



ଦୂର ଓ ଅନୁଲାଇନ୍ ଶିକ୍ଷା କେନ୍ଦ୍ର, ଉତ୍କଳ ବିଶ୍ୱବିଦ୍ୟାଳୟ  
CENTRE FOR DISTANCE AND ONLINE EDUCATION  
UTKAL UNIVERSITY

# BACHELOR OF ARTS IN EDUCATION

SEMESTER – III

CORE - 5: EDUCATIONAL ASSESSMENT AND  
EVALUATION

CREDIT: 6

BLOCK:1,2,3,4

AUTHOR:  
PROF.RAMAKANTA MOHALIK



UTKAL UNIVERSITY,  
Accredited with Grade – A+ by NAAC  
VANIVIHAR, BHUBANESWAR, ODISHA-751004

**CENTRE FOR DISTANCE AND ONLINE EDUCATION (CDOE), UTKAL  
UNIVERSITY, VANIVIHAR, BHUBANESWAR**

Programme Name: **Bachelor of Arts in Education, Programme Code -010105**

Course Name: **EDUCATIONAL ASSESSMENT AND EVALUATION**

SEMESTER: - III

CREDIT – 4

BLOCK – 1 TO 4

UNIT NO- 1 to 20

---

**EXPERT COMMITTEE**

---

**Prof. S. P. Mishra**

Retd. Professor, Regional Institute of Education, NCERT, Bhubaneswar

**Prof. Krutibash Rath**

Retd. Professor, Regional Institute of Education, NCERT, Bhubaneswar

**Prof. Smita Mishra**

Retd. Professor, Former Principal, Radhanath Institute of Advanced Studies in Education,  
Cuttack

**Dr. Dhiren Kumar Mohapatra**

Retd. Associate Professor, B.J.B Autonomous College, Bhubaneswar

---

**COURSE WRITER**

---

**PROF. RAMAKANTA MOHALIK**

PROFESSOR, REGIONAL INSTITUTE OF EDUCATION, NCERT, NEW DELHI.

---

**COURSE STRUCTURE EDITORS**

---

**Dr. Diptansu Bhusan Pati**, Faculty, Department of Education, CDOE, Utkal University,  
Vanivihar, Bhubaneswar

**Ms. Anita Nath**, Faculty, Department of Education, CDOE, Utkal University, Vanivihar,  
Bhubaneswar

---

**MATERIAL PRODUCTION**

---

**BY CDOE, UTKAL UNIVERSITY**



**DDCE,**

**EDUCATION FOR ALL**

**CENTRE FOR DISTANCE AND ONLINE EDUCATION (CDOE),  
UTKAL UNIVERSITY, VANIVIHAR, BHUBANESWAR-751007**

### **From the Director's Desk**

The Centre for Distance and Online Education, originally established as the University Evening College way back in 1962 has travelled a long way in the last 62 years. 'EDUCATION FOR ALL' is our motto. Increasingly the Open and Distance Learning institutions are aspiring to provide education for anyone, anytime and anywhere. CDOE, Utkal University has been constantly striving to rise up to the challenges of Open Distance Learning system. Nearly ninety thousand students have passed through the portals of this great temple of learning. We may not have numerous great tales of outstanding academic achievements but we have great tales of success in life, of recovering lost opportunities, tremendous satisfaction in life, turning points in career and those who feel that without us they would not be where they are today. There are also flashes when our students figure in best ten in their honours subjects. In 2014 we have as many as fifteen students within top ten of honours merit list of Education, Sanskrit, English and Public Administration, Accounting and Management Honours. Our students must be free from despair and negative attitude. They must be enthusiastic, full of energy and confident of their future. To meet the needs of quality enhancement and to address the quality concerns of our stakeholders over the years, we are switching over to self-instructional material printed courseware. Now we have entered into public private partnership to bring out quality SIM pattern courseware. Leading publishers have come forward to share their expertise with us. A number of reputed authors have now prepared the course ware. Self-Instructional Material in printed book format continues to be the core learning material for distance learners. We are sure that students would go beyond the course ware provided by us. We are aware that most of you are working and have also family responsibility. Please remember that only a busy person has time for everything and a lazy person has none. We are sure you will be able to chalk out a well-planned programme to study the courseware. By choosing to pursue a course in distance mode, you have made a commitment for self-improvement and acquiring higher educational qualification. You should rise up to your commitment. Every student must go beyond the standard books and self-instructional course material. You should read number of books and use ICT learning resources like the internet, television and radio programmes etc. As only limited number of classes will be held, a student should come to the personal contact programme well prepared. The PCP should be used for clarification of doubt and counselling. This can only happen if you read the course material before PCP. You can always mail your feedback on the course ware to us. It is very important that you discuss the contents of the course materials with other fellow learners.

We wish you happy reading.

DIRECTOR

## CORE 5- EDUCATIONAL ASSESSMENT AND EVALUATION

### Brief Content

Block No	Block	Unit No	Unit
1	<b>ASSESSMENT AND EVALUATION IN EDUCATION</b>	1	Understanding the meaning and purpose of test, measurement, assessment and evaluation
		2	Scales of measurement- nominal, ordinal, interval and ratio
		3	Types of test- teacher made and standardized, Approaches to evaluation placement, formative, diagnostic and summative
		4.	Types of evaluation- norm referenced and criterion referenced
		5.	Concept and nature of continuous and compressive evaluation
Block No	Block	Unit No	Unit
2	<b>INSTRUCTIONAL LEARNING OBJECTIVES</b>	6.	Taxonomy of instructional learning objectives with special reference to cognitive domain
		7.	Criteria of selecting appropriate learning objectives
		8.	stating of general and specific instructional learning objectives
		9.	Relationship of evaluation procedure with learning objectives
		10.	Difference between objective based objective type test and objective based essay type test
Block No	Block	Unit No	Unit
3	<b>TOOLS AND TECHNIQUES OF ASSESSMENT AND CONSTRUCTION OF TEST</b>	11.	Steps of test construction: planning, preparing, trying out and evaluation
		12.	Principles of construction of objective type test items, matching, multiple choice, completion and true – false
		13.	Principles of construction of essay type test
		14.	Non- standardized tools: Observation schedule, interview schedule, , check list,
		15.	Non- standardized tools: portfolio and rubrics, rating scale
Block No	Block	Unit No	Unit
4	<b>Characteristics of a good Test</b>	16.	Validity-concept, types and methods of validation
		17.	Reliability- concept and methods of estimating reliability
		18.	Objectivity- concept and methods of estimating objectivity
		19.	Usability- concept
		20.	Usability- factors ensuring usability

## CORE 5- EDUCATIONAL ASSESSMENT AND EVALUATION

### Contents

BLOCKS/UNITS	Page No
<b>BLOCK 01: ASSESSMENT AND EVALUATION IN EDUCATION</b>  <b>Unit 01:</b> Understanding the meaning and purpose of test, measurement, assessment and evaluation <b>Unit 02:</b> Scales of measurement- nominal, ordinal, interval and ratio <b>Unit 03:</b> Types of test- teacher made and standardized , Approaches to evaluation- placement, formative, diagnostic and summative <b>Unit 04:</b> - Types of evaluation- norm referenced and criterion referenced <b>Unit 05:</b> - Concept and nature of continuous and compressive evaluation	<b>1-52</b>
<b>BLOCK 02: INSTRUCTIONAL LEARNING OBJECTIVES</b>  <b>Unit 06:</b> Taxonomy of instructional learning objectives with special reference to cognitive domain <b>Unit 07:</b> Criteria of selecting appropriate learning objectives <b>Unit 08:</b> stating of general and specific instructional learning objectives <b>Unit 09:</b> Relationship of evaluation procedure with learning objectives <b>Unit 10:</b> Difference between objective based objective type test and objective based essay type test	<b>53-87</b>
<b>BLOCK 03: TOOLS AND TECHNIQUES OF ASSESSMENT AND CONSTRUCTION OF TEST</b>  <b>Unit 11:</b> Steps of test construction: planning, preparing, trying out and evaluation <b>Unit 12:</b> Principles of construction of objective type test items, matching, multiple choice, completion and true – false <b>Unit 13:</b> Principles of construction of essay type test <b>Unit 14:</b> Non- standardized tools: Observation schedule, interview schedule, , check list, <b>Unit 15:</b> Non- standardized tools: portfolio and rubrics, rating scale	<b>88-164</b>
<b>BLOCK 04: CHARACTERISTICS OF A GOOD TEST</b>  <b>Unit 16:</b> Validity-concept, types and methods of validation <b>Unit 17:</b> Reliability- concept and methods of estimating reliability <b>Unit 18:</b> Objectivity- concept and methods of estimating objectivity <b>Unit 19:</b> Usability- concept <b>Unit 20:</b> Usability- factors ensuring usability	<b>165-201</b>

---

## **BLOCK-I**

### **ASSESSMENT AND EVALUATION IN EDUCATION**

---

**Unit 01:** Understanding the meaning and purpose of test, measurement, assessment and evaluation

**Unit 02:** Scales of measurement- nominal, ordinal, interval and ratio

**Unit 03:** Types of test- teacher made and standardized, Approaches to evaluation- placement, formative, diagnostic and summative

**Unit 04:** - Types of evaluation- norm referenced and criterion referenced

**Unit 05:** - Concept and nature of continuous and compressive evaluation

## UNIT-1

# ASSESSMENT AND EVALUATION IN EDUCATION

### STRUCTURE

1.1.Learning Objectives

1.2.Introduction

1.3. Concept, Scope and Need of Measurement and Assessment

1.3.1. Concept of Measurement

1.3.2. Scope of Measurement

1.3.3. Concept of Assessment

1.3.4. Concept of Evaluation

1.3.5. Need of measurement and assessment

1.3.6. Characteristics of Assessment

1.3.7. Scope of Assessment

1.3.8. Scope of Assessment based on roles in teaching learning process

1.3.9. Interrelationship between measurement and assessment

1.3.10. Norm-referenced and criterion referenced measurement

1.4. Summary/Key Points

1.5.Unit End Exercises

1.6.Further Reading

---

### 1.1.LEARNING OBJECTIVES

---

After reading the unit, the learners shall be able to;

- Define the term measurement and assessment.
- Differentiate between measurement and assessment.
- Explain different types of assessment and its uses.
- Distinguish between uses of norm-referenced and criterion-referenced tests.

