coursework completed at the same level at domestic institutions. The challenge, therefore, becomes consistency across campus in converting foreign credential outcomes to credits and making transfer recommendations. There is no universal recommendation on transfer credit for such programmes.

In summary, transferring credit is a complex process. It begins with consistently thorough and accurate credential evaluation, but affects many stakeholders beyond the admissions office including registrars, academic faculty, study-abroad offices, prospective applicants, and current students themselves. As such, one of the most important aspects in the process becomes consistency: consistency in the application of evaluation policies on campus, consistency with transferring credit across university offices, and consistency with peer institutions.

6.6 Summary

The traditional examination system is being replaced by modern evaluation processes. Grading is one of the important methods of replacing the traditional marking system. Grades describe the level of achievement. It addresses the limitations on the marking system. Marks denote a qualification of a performance, while grades denote qualitative banding of clustering of performance. The grading can be done by using five points or seven point scales. In the recent past grading system is normalized to reduce the scores variability in almost all major public examinations in India.

A pen-and-paper assessment requires extensive human effort, time and money in order to maintain the reliability, integrity and accuracy of the assessment. An online

assessment once properly designed can quickly deliver accurate and reliable results with minimum human intervention. Whereas, an online assessment acts as a first-level screening to select the right talent from the pool of applicants. It delivers quick, reliable and credible results, making the whole admission process absolutely convenient for the institution. Therefore an online assessment helps candidates perform their best and saves them from unnecessary travel and expense. Online assessments have some pros and cons. Although some financial investment in online assessment software is needed, the benefits outweigh the costs by far. It is a value for money acquisition and inevitable in the post pandemic era. Online assessments also allow the test takers to take on mobile learning, which means, they are able to take the assessments anywhere, anytime. And they receive their results immediately by email or Google form. They can see their results and download them into an Excel file. A percentile rank is defined as the proportion of scores in a distribution that a specific score is greater than or equal to. Same grade in two classes does not necessarily mean that one did equally well in both the classes. Percentile Ranking can help us in precisely measuring the difference.

Grading is an abstract concept which we attempt to quantify. This is done by isolating the few absolute criteria in any type of assessment system - the highest level of achievement and the lowest level of achievement required to 'pass'. A GPA is the mathematical average of all grades received, with consideration for the weighting of each course. Unlike overall grades used in foreign systems, a GPA is purely a mathematical average and does not place any arbitrary weighting on specific subjects or terms of study beyond the weighting indicated by the credit values. This is the standard by which domestic students' academic performance is compared, necessitating a similar calculation for a consistent assessment of international applicants. Semester credit system is another important method of the modern evaluation process. A credit system is a systematic way of describing an educational programme by attaching credits to its components. The semester credit hour is defined as a unit for expressing quantitatively the content of a course at the level of higher education. Averaging is a process of dividing the sum total of scores by the number of test takers. Central tendency measures help us in calculating average. Averaging helps teachers to identify the gifted and problem children in the class. A transcript is basically a marksheet, marklist or report card of a student that contains the full history at educational institution throughout the course. Based on the recommendations on Letter grades, grade points and SGPA and CGPA, the higher education institutions may issue the

transcript for each semester and a consolidated transcript indicating the performance in all semesters. Transcripts are important for transferring credit universally.

6.7 Self Assessment Questions

1. What are the effects of normalization in scoring?

2. Mention the advantages of online assessment.

3. What is the credit system in educational evaluation?

4. What is the need of Percentile Ranking in educational evaluation?

5. State the difference between absolute and relative grading systems?

6. What is letter grading?

7. Elaborate the computation of SGPA and CGPA

8. Discuss the significance of Grading.

9. Mention the uses of online assessment.

10. What is transcript evaluation?

6.8 References Books

- Dash, M. (2004): Education in India: Problems and Perspectives, Atlantic Publishers and Distributors, New Delhi, India, ISBN 81-7156-881-5, page no 204-208

- Sarkar, Amal kanti (2010): Measurement and Evaluation in Education, Kolkata, India, Rita Book House, page no 158-162, 217-219

- Lovely Professional University (2012): Educational Management and Evaluation DEDU504, USI Publications, New Delhi, page no. 209-234

Webliography:

- Abdul Gafoor, K and Jisha, P (2014) Study of Reliability of Marking and Absolute Grading in Secondary schools, Guru Journal of Behavioral and Social Sciences, Volume 2, Issue 2 (2014) ISSN: 2320-9038, Page292   ,

- https://www.researchgate.net/publication/273141548_A_Study_ of_Reliability_of_ Marki ng_and_Absolute_Grading_In_Secondary_Schools, retrieved on 16.07.2020

- https://onlinetyari.com/blog/effects-normalisation-marks-government-exams/ retrieved on 17.07.2020

- https://vikaspedia.in/education/policies-and-schemes/choice-based-credit-system-cbcs#:~:text=Cumulative%20Grade%20Point%20Average%20(CGPA)%3A%20It%20is%20a%20measure,courses%20in%20all%20the%20semesters., retrieved on 18.07.2020

- https://www.cs.odu.edu/~zeil/cs333/f13/Public/grading/grading-htmlse1.html visited on 19.07.2020

- https://www.merittrac.com/guides/what-is-online-assessment, retrieved on 05.07.2020

- https://www.onlineassessmenttool.com/knowledge-center/assessment-knowledge-center/a dvantages-and-disadvantages-of-online-assessments/item12518 retrieved on 10.07.2020

- https://www.upjs.sk/en/university/education/credit-system/(visited on 17.11.2017)

- https://www.verywellfamily.com/what-makes-a-test-score-average-in-special-education-2 161864, retrieved on 20.07.2020

- https://www.presiuniv.ac.in/web/exam_assessment.php, retrieved on 12.07.2020

- https://www.scholaro.com/reports/Transcript-Evaluation, retrieved on 16.07.2020

- http://www.cbse.nic.in/Normalisation%20Formula%20v8_2013.pdf, retrieved on 15.07.2020

- http://www.lbwcc.edu/Content/Uploads/lbwcc.edu/files/Transcripts.pdf, retrieved on 15.07.2020

- https://en.wikipedia.org/wiki/Grading_(education) (visited on 17.11.2017)

- https://link.springer.com/referenceworkentry/10.1007%2F978-0-387-73003-5_767, visited on 15.07.2020

Notes

Notes

Notes