---

### 1.2.INTRODUCTION

---

One of the significant aspects of formal education is measurement and evaluation which gives feedback to the entire system of education for its quality improvement. Hence, all the stakeholders of education starting from parents to education administrators are very much

concerned about the assessment practices followed in school and colleges. It is very much essential for teachers and educators to make reliable and valid assessments and communicate results to parents and students. To conduct assessment in a qualitative manner, educators are required to have knowledge of theory and practices of assessment and related concepts. In this unit, the basic concepts of assessment such as measurement, assessment, types of assessment, role of educational objectives etc. are discussed with real examples. It also contains the nature, principles and functions of different types of assessment relevant to formal education.

---

### **1.3.CONCEPT, SCOPE AND NEED OF MEASUREMENT AND ASSESSMENT**

---

#### **1.3.1. Concept of Measurement**

Measurement is the process of assigning numerical values or quantifying attributes, characteristics, quantities, or qualities of objects, events, phenomena, or individuals based on established standards, criteria, or units. Measurement involves comparing an unknown quantity with a known quantity. For example, when we measure the width of a block, we compare the unknown quantity which is the width of the block with a known one which is the measuring tape.

Measurement in the educational context refers to the process of quantifying or assigning numerical values to various attributes, characteristics, or aspects related to teaching, learning, and educational outcomes. For example, a student scored 60 in mathematics. Here we are quantifying the student's achievement in one subject. Effective measurement in education is crucial for understanding student progress, evaluating the effectiveness of educational programs, and making informed decisions to improve the learning environment.

Various experts have provided different definitions based on their perspectives and areas of expertise. Here are some notable definitions of measurement:

- "Measurement is the process of assigning numbers to objects or events in such a way as to represent quantities of attributes." - **Gronlund**
- Measurement is the description of data in terms of numbers. -**Guildford**
- "Measurement is the assignment of numerals to objects or events according to rules."- **Stevens**
- "Measurement is the quantification of attributes of an object or event, which can be used to compare with other objects or events." - **Nunnally**



- “Measurement is the process of assigning symbols to dimensions or phenomena in order to characterize the status of the phenomenon as precisely as possible”. – **Bradfield & Moredock**

To sum up we can say that educational measurement is a descriptive process that involves the assignment of numbers to express the degree to which a student possesses a certain trait or characteristics in numerical terms. Hence it is always quantitative in nature.

### **1.3.2. Scope of measurement**

Measurement is the process of quantifying the characteristics/ attributes of individual learners on the basis of certain rules. It is the first step in the assessment process of formal education. It provides quantitative evidence to the assessor/teacher for making qualitative judgment or taking some decisions. The scope of the measurement covers the entire education system. Broadly, the scope can be explained in three levels as follows.

- **Direct measurement-** Direct measurement in education refers to assessing student learning or performance using tools and methods that directly measure the intended learning outcomes. It involves the straightforward evaluation of what a student knows or can do. For example, direct measurement methods are usually not used in educational setting as most of traits of learners can't be measured directly. These measurement methods are used in physical sciences where standard tools are available and can be directly used to measure traits. For example; weight can be measured by using the standard of kilogram. But in educational context some of the physical characteristics of learners such as Height, weight etc. can be measured with direct measurement.
- **Indirect measurement-** Indirect measurement in education involves assessing students through methods that require interpretation or inference based on multiple pieces of evidence or criteria. For example, while assessing some psychological traits of students cannot be measured directly, rather they are measured indirectly by inferring from the observable behaviour.
- **Relative measurement-** Relative measurement in education involves comparison of the performance or achievement of a student to that of peers or a group of reference students. It assesses how the performance of a student relates to the performance of others. Grading on a curve, norm-referenced tests, and percentile rankings are

examples of relative measurements in education. These methods gauge a student's standing within a specific reference group.

The measurement can be direct, indirect and relative depending on qualities to be measured. But in education, most of the measurement is indirect and relative in nature as it deals with human qualitative which can not be observed directly.

***Self-check Exercise-1.1***

Write three examples of measurement from a school situation as a teacher.

-----  
-----  
-----  
-----  
-----

**1.3.3. Concept of Assessment**

Assessment refers to the process of collecting, analyzing, and interpreting the information about students' performance by using testing and non-testing devices. It is a tool that provides teachers with data to improve their teaching methods and guide and motivate students to actively participate in their learning. It offers crucial feedback to both teachers and pupils, giving them essential information on what the pupils are learning and how well teachers are achieving their teaching goals. The real potential of assessments lies in utilizing them to give feedback to students. Enhancing the standard of education in courses means not just assessing how proficiently students have comprehended course material at the end of the academic year, but also tracking their advancement throughout the year. Assessment should not only provide teachers with valuable information about pupils' learning, but also assist pupils diagnose their own learning. That is, assessment should assist pupils "become more effective, self-assessing, self-directed learners" (Angelo & Cross, 1993). In reality, assessment is a planned and systematic procedure of collecting and utilizing factual data to measure students' knowledge, abilities, attitudes, and convictions to make informed teaching decisions. It gives comprehensive, valid, and reliable information that facilitates decision-making about students' progress. Assessment can be done during as well as at the conclusion of the academic session in the form of formative assessment and summative assessment by using both conventional and unconventional tools.

**1.3.4. Concept of Evaluation**

Evaluation is an act or process of assigning value judgment to a student's performance on the basis of certain criteria. It means to find out the value of or to judge the worth of. It

involves the process of comparing with certain standards or with a group of students. It is always based on measurement or quantification of students' performance in subjects or courses. Evaluation is a comprehensive term that encompasses all the techniques used to determine the outcomes of a particular intervention or teaching. It involves the methodical assessment of the value or significance of an object. Evaluation also involves the systematic gathering and analysis of data to offer constructive feedback regarding the performance of students.

Evaluation adds value judgment to assessment. It often involves providing suggestions for constructive action. It is a qualitative measure of effectiveness, suitability, and goodness. Evaluation is the process of assessing whether a program's parts, processes, or outcomes meet its stated objectives or an established standard of excellence. Evaluation involves inferring from the data collected through multiple sources. It is defined as "the process of collecting, interpreting and synthesising information in order to make decisions" (Gage and Berliner, 1991). The concept of evaluation can be articulated as follows:

<b>Evaluation</b> =	Value	Quantitative measure	Qualitative description
	of	+	of + Judgement
	Student's achievement		student's achievement by

#### 1.3.5. Need of Measurement and Assessment

The main purpose of measurement and assessment is to improve the quality of teaching learning. According to Oguniyi (1984), educational assessment and evaluation is carried out from time to time in the school for the following purposes:

- To determine the effectiveness of the curriculum, textbooks, and other learning resources in terms of students' behavioral output.
- To make valid and reliable decisions about educational planning.
- To determine the value of time, energy, and resources dedicated to the teaching-learning process.
- To recognize pupils' growth or absence of growth in acquisition of desired knowledge, skills, perceptions, and values.
- To support educators in assessing the efficiency of their teaching methods and educational resources.

- To motivate students to enhance their learning by recognizing their advancements or areas where improvement is needed in a particular subject.
- To cultivate discipline and adopt systematic study habits among students.
- To furnish educational administrators with sufficient information regarding teachers' performance and the requirements of the school.
- To familiarize parents or guardians with their children's performances in various learning areas.
- To predict the general trend in the development of the teaching-learning process.
- To offer an impartial basis for deciding the promotion of pupils from one grade to another and the issuance of certificates.

Measurement and assessment are an essential component of the teaching-learning process. It determines the extent to which learning objectives are achieved by the learners and provides feedback to teachers about the effectiveness of teaching. Many instructional decisions such as admission, grouping, scholarship, promotion, remediation, certification, and curriculum renewal are based on the assessment and evaluation results.

**Self-check Exercise-1.2**

What are the differences between assessment and evaluation?

---

---

---

---

---

---

---

---

**1.3.6. Characteristics of Assessment**

Assessment in education has several key characteristics defining its nature and purpose within the educational context. Understanding these characteristics is essential for making informed decisions about teaching and learning. Here are some of the prominent characteristics of assessment in education:

- **Purposeful:** Assessment in education serves specific purposes, such as evaluating student learning, informing instruction, providing feedback, and making decisions about curriculum and instructional improvement.
- **Multifaceted:** Educational assessment takes many forms, including quizzes, tests, projects, presentations, observations, and portfolios. These various assessment

methods allow educators to measure different aspects of student learning and performance.

- **Systematic:** Assessment is conducted systematically, often aligned with educational objectives and standards. It follows established procedures and criteria to ensure fairness, consistency, and reliability.
- **Ongoing:** Assessment is not limited to end-of-term exams but includes formative assessments conducted throughout the learning process. These ongoing assessments help teachers gauge student progress and adapt their teaching accordingly.
- **Diagnostic:** Assessments are used diagnostically to identify students' strengths and weaknesses in specific areas of knowledge or skills. This information guides instructional planning and support.
- **Objective:** Educational assessments aim to be as objective as possible, reducing subjectivity in grading and evaluation. Clear criteria and rubrics are often used to ensure consistency.
- **Valid:** Validity is a crucial characteristic of educational assessment. Valid assessments accurately measure the intended educational constructs or objectives, ensuring that the results are meaningful.
- **Reliable:** Reliability ensures that assessments produce consistent results when administered multiple times or by different assessors. Reliable assessments are trustworthy and dependable.
- **Feedback-Oriented:** Assessment provides feedback to both teachers and students. It helps teachers understand their students' needs and informs instructional decisions. Students receive feedback on their performance, allowing them to identify areas for improvement.
- **Continuous Improvement:** Assessment is a tool for continuous improvement in education. Data from assessments help identify areas in need of enhancement and inform strategies for educational improvement.

Understanding these characteristics helps educators and stakeholders use assessment effectively as a means of promoting learning, supporting educational goals, and ensuring the quality of education provided to students.

### 1.3.7. Scope of Assessment

Assessment is a very comprehensive and detailed process of gathering, analyzing, interpreting, and using test results for the improvement of learning. Its scope is very broad and all inclusive. It can start before the academic session, can continue throughout the teaching and end at the end of the academic session. Hence, scope of assessment covers the entire education system. Assessment in education can be categorized into different categories based on different criteria. Here are some common types of assessment based on different criteria in education:

### **Scope of Assessment on the Basis of Purpose**

#### **Placement Assessment**

Placement assessment, also referred to as initial assessment, serves as a crucial tool in education, especially at the commencement of a course or educational program. Its primary objective is to gauge a student's foundational knowledge, skills, and competencies within a particular subject area. This evaluation holds paramount significance as it ensures that students are positioned appropriately in courses that align with their current abilities, thereby preventing situations of either over-challenging or under-challenging students throughout their learning journey. Placement assessments provide educators with invaluable insights to make informed decisions about class placement, course selection, and instructional differentiation, all of which should be tailored to individual student readiness. The importance of this assessment lies in the creation of an optimal learning environment, where students receive instruction aligned with their current level of proficiency, reducing frustration and fostering a positive and productive learning experience. For instance, a typical placement assessment example includes the administration of a mathematics placement test to determine a student's readiness for advanced math courses, exemplifying its role in ensuring students are appropriately positioned for their educational endeavors.

#### **Formative assessment**

Formative assessment is a dynamic and interactive process that plays a pivotal role in education. It is an ongoing practice of collecting information during the learning journey, with the primary aim of providing valuable feedback and guiding instructional strategies. This assessment type holds the key purpose of not only monitoring student progress but also delivering timely feedback to both educators and students. By doing so, it enables real-time adjustments to teaching and learning strategies, ensuring that the educational journey is fine-tuned to the specific needs of each student. The significance of formative assessment lies in

its capacity to promote self-regulation of learning and empower students to recognize their strengths and areas that require improvement. Furthermore, it equips educators with the tools to make precise instructional adjustments, ultimately resulting in improved learning outcomes. In practice, formative assessment takes the form of in-class quizzes, classroom discussions, peer feedback, and teacher observations. For example, a teacher may conduct a brief quiz following a lesson to gauge student comprehension, thus allowing for tailored adjustments to the next day's lesson based on the outcomes, thus exemplifying the value of formative assessment in educational progress.

### **Diagnostic Assessment**

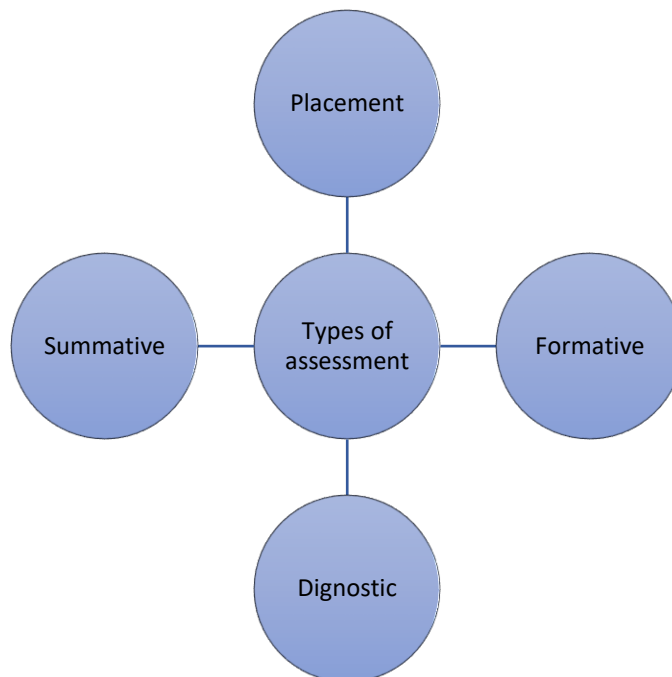
Diagnostic assessment stands as a comprehensive and indispensable element in the educational realm. Unlike placement and formative assessments, diagnostic assessments delve deeper into the students' specific strengths and weaknesses within particular subject areas or skills. The primary aim of this assessment type is to unearth individual learning needs, misconceptions, and areas that necessitate precise support or remediation. The wealth of insights provided by diagnostic assessments fuels the development of personalized learning plans tailored to the unique requirements of each student. These assessments are administered at various points in the learning process, delivering essential information for educators to offer highly precise instruction and targeted intervention. The importance of diagnostic assessment lies in its ability to ensure students receive tailored support to overcome their distinctive learning challenges. It empowers educators to provide the most effective instruction, thus optimizing student success. For example, when a teacher conducts pre-assessments at the start of a unit or course to determine what students know and need to learn are diagnostic in nature. In-depth interviews and surveys may also reveal students' prior knowledge and misconceptions.

### **Summative Assessment**

Summative assessment holds a pivotal role in the educational landscape, serving as a conclusive evaluation conducted at the culmination of a specific instructional period. Its primary objective is to offer a comprehensive overview of a student's overall performance in a particular subject or course. The purpose of summative assessment is multifaceted, aiming to make judgments about student progress, assign grades, assess the effectiveness of educational programs, and uphold accountability. These assessments are typically administered at the end of a course, term, or academic year, serving as significant

benchmarks for making decisions regarding student promotion or graduation. The importance of summative assessment is profound; it supports the grading process, aids in evaluating program effectiveness, and ensures educational accountability. By providing a holistic view of student achievement, summative assessments offer an indispensable tool for educators and institutions. Common examples of summative assessment include final exams, end-of-year standardized tests, term papers, and comprehensive projects. For instance, a final exam in a science course not only evaluates a student's overall knowledge of the subject but also contributes significantly to their final grade, exemplifying the critical role of summative assessment in education.

Educational assessment encompasses a range of functions and types, each contributing to the understanding and improvement of student learning. Placement assessment ensures students start their educational journey at an appropriate level, while formative assessment guides ongoing learning and adjustment. Diagnostic assessment uncovers individual learning needs, and summative assessment provides a comprehensive evaluation of overall achievement. The four types of assessment is presented graphically below.



***Self-check Exercise: 1.3***

What are the purposes of formative assessment?

---

---

---

---



-----
-----
-----
-----
-----

**Difference between Formative and Summative Assessment**

<b>Formative Assessment</b>	<b>Summative Assessment</b>
Formative assessment takes place throughout the teaching learning process.	Summative assessment takes place at the end of a specific time period.
Its primary purpose is to provide feedback, guidance for future teaching strategies to support the ongoing learning process.	Its primary purpose is to evaluate overall performance, assigning grades and assess the effectiveness of a programme.
The feedback provided during assessment allows for real time adjustment to the teaching learning strategies.	The feedback provided does not allow real time adjustments but helps in future planning and programme evaluation.
The frequency of such assessment is regular and frequent.	The frequency of summative assessment is less than formative assessment as it is done after the learning period is over.
Example-: In-class quiz, classroom discussion, peer feedback, teacher observations etc.	Example-: Final exams, term papers, comprehensive projects etc.
Teacher made test and informal tools are used for formative assessment	Teacher made test/school made test and test developed by the schools are used.

**1.3.8. Scope of Assessment Based on Roles in Teaching-learning Process**

Assessment of learning, assessment for learning, and assessment as learning are categorized based on their primary purposes and the roles they play in the teaching and learning process. These categorizations are primarily focused on the educational objectives and the roles of assessments in achieving those objectives:

- **Assessment of learning**

Assessment of learning is frequently referred to as summative assessment and it generally takes place at the end of the academic year. The main purpose is to promote or certify learners on the basis of assessment results. It gives an overall picture of students' performance at the end of the academic session. It also informs about the effectiveness of

curriculum, textbooks, and other resource materials so that initiatives can be taken by educational administrators for improvement. It provides the concluding marks for the students which are given by the teachers to the students. Through these parents know how qualitatively and quantitatively each student has scored in various tasks and activities of learning. It provides important information about student achievement; it often has little impact on learning. Assessment of Learning is performance-based and overall judgment, multi-tasking observation of a single student, providing certificate/ grade sheet, giving award or promotion from an existing class, shows final efforts of students in all subjects.

Both teacher-made tests as well as standardized tests can be used for assessment for learning. It uses both written and practical tests covering the curriculum of the entire academic year. The results of the student's performances are reported to different stakeholders such as teachers, school administrators, policymakers, and students. Assessment of learning has grave and extensive consequences and affects the future of students. That's why, it is of utmost importance that the tests are designed with care to measure the student's learning systematically and the results are far from undue influence and bias.

### **Assessment for learning**

The assessment done during the instruction is known as assessment for learning. It is an ongoing assessment process that allows teachers to monitor students on day-to-day basis tasks and modify their teaching needs to be successful. It provides students with specific feedback at a specific time that they need to make adjustments for their learning.

It is the process of seeking and interpreting evidence for use by learners and teachers to decide where learners are in their learning, where they need to go, and how best to get there (Assessment Reform Group, 2002). The aim of assessment for learning is to investigate what students have learned so far. Assessment for learning is very beneficial for teachers as it informs and impacts their decisions and instruction to decide their future teaching strategies for enhanced learning outcomes. It empowers students throughout the learning process and makes them the proprietors of their work. Students are encouraged and motivated for their attempts to complete challenging cognitive tasks. A healthy classroom environment is developed in which the instructor and students work together to achieve the learning objectives. Its primary goal is to improve teaching and learning through facilitating learners. It occurs concurrently with the teaching-learning process in the classroom. It is increasingly common and usually unstructured, frequently referred to as "formative assessment." Assessment for learning is school-based and essential to the teaching-learning process. It relies on diverse evidence, emphasizes a comprehensive assessment approach, is attuned to

individual learning requirements, monitors changes in learning progression over time, aids teachers in reviewing and adjusting the teaching-learning process, and contributes to addressing learning disparities.

### **Assessment as Learning**

Assessment as learning focuses on the advancement of learners' metacognitive skills. Writing is not a simple activity; rather, it requires a variety of cognitive and metacognitive abilities to produce a piece of writing. The goal of assessment as learning is to develop writers into critical assessors of their work integrating assessment and learning. Students learn how to organize their thoughts in a particular context.

Students are taught to make sense of current information and to combine it with past knowledge to create new information and new concepts. This assessment method emphasizes reflection and examination of one's own work. An essential purpose of assessment as learning is to nourish learning habits of mind in order to use crucial cognitive skills such as synthesis, analysis, restructuring, and so on. This instills confidence in learners, and they learn to guide their own learning and make decisions after reflecting on their work. However, instructors' responsibilities grow as they must now supply adequate examples of practice to pupils in molding their brains to develop the specific critical talents of mind. As a consequence, children become more adaptive, flexible, and self-sufficient learners.

When learners assess their performance on their own, they use a variety of assessment techniques and strategies that help learners to identify their knowledge gaps, adopt appropriate learning strategies, and use assessment as a tool for new learning.

In the realm of education, the integration of these three types of assessment ensures a well-rounded and effective approach to supporting students in their learning journeys. They not only serve to evaluate performance but also to drive improvements, encourage independent learning, and foster the development of critical cognitive skills. Ultimately, the convergence of these assessment approaches aims to enhance the overall quality of education and students' educational experiences.



### 1.3.9. Interrelationship between Measurement and Assessment

The interrelationship between assessment and measurement is fundamental in the field of education, where both concepts are closely intertwined to evaluate and enhance student learning and instructional effectiveness.

Assessment encompasses the broader process of collecting, analyzing, and interpreting information about students' knowledge, skills, and abilities. It involves determining what should be assessed, the methods used, and the purpose of assessment. Assessment focuses on making informed judgments about student performance, both in terms of formative assessment to improve ongoing learning and summative assessment to gauge overall achievement.

Measurement, on the other hand, is the systematic process of quantifying the attributes assessed in a reliable and valid manner. It involves using specific tools and instruments to assign numerical or qualitative values to the outcomes of assessment. Measurement provides the data necessary for educators to make objective and informed decisions about students' achievements. It helps ensure that assessments are standardized, consistent, and fair.

The relationship between assessment and measurement is reciprocal. Assessment guides the need for measurement by defining what aspects of student performance should be evaluated and why. It sets the educational objectives and learning outcomes that need to be measured. In this way, assessment serves as the driving force behind the measurement process, specifying what should be measured and what methods are appropriate for the task. Conversely, measurement informs the assessment process by providing the means to collect data and make sense of it. It transforms abstract educational objectives into concrete data that can be analyzed and interpreted. The data generated through measurement inform the

assessment process, allowing educators to make informed decisions about students' progress, identify areas for improvement, and adapt their teaching strategies accordingly.

In essence, assessment and measurement are inextricably linked in education. They work together to ensure that the evaluation of student learning is accurate, reliable, and aligned with educational objectives. This symbiotic relationship is essential for driving educational improvement, enhancing teaching and learning, and ultimately, achieving the goals of education systems around the world. In fact measurement is the starting point of assessment which provides quantitative information about students learning progress and performance. Assessment goes beyond measurement in and adds qualitative descriptions by gathering information from multiple sources like teachers rating, peer views, laboratory works etc. The difference between measurement and assessment is presented in the tabular form.

#### **Difference Between Measurement and Assessment**

<b>Measurement</b>	<b>Assessment</b>
It is the systematic process of quantifying attributes in a reliable and valid manner, using specific tools and instruments.	It is the broader process of collecting, analyzing, and interpreting information about students' knowledge, skills, and abilities.
It aims to assign numerical or qualitative values to the outcomes of assessment.	It aims to make informed judgments about student performance and progress.
Measurement is a smaller concept and part of the assessment process.	Assessment is the broader concept and uses the information collected through measurement.
Measurement focuses on quantitative data, typically numeric values.	Assessment includes both quantitative and qualitative data, such as scores, observations, and feedback.
It uses standardized tools and instruments, like tests and scales	It utilizes a variety of methods, including tests, assignments, observations, and self-assessments.
Measurement contributes to objective and data-driven decision-making, such as grading	Assessment supports understanding, improvement, and feedback, guiding educational decisions and strategies.

## **Norm-Referenced and Criterion-Referenced Measurement**

In the realm of education, measurement plays a pivotal role in gauging individual performance and determining the effectiveness of educational endeavors. Measurement can be classified as norm-referenced measurement and criterion-referenced measurement on the basis of the interpretation. These two approaches serve divergent purposes and shed light on varying facets of an individual's abilities and achievements in relation to the content and norms.

### **Norm-Referenced Measurement**

A norm-referenced measurement is an assessment approach that evaluates an individual's performance by comparing it to a larger group, known as the reference group. In this type of measurement performance of an individual is interpreted in terms or in reference with the norm. A norm is nothing but the typical or average performance of a group. So, through the lens of norm-referenced measurement, individuals are positioned relative to their peers, with scores commonly expressed in percentiles or standardized units. In this regard following definitions may be taken into consideration.

Gronlund (1976) stated that Norm-referenced tests are “designed to rank students in order of achievement, from high to low so that decisions based on relative achievements (selection, grading, grouping) can be made with greater confidence.” According to Bormuth it is designed “to measure the growth in a student’s attainment and to compare his level of attainment with the levels reached by other students and norm groups.

Example: Consider the UGC-NET examination which is a nationwide standardized assessment for master’s students. The test is administered to a large group of students across the nation. After the test, each student receives a percentile score, which indicates their ranking relative to the reference group master’s students who took the same test. For example, if Lisa got a percentile score of 90 then she performed better than 90% of the reference group.

### **Characteristics of Norm-Referenced Measurement**

**Comparative Assessment:** Norm-referenced assessment primarily focuses on comparing an individual's performance to that of a reference group. This group is usually a representative sample of individuals who have previously taken the same test or assessment.

### **Standardized Scores:**

Test scores in norm-referenced assessments are often transformed into standardized units such as percentiles, z-scores, or T-scores. These standardized scores indicate where the individual's performance falls in comparison to the reference group.

**Bell Curve Distribution:** In norm-referenced testing, scores tend to follow a bell curve distribution, also known as a normal distribution. This means that there will be a few high achievers, a majority of average performers, and a few low achievers.

**Ranking and Selection:** Norm-referenced measurement is frequently used in situations where ranking or selection is crucial. For example, it is commonly employed in standardized tests for college admissions, employment assessments, or competitive exams.

**No Fixed Passing Score:** Norm-referenced assessments typically do not have a fixed passing score. Instead, performance is evaluated relative to others in the reference group. The cutoff scores for different categories (e.g., "above average," "average," and "below average") are determined based on the distribution of scores in the reference group.

### **Advantages of Norm-Referenced Measurement**

Norm-referenced measurement provides an array of advantages such as:

- Norm-referenced measurement allows for the comparison of an individual's performance to that of a larger group, providing valuable insights into how they rank relative to their peers. This can be particularly useful in competitive contexts.
- It offers a standardized way to benchmark an individual's performance against national or regional averages, helping to identify strengths and weaknesses in educational systems.
- Norm-referenced tests can be helpful in situations where selection and differentiation are essential, such as college admissions, job recruitment, or awarding scholarships. They allow organizations to choose top performers from a pool of candidates.
- In some cases, norm-referenced assessments can serve as a quality control measure by comparing the performance of individuals or groups over time, identifying trends, and making improvements accordingly.

### **Limitations of Norm-Referenced Measurement**

Norm-referenced measurement offers valuable comparative information but has some limitations such as

- Norm-referenced measurement only provides information about how an individual's performance compares to others. It doesn't provide an absolute measure of their knowledge or skills. This can be limiting in understanding an individual's true proficiency.
- Norm-referenced tests often focus on a specific subset of content or skills, which may not align with the full range of educational objectives. This can lead to a narrow curriculum and neglect of important topics.
- High-stakes norm-referenced tests can create pressure and stress for individuals, especially in situations where their future opportunities are contingent on their performance. This can lead to test anxiety and may not accurately reflect their abilities.
- Depending on the performance of the reference group, norm-referenced assessments can result in grade inflation (if the reference group performs poorly) or grade deflation (if the reference group performs exceptionally well), which can distort the interpretation of scores.
- Norm-referenced tests typically do not provide diagnostic information about specific areas of strength or weakness, making it challenging for educators to tailor instruction to individual needs

While norm-referenced measurement offers several advantages, it is also important to recognize its limitations. Therefore, it should be used judiciously and in conjunction with other assessment methods to provide a more comprehensive view of an individual's abilities and skills.

### **Criterion-Referenced Measurement**

Criterion-referenced measurement is rooted in the assessment of absolute mastery. Here, individuals' performances are evaluated based on predetermined criteria or standards that reflect specific learning objectives or content proficiency. This approach, prevalent in classroom assessments and skill-based evaluations, answers the fundamental question of whether a learner has achieved the predefined goals. In this regard following definitions can be taken into consideration.



According to Glaser (1963), Criterion-referenced tests are designed as procedures which allow educators to monitor an individual's strengths and weaknesses in a given area, even though the tests may be administered in a group situation. According to Gronlund (1985), 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.”

Example: Let's consider a criterion-referenced test in the context of a high school biology class. The test is designed to assess whether students have mastered specific learning objectives related to cellular biology, such as understanding the process of mitosis and identifying the structures involved. The test questions are directly aligned with these objectives, and students are assessed on their ability to meet these criteria. A student who correctly answers all questions related to mitosis is considered to have met the criterion for that topic.

### **Characteristics of Criterion-Referenced Measurement**

- Criterion-referenced measurements are designed to measure whether a test-taker has met predetermined criteria or standards, often aligned with learning objectives or content standards.
- The primary focus is on determining whether an individual has mastered specific content or skills rather than comparing them to others.
- Results are often reported as "pass" or "fail" or different categories. These categories indicate the degree to which the criteria have been met.
- The assessment criteria are typically well-defined and objective, making it clear what a person needs to achieve to meet the standard.
- Results can help identify areas where students need additional support or enrichment.

### **Advantages of Criterion-Referenced Measurement**

- Criterion-referenced tests directly assess whether learning objectives have been achieved. This information is valuable for educators in gauging the effectiveness of instruction.
- Results from criterion-referenced tests provide clear and actionable feedback to students and educators. It highlights areas where students have succeeded and areas that require further attention.
- Criterion-referenced assessments are closely aligned with curriculum goals and content standards, ensuring content validity and relevance.

- Based on the results of criterion-referenced tests, educators can tailor instruction to address specific areas of weakness or provide enrichment for areas of strength.

### **Limitation of Criterion-Referenced Measurement**

Criterion-referenced tests have some disadvantages such as:

- Creating valid and reliable tests can be time-consuming and expensive.
- Moreover, generalizing results beyond the particular course or program is not possible.
- The integrity of these tests may be compromised if students gain access to exam questions beforehand.
- Criterion-referenced tests are tailored to a specific program, making them unsuitable for assessing the performance of large groups.
- Criterion-referenced assessments do not support comparisons among individuals or groups as their emphasis is on absolute mastery.

In summary, criterion-referenced tests are designed to determine whether individuals have met specific learning objectives or criteria. They provide clear and objective feedback on mastery but may not capture the complexity of real-world skills and can be challenging to develop with precision. These assessments are valuable tools for aligning instruction with educational goals and measuring content proficiency.

#### ***Self-check Exercise:1.4***

What are the uses of criterion-referenced measurement?

---

---

---

---

---

---

---

---

Our educational system uses both norms referenced and criteria referenced measurement as per the requirements. The classroom test used criteria referenced measurement whereas school board used norm referenced measurement. The differences between both are explained in the following table.

### **Difference between Criterion-Referenced Measurement and Norm-Referenced Measurement**

<b>Norm-Referenced Measurement</b>	<b>Criterion-Referenced Measurement</b>
Compare individuals' performance to a group or norm	Determine if individuals meet specific criteria or standards.
Focus is on relative performance (how one performs compared to others).	Focus is on absolute performance (whether predefined criteria are met).
No specific benchmark or standard.	Pre-defined benchmarks or standards are set
Often used for grading on normal probability curve	Typically used for pass/fail or mastery-based grading.
A high score may not represent mastery if the reference group performs poorly.	A high score indicates mastery, regardless of others' performance.
Typically designed to differentiate between test-takers	Designed to measure whether a person meets predetermined criteria.

---

#### **1.4.SUMMARY/KEY POINTS**

---

Assessment is an essential component of the teaching learning process. It is a process of collecting, analyzing, interpreting the students' performance to make instructional decisions. It starts with the measurement which is a process of quantifying students' performance in terms of certain rules. Measurement can be direct, indirect and relative depending on the characteristics to be measured. But in educational measurement, we use indirect and relative measurement as measuring qualities are not observable. Assessment is of many types as per the criteria. The assessment can be divided into four types such as placement, formative, diagnostic, and summative on the basis of purpose. Further, assessment can be assessment for learning, assessment of learning and assessment as learning. The scope of assessment is continuous and regular. Some key points are given below.

**Measurement** - Measurement is the process of assigning numerical values or quantifying attributes, characteristics, quantities, or qualities of objects, events, phenomena, or individuals based on established standards, criteria, or units.

**Assessment** - Assessment refers to the process of collecting, analyzing, and interpreting the information about students' performance by using testing and non-testing devices. It is a tool that provides teachers with data to improve their teaching methods and guide and motivate students to actively participate in their learning

**Evaluation** - is an act or process of assigning value judgment to a student's performance on the basis of certain criteria. It means to find out the value of or to judge the worth of. It involves the process of comparing with certain standards or with a group of students.

**Formative assessment** takes place throughout the teaching learning process.

**Summative assessment** takes place at the end of a specific time period.

**Assessment as learning** focuses on the advancement of learners' metacognitive skills.

**Assessment for Learning** - The assessment done during the instruction is known as assessment for learning. It is an ongoing assessment process that allows teachers to monitor students on day-to-day basis tasks and modify their teaching needs to be successful.

**Norm reference group** - Through the lens of norm-referenced measurement, individuals are positioned relative to their peers, with scores commonly expressed in percentiles or standardized units.

**Norm** - A norm is nothing but the typical or average performance of a group.

**Criterion-referenced measurement** is rooted in the assessment of absolute mastery. Here, individuals' performances are evaluated based on predetermined criteria or standards that reflect specific learning objectives or content proficiency.

---

### 1.5.UNIT END EXERCISE

---

- Defining meaning and scope and need of measurement. Explain
- Discuss about difference between assessment and measurement. Explain
- What are the scopes of assessment based on roles in teaching learning process?
- Differentiate between norm-referenced and criterion-referenced measurement. Discuss its relevance.

---

## 1.6.FURTHER READING

---

- Gronlund, N. E. (1965). *Measurement and evaluation in teaching*.  
<http://ci.nii.ac.jp/ncid/BA12623208>
- Goswami, M. (2013). *Measurement and Evaluation in Psychology and Education*
- Lee, W. Y. (2010). *Assessment and evaluation in higher education*.  
<http://ci.nii.ac.jp/ncid/BB11596810>
- Patel, R. N. (2014). *Educational evaluation theory and practice*.  
<http://ci.nii.ac.jp/ncid/BA29677030>

## UNIT -II

### FUNCTIONS AND BASIC PRINCIPLES OF ASSESSMENT

#### Structure

- 2.1. Learning Objectives
- 2.2. Introduction
- 2.3. Functions of Assessment
  - 2.3.1. Evaluation of learning
  - 2.3.2. Diagnosis of learning Needs
  - 2.3.3. Feedback for improvement
  - 2.3.4. Monitoring Progress
  - 2.3.5. Accountability and Quality Assurance
  - 2.3.6. Curriculum and Improvement
  - 2.3.7. Motivation and Goal Setting
  - 2.3.8. Placement and Differentiation
  - 2.3.9. Research and Education Policy
  - 2.3.10. Self-Assessment
- 2.4. Basic Principles of Assessment
  - 2.4.1. Validity
  - 2.4.2. Authenticity
  - 2.4.3. Fairness
  - 2.4.4. Flexibility
  - 2.4.5. Transparency
  - 2.4.6. Feedback and Improvement
  - 2.4.7. Ethical Consideration
- 2.5. Summary
- 2.6. Unit End Exercise
- 2.7. Further Reading

---

#### 2.1. LEARNING OBJECTIVES

---

After reading the unit, the learners shall be able to;

- Elaborate the functions of assessment.

- Explain basic principles of assessment.
- Define the implications of assessment

---

## **2.2. INTRODUCTION**

---

Assessment in education plays very critical role for improvement of education system in general. In particular it focuses on evaluation of learning, diagnosis of the learning needs, kinds of feedback received for improvement, monitoring progress, ensuring accountability and quality assurance, curriculum development, and motivation and goal setting. Overall improvement of education system depends on the proper assessment which ensures credibility. True assessment makes effort to develop every component of the system. It details all aspects given below.

---

## **2.3. FUNCTIONS OF ASSESSMENT**

---

Assessment in education is a multifaceted process that goes beyond merely assigning grades. It serves a wide range of functions, each contributing to the improvement of teaching and learning. This section will explore the diverse functions of assessment in education, highlighting their significance for students, educators, and the educational system as a whole.

### **2.3.1. Evaluation of Learning**

One of the fundamental functions of assessment is to evaluate what students have learned. It provides a structured and systematic way to measure their knowledge, skills, and understanding of specific content or concepts. Through various assessment methods such as tests, quizzes, and assignments, educators gauge students' comprehension and mastery of the subject matter. This evaluation helps determine whether instructional objectives have been met and to what extent.

### **2.3.2. Diagnosis of Learning Needs**

Assessment serves as a diagnostic tool to identify individual student learning needs. By analyzing assessment results, educators can pinpoint areas of strength and weakness for each student. This information goes beyond the mere quantification of performance; it reveals why a student may be struggling in a particular area or excelling in another. With this insight, educators can develop personalized instructional plans to address specific learning gaps and challenges effectively.

### **2.3.3. Feedback for Improvement**

Assessment provides timely and constructive feedback to both students and educators. Students receive feedback on their performance, allowing them to reflect on their strengths and areas requiring improvement. This feedback loop is essential for promoting metacognition and self-regulation in learners, as they become more aware of their learning processes. For educators, assessment data offer insights into the effectiveness of their instructional strategies. It allows them to make informed decisions about adjusting their teaching methods, curriculum, and instructional materials to better meet the needs of their students.

### **2.3.4. Monitoring Progress**

Assessment functions as a means to monitor student progress over time. Formative assessments, conducted throughout the learning process, allow educators to track how students are progressing toward learning goals. This ongoing monitoring serves several purposes, including identifying students who may need additional support, adapting instruction as needed, and providing students with a sense of their own growth and development.

### **2.3.5. Accountability and Quality Assurance**

Assessment is a tool for accountability at various levels of the education system. It ensures that educational institutions and educators are responsible for the quality of education they provide. High-stakes assessments, such as standardized tests, are often used for accountability purposes to assess school and system-wide performance. By holding stakeholders accountable for their roles in education, assessment contributes to maintaining and improving the overall quality of education.

### **2.3.6. Curriculum Development and Improvement**

Educational assessments help inform the development and improvement of curricula and instructional materials. By analyzing assessment results and identifying areas where students struggle, curriculum developers can refine educational content to enhance its effectiveness. This iterative process ensures that curricula align with educational goals and adapt to the evolving needs of students.

### **2.3.7. Motivation and Goal Setting**



Assessment can be a motivating force for students. When assessment is used to set clear learning goals and provide opportunities for students to track their progress toward those goals, it can inspire greater engagement and effort. As students see their efforts translate into improved assessment results, they are more likely to be motivated to continue their learning journey.

### **2.3.8. Placement and Differentiation**

Assessment plays a crucial role in determining appropriate placement and differentiation strategies for students. By assessing students' knowledge and skills, educators can place them in the right courses or programs, ensuring that they are appropriately challenged. Differentiation strategies, such as providing tailored instruction or additional support, can be based on assessment data to meet individual students' needs effectively.

### **2.3.9. Research and Educational Policy**

Assessment data are valuable for educational research and policy development. Researchers use assessment results to study educational trends, evaluate the effectiveness of interventions, and inform educational policies and practices. Government agencies and educational institutions rely on assessment data to make data-driven decisions about curriculum changes, resource allocation, and educational reform efforts.

### **2.3.10. Self-assessment**

Assessment encourages students to engage in self-assessment and develop metacognitive skills. Through reflection on their performance, students can better understand their learning processes, strengths, and areas for improvement. This heightened self-awareness promotes independence and self-directed learning, empowering students to take an active role in their education.

Assessment in education serves a multitude of functions, each contributing to the improvement of teaching and learning. Whether it's evaluating learning, diagnosing individual needs, providing feedback, monitoring progress, or informing policy and practice, assessment plays a central role in the educational process. When approached thoughtfully and strategically, assessment empowers both educators and students on their educational journeys, fostering growth, accountability, and the continuous improvement of educational systems.

**Functions of  
assessment**

- Evaluation of learning
- Diagnosing learning needs
- Feedback for improvement
- Monitoring progress
- Accountability and quality insurance
- Curriculum development and improvement
- Placement and differentiation
- Motivation and goal setting
- Self assessment
- Research and educational policy

***Self-check Exercise-1.5***

How assessment helps in monitoring learners' progress?

---

---

---

---

---

---

---

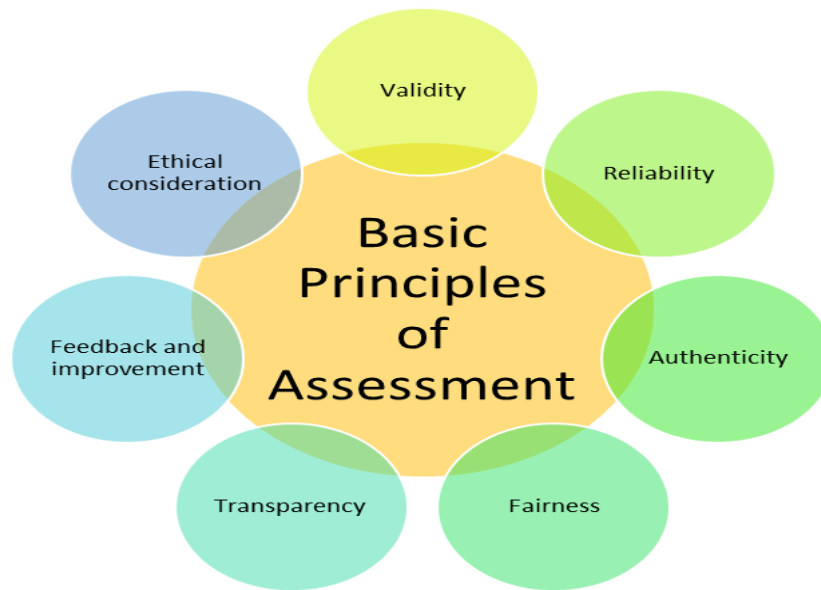
---

---

**2.4. BASIC PRINCIPLES OF ASSESSMENT**

---

Assessment in education is a complex process that involves collecting, interpreting, and using information to make informed decisions about teaching and learning. To ensure the effectiveness and fairness of assessments, a set of basic principles has been established. These principles serve as guiding standards that underpin the design, administration, and interpretation of assessments. Let's explore the basic principles of assessment in education, which serve as a foundation for designing, implementing, and interpreting assessments.



#### **2.4.1. Validity**

Validity is the cornerstone of assessment. It refers to the extent to which an assessment accurately measures what it is intended to measure. A valid assessment instrument aligns with specific learning objectives or outcomes and provides evidence that these objectives have been met. To establish validity, assessments should undergo thorough validation processes, including content validity, criterion-related validity, and construct validity.

#### **2.4.2. Reliability**

Reliability is the consistency and stability of assessment results. A reliable assessment produces consistent outcomes when administered to the same group of students or individuals under similar conditions. Reliability is essential to ensure that assessment results are not influenced by random factors, allowing educators to make dependable decisions based on the data.

#### **2.4.3. Authenticity**

Authenticity refers to the extent to which assessment tasks mirror real-world situations and require students to demonstrate meaningful, practical knowledge and skills. Authentic assessments are relevant and engaging, connecting classroom learning to real-life applications. They encourage critical thinking and problem-solving rather than rote memorization.

#### **2.4.4. Fairness**

Assessments must be fair and free from bias. Fairness ensures that all students have an equal opportunity to demonstrate their knowledge and skills, regardless of their background, characteristics, or circumstances. Educators should consider cultural, linguistic, and accessibility factors to create fair assessments that do not disadvantage any particular group of students.

#### **2.4.5. Flexibility**

The flexibility principle of assessment emphasizes the importance of adapting assessments to accommodate the diverse needs and circumstances of individuals. This principle acknowledges that a one-size-fits-all approach to assessment may not effectively capture the full range of students' abilities, experiences, and learning styles. Instead, flexibility in assessment design and administration allows for fair and meaningful evaluations while accommodating individual differences.

#### **2.4.6. Transparency:**

The transparency principle of assessment is a fundamental concept that underscores the importance of clarity, openness, and communication in the assessment process. This principle emphasizes that all stakeholders involved in the assessment, including students, educators, and administrators, should have a clear and comprehensive understanding of the assessment's purpose, expectations, and criteria. Transparency in assessment ensures fairness, motivates learners, and fosters trust in the educational system.

#### **2.4.7. Feedback and improvement**

Assessment should provide actionable feedback to students, educators, and stakeholders. Feedback helps students understand their strengths and areas for improvement, guiding their efforts. Educators use assessment data to adapt instruction, address individual student needs, and enhance teaching practices. Assessment results should inform continuous improvement in curriculum, instruction, and assessment design.

#### **2.4.8. Ethical Consideration**

The ethical consideration principle of assessment emphasizes the importance of conducting assessments with integrity, fairness, and respect for the rights and well-being of all individuals involved in the assessment process. Ethical assessment practices uphold the

values of honesty, transparency, equity, and privacy while ensuring that assessments serve their intended purposes without causing harm or discrimination.

The basic principles of assessment in education are essential guidelines that underpin the design, implementation, and interpretation of assessments. Adhering to these basic principles of assessment ensures that assessments are meaningful, reliable, and fair, promoting effective teaching and learning while upholding ethical standards.

***Self-check Exercise-1.6***

In which way validity is differentiated from reliability? Please elaborate

---

---

---

---

---

---

---

## 2.5. SUMMARY/KEY POINTS

The main function of assessment is to improve the quality of teaching learning, enhance the learning outcomes, suggest modifications of textbook and curriculum. Some of the key points are given below.

- **Formative assessments**, conducted throughout the learning process, allow educators to track how students are progressing toward learning goals.
- **Assessment serves** as a diagnostic tool to identify individual student learning needs. By analyzing assessment results, educators can pinpoint areas of strength and weakness for each student.
- **Assessment encourages** students to engage in self-assessment and develop metacognitive skills. Through reflection on their performance, students can better understand their learning processes, strengths, and areas for improvement.
- **Validity is the** cornerstone of assessment. It refers to the extent to which an assessment accurately measures what it is intended to measure.
- **Reliability** is the consistency and stability of assessment results. A reliable assessment produces consistent outcomes when administered to the same group of students or individuals under similar conditions.
- **Authenticity** refers to the extent to which assessment tasks mirror real-world situations and require students to demonstrate meaningful, practical knowledge and skills.
- The **ethical consideration principle** of assessment emphasizes the importance of conducting assessments with integrity, fairness, and respect for the rights and well-being of all individuals involved in the assessment process

## 2.6. UNIT END EXERCISE

- What are the important functions of assessment? Elaborate
- What are the basic principles of assessment?
- How assessment does help in teaching learning process? explain

## 2.7. FURTHER READING

- Gronlund, N. E. (1965). *Measurement and evaluation in teaching*.  
<http://ci.nii.ac.jp/ncid/BA12623208>
- Goswami, M. (2013). *Measurement and Evaluation in Psychology and Education*
- Lee, W. Y. (2010). *Assessment and evaluation in higher education*.  
<http://ci.nii.ac.jp/ncid/BB11596810>
- Patel, R. N. (2014). *Educational evaluation theory and practice*.  
<http://ci.nii.ac.jp/ncid/BA29677030>

## UNIT – III

### TAXONOMY OF EDUCATIONAL OBJECTIVES

#### STRUCTURE

- 3.1. Learning Objectives
- 3.2. Introductions
- 3.3. Concepts of Taxonomy of Educational Objectives
- 3.4. Domains of Taxonomy of Objectives
  - 3.4.1. Cognitive domain
  - 3.4.2. Affective Domain
  - 3.4.3. Psychomotor Learning
- 3.5. Summary
- 3.6. Unit End Exercise
- 3.7. Further Reading

---

#### 3.1. LEARNING OBJECTIVES

---

After reading the unit, the learners shall be able to;

- Define the taxonomies of educational objectives
- differentiate between cognitive and psychomotor domain of taxonomy of objectives
- explain different taxonomies of objectives and its uses.
- distinguish between uses of norm-referenced and criterion-referenced tests.

---

#### 3.2. INTRODUCTION

---

Educational objectives serve as the foundation of effective teaching and learning. They articulate what students are expected to know, understand, and be able to do as a result of their educational experiences. Educators often turn to taxonomies to provide a systematic framework for defining these objectives. Taxonomies of educational objectives categorize and organize learning outcomes into hierarchical structures, offering educators a valuable tool for curriculum design, assessment, and instructional planning.

---

#### 3.3. CONCEPT OF TAXONOMY OF EDUCATIONAL OBJECTIVES

---

A taxonomy of educational objectives is a classification system that categorizes learning outcomes based on their cognitive complexity, specificity, and depth. The term "taxonomy" originates from the Greek words "*taxis*," meaning arrangement, and "*nomos*," meaning law or

system. The purpose of a taxonomy is to provide a systematic way to describe and structure educational goals and objectives.

The historical background of the taxonomy of educational objectives is closely tied to the work of Benjamin S. Bloom and his colleagues, who developed what has become one of the most well-known taxonomies in the field of education. The taxonomy was initially conceived in the mid-20th century and has since become a foundational framework for educators and curriculum developers worldwide.

In 1956, a team of educators led by Benjamin S. Bloom published the book "Taxonomy of Educational Objectives: The Classification of Educational Goals." This book introduced what is commonly referred to as "Bloom's Taxonomy." The primary purpose of Bloom's Taxonomy was to provide a systematic framework for classifying and categorizing educational goals and objectives. This taxonomy classified educational objectives into three domains: cognitive, affective, and psychomotor. The cognitive domain, which received the most attention, was further divided into six levels of cognitive complexity: knowledge, comprehension, application, analysis, synthesis, and evaluation.

While Bloom's Taxonomy is widely recognized, other educators and researchers have developed taxonomies for specific domains or purposes. For instance, Krathwohl, Bloom, and Masia (1964) extended Bloom's work by addressing the affective domain, which deals with emotions, attitudes, and values. Additionally, Dave in 1970 and Simpson in 1972 proposed taxonomies for the psychomotor domain, which involves physical skills and coordination.

---

### **3.4. DOMAINS OF TAXONOMY OF OBJECTIVES**

---

Taxonomy of educational objectives is a powerful framework that categorizes and organizes learning goals into distinct domains. These domains, each focusing on different aspects of human learning and development, provide educators with a structured approach to defining, assessing, and facilitating student growth. The Taxonomy of Educational Objectives comprises three primary domains: the Cognitive Domain, the Affective Domain, and the Psychomotor Domain. Each of these domains offers a unique lens through which to view the diverse dimensions of education. Below is a detailed discussion of these taxonomies within their respective domains:

#### **3.4.1. Cognitive domain**



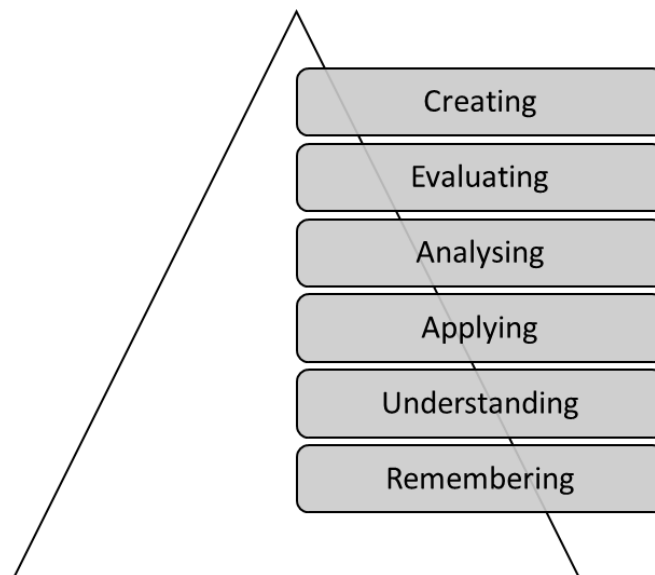
The Cognitive Domain focuses on intellectual skills and the mental processes involved in learning developed by Benjamin S. Bloom and his colleagues in 1956, the Taxonomy of Educational Objectives in the Cognitive Domain, often referred to as Bloom's Taxonomy, has become a cornerstone in education. It categorizes learning objectives into six hierarchical levels:

Bloom's Taxonomy is a widely recognized framework for categorizing educational objectives and learning outcomes, particularly in the cognitive domain. It was developed by a group of educators led by Benjamin S. Bloom in 1956 and has since become a foundational tool in education for setting clear learning objectives, designing curriculum, and assessing student progress. The taxonomy organizes cognitive skills into a hierarchical structure, with each level representing a different level of cognitive complexity. The original Bloom's Taxonomy consists of six levels:

- **Knowledge:** At the base of the pyramid, knowledge involves the recall of facts, information, or concepts. Learners are expected to remember and recognize information.
- **Comprehension:** This level requires students to understand the meaning and interpretation of information. They should be able to explain concepts or ideas in their own words.
- **Application:** Application involves using acquired knowledge and comprehension to solve problems, make predictions, or apply concepts to new situations. Learners apply what they have learned in practical ways.
- **Analysis:** Analytical thinking requires breaking down complex ideas or concepts into smaller components. Students examine relationships and implications, deconstructing information to gain deeper insights.
- **Synthesis:** Synthesis entails combining elements or ideas to form a new, integrated whole. It involves creative thinking and the generation of novel solutions or ideas.
- **Evaluation:** At the pinnacle of the taxonomy, evaluation requires students to assess the value, significance, or quality of ideas, concepts, or solutions. They make judgments and provide evidence to support their conclusions.

Anderson and Krathwohl revised the taxonomy of the cognitive domain in 2001, which is an adaptation of the original Bloom's Taxonomy. It offers a more contemporary and comprehensive framework for categorizing cognitive learning objectives and outcomes. The six cognitive categories in Anderson and Krathwohl's revised taxonomy are as follows:

- **Remembering-** At the lowest level, remembering involves recalling factual information, concepts, or previously learned material. It focuses on recognizing and retrieving knowledge.
- **Understanding-** Understanding goes beyond mere recall. It requires students to comprehend and interpret information by explaining it in their own words, summarizing, or paraphrasing.
- **Applying-** Applying involves using knowledge and comprehension to solve problems, make predictions, or apply concepts to new or unfamiliar situations. It requires practical application of what has been learned.
- **Analyzing-** Analyzing entails breaking down complex ideas or concepts into smaller components. Students examine relationships, identify patterns, and explore the structure of information.
- **Evaluating-** Evaluation requires students to assess the value, significance, or quality of ideas, concepts, or solutions. They make judgments and provide evidence to support their conclusions.
- **Creating-** Creating is the highest level in the revised taxonomy, emphasizing the generation of novel ideas, products, or solutions. It involves originality, creativity, and the ability to synthesize knowledge.

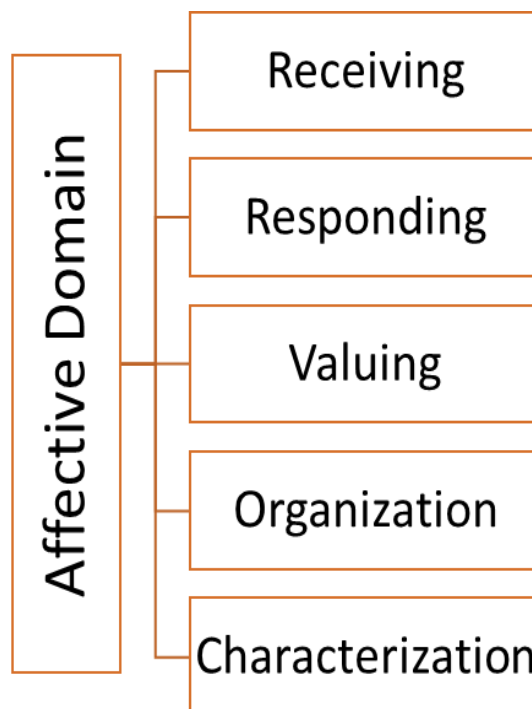


### 3.4.2. Affective Domain

The Affective Domain addresses emotions, attitudes, and values. It categorizes learning objectives related to feelings, beliefs, and behavioral intentions. The Taxonomy of

Educational Objectives in the Affective Domain, developed by Krathwohl, Bloom, and Masia in 1964, includes five hierarchical levels:

- **Receiving**-This means that the individual has become aware of certain stimuli in their surroundings. Teachers often create simulated scenarios to help students understand what they should focus on and what they should ignore.
- **Responding**- When receiving leads to selective responses to certain stimuli, individuals derive pleasure and remain actively engaged.
- **Valuing**- The third category in this domain is about valuing, which is the process of assessing the worth of a thing or activity. As a person internalizes values over time, they gradually build up a value system.
- **Organization**- It involves relating the new value to that one already holds and bringing it into a harmonious and internally consistent philosophy.
- **Characterization**- Finally, an ideal organization of the person's value systems results in his distinctive characterization. Characterization refers to the organization of value in an internally consistent system. It is the final step on the emotional ladder.



### 3.4.3. Psychomotor Domain

The Psychomotor Domain deals with physical skills, motor abilities, and the performance of physical tasks. It categorizes learning objectives related to physical actions and

behaviors. There are several popular taxonomies and frameworks for the psychomotor domain that have been developed to categorize and assess physical skills and abilities. However, we will discuss two notable taxonomies of the psychomotor domain in detail.

**R.H. Dave in 1970** developed the following taxonomy which included five levels:

- **Imitation:** At this level, learners can observe and replicate basic physical actions or movements performed by others. Imitation involves mimicking the actions without necessarily understanding the underlying principles.
- **Manipulation:** Manipulation represents a more advanced stage where learners can perform physical tasks with some degree of skill and coordination. They can manipulate objects or perform actions based on instruction or demonstration.
- **Precision:** Precision involves the ability to perform physical tasks with a high degree of accuracy and consistency. Learners can control their movements and actions with finesse and can make precise adjustments as needed.
- **Articulation:** Articulation signifies the ability to adapt and modify physical skills and actions to suit different situations or contexts. Learners can articulate and apply their skills creatively and flexibly.
- **Naturalization:** At the highest level, naturalization represents the integration of physical skills into one's natural or automatic behavior. Learners can perform complex tasks effortlessly, almost as second nature.

Another notable taxonomy of the psychomotor domain was developed by Simpson in 1972. This taxonomy has 7 hierarchy levels:

- **Perception:** At the lowest level, learners develop the ability to detect and recognize sensory stimuli related to physical skill. This involves using sensory cues to identify relevant information.
- **Set:** The second level involves preparing and getting ready to perform a physical skill. Learners establish a mindset or attitude that is conducive to the skill they are about to execute. This level focuses on mental and emotional readiness.
- **Guided Response:** At this level, learners begin to imitate or mimic a physical skill based on observation or instruction. They are guided by external cues, demonstrations, or step-by-step guidance from an instructor.
- **Mechanism:** Mechanism represents a higher level of skill development where learners can perform a physical task without external guidance or cues. This level emphasizes the coordination of movements and actions required for the skill.

- **Complex Overt Response:** This level involves executing complex physical skills with precision and accuracy. Learners can perform the skill effectively and adapt it to various situations
- **Adaptation-** Adaptation is the ability to modify the learned skills to meet new or special requirements. The skills are so well developed that one can modify movement patterns to fit special requirements.
- **Origination-** Origination is the ability to create new movements for unique situations or problems. It involves developing an original skill from a learned one and emphasizes creativity as a learning outcome.

***Self-check Exercise-1.7***

What are the different levels of mental operation needed in the cognitive domain of educational objectives?-----

-----  
-----  
-----  
-----  
-----  
-----

---

**3.5.SUMMARY/KEY POINTS**

---

The educational taxonomy has a great role to play in the assessment process. The educational objectives are divided in three domains such as cognitive, affective and psychomotor. The cognitive domain deals with mental operations such as remembering, understanding, applying, analysing, evaluating and creating. Similarly, the affective and psychomotor domain has many mental operations. The assessor must keep in mind the instructional objectives when planning and interpreting assessment. Some of the key points are given below.

- A **taxonomy of educational objectives** is a classification system that categorizes learning outcomes based on their cognitive complexity, specificity, and depth.
- The term "**taxonomy**" originates from the Greek words "*taxis*," meaning arrangement, and "*nomos*," meaning law or system. The **purpose of a taxonomy** is to provide a systematic way to describe and structure educational goals and objectives.
- In 1956, a team of educators led by Benjamin S. Bloom published the book "**Taxonomy of Educational Objectives: The Classification of Educational Goals.**" This book introduced what is commonly referred to as "**Bloom's Taxonomy.**"

- This taxonomy classified **educational objectives into three domains: cognitive, affective, and psychomotor**. The cognitive domain, which received the most attention, was further divided into six levels of cognitive complexity: **knowledge, comprehension, application, analysis, synthesis, and evaluation**.
- The Cognitive Domain focuses on intellectual skills and the mental processes involved in learning **developed by Benjamin S. Bloom** and his colleagues in 1956
- The **Affective Domain** addresses emotions, attitudes, and values. It categorizes learning objectives related to feelings, beliefs, and behavioral intentions.
- The **Psychomotor Domain** deals with physical skills, motor abilities, and the performance of physical tasks

---

### 3.6. UNIT END EXERCISE

---

1. Explain the taxonomy of educational objectives in the cognitive domain.
2. Explain the role of taxonomy of educational objectives in assessment.
3. Discuss the taxonomy of educational objectives in affective domain.
4. Explain the taxonomy of educational objectives in the psychomotor domain.

---

### 3.7. FURTHER READINGS

---

- Gronlund, N. E. (1965). *Measurement and evaluation in teaching*. <http://ci.nii.ac.jp/ncid/BA12623208>
- Goswami, M. (2013). *Measurement and Evaluation in Psychology and Education*
- Lee, W. Y. (2010). *Assessment and evaluation in higher education*. <http://ci.nii.ac.jp/ncid/BB11596810>
- Patel, R. N. (2014). *Educational evaluation theory and practice*. <http://ci.nii.ac.jp/ncid/BA29677030>

## **UNIT - 4: TYLER'S MODEL OF ASSESSMENT**

### **Structure**

- 4.1. Learning Objectives
- 4.2. Introductions
- 4.3. Concept of Tyler's Model of Assessment
- 4.4. Stages of Tyler's Model of Assessment
  - 4.4.1. Establishing broad goals or objectives
  - 4.4.2. Classifying objectives
  - 4.4.3. Stating objectives in behavioural terms
  - 4.4.4. Finding situations in which achievements of objectives can be observed
  - 4.4.5. Developing or selecting measurement techniques
  - 4.4.6. Collecting students' performance data
  - 4.4.7. Comparing data with behavioural stated objectives
  - 4.4.8. Uses and Limitations
- 4.5. Summary/Key Points
- 4.6. Unit End Exercise
- 4.7. Further Reading

---

### **4.1. LEARNING OBJECTIVES**

---

After reading the unit, students shall be able to;

- Explain stages of Tyler's Model of Educational Assessment.
- Elaborate the Tyler's Model of Educational Assessment, its uses and implication
- Define the practical applicability of Tylers Models of Educational Assessment in education

---

### **4.2. INTRODUCTIONS**

---

Assessment is the integral part of the teaching and learning process. It determines the effectiveness of teaching of teachers, leadership of school heads, quality of textbooks, assessment practices and teacher education programmes. The achievement of learning outcomes by learners is ascertained by using assessment data. Hence, all the stakeholders of education from school to university are very much interested about the modalities and practices of assessment followed. So, the assessment must provide valid and reliable data to the policy makers and stakeholders as well. It must contribute to the quality improvement of education. To make the assessment effective in helping the education process, it must follow

models of assessment advocated by many psychometricians and educationists. Model gives a framework and step by step process of conducting assessment so that it gives a valid and reliable result. Model makes assessment organised, systematic, usable, and objective. All the teachers and educators must understand the nitty gritty of different educational models so that they can conduct assessment. There are many models of assessment proposed by different authors. In this unit, the Tyler's Model of Educational Assessment are discussed in detail with its uses and limitations

---

### 4.3. CONCEPT OF TYLER'S MODEL OF ASSESSMENT

---

Model of educational assessment gives a systematic procedure for conducting assessment to practitioners. The practitioners may be a teacher, school heads, curriculum developers, textbook writers, policy makers and educational administrators. Let's discuss different models of assessment and its uses in educational settings.

Ralph W. Tyler (1902-1994) was an educator from America who contributed for the quality improvement of assessment and curriculum in his book *Basic Principles of Curriculum and Instruction*. He came up with the first recognised model for curriculum evaluation during 1933 and 1942. His model is a systematic, structured and linear approach to curriculum evaluation. It focuses on purposes that schools seek to attain during an academic year. The steps focus on educational experience that can be provided to attain the utmost achievement. It plans for educational experiences to be effectively organised. The model helps to determine whether the purposes of education have been attained or not by the learners. In his model of assessment, he has suggested eight steps which to be followed in assessment. His model of assessment is very much useful for curriculum developers who can ascertain if the objectives are achieved by the learner or not. The steps of assessment proposed by Tyler are described in the following pages.

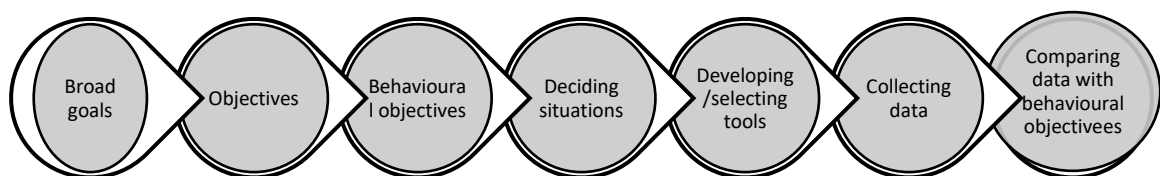


Figure-2.1: Steps of Tyler model of assessment

---

### 4.4. STAGES OF TYLER'S MODEL OF ASSESSMENT

---

#### 4.4.1. Establishing broad goals or objectives



The purpose of education is to train the human mind to achieve all round development in cognitive, affective and psychomotor domains through sense organs. For achieving optimum development and bringing out the best possibilities within learners, the educational planners and administrators need to come up with a curriculum that can be utilised as an instrument in the hand of the teacher. Curriculum provides the totality of experience to the learners inside the school campus. A curriculum is designed critically by considering different stages of growth and maturation. Tyler suggested establishing the broad goals as the first step of curriculum development and assessment. The first step in the assessment is to determine the goals or objectives of education.

#### **4.4.2. Classifying objectives**

The next step of Tyler's model is categorising the objective into dimension. Objectives may be directed towards developing different dimensions of development, such as communication skill, demonstration skill, activity skills. The objectives can be;

- To promote socio-cultural and moral values.
- Development of reasoning and logical thinking.
- Like skill learning
- Fostering creativity among students.
- To bring out positive behavioural changes among students.
- To develop scientific attitude and citizenship among students.

This step holds good, since specific objectives require individualised learning instructions and methods of teaching. In these steps, the different dimensions of learning to be decided. The dimensions can be cognitive, affective and psychomotor as the goal of education is all round development of learners.

#### **4.4.3. Stating objectives in behavioural terms**

Assessment requires the learning objectives to be stated in behavioural terms so that it can be observed and measured. In this step, objectives of learning in different subjects must be stated in behavioural terms or in action verbs and the changes that educator want to bring among the learners. The behavioural changes may result in with some action verbs like defining, describe, least record, repeat, discuss, explain, translate, analyse, calculate, compare, criticize, differentiate, distinguish, estimate, measure, value, score, contrast. The example of objectives in action verb is learners will be able to explain the photosynthesis process. Similarly, the assessor must state all the objectives in behaviour terms for assessment. Because, objectives are like target to be achieved by the learners. Without deciding target, it is difficult to say whether it is met or not. By end of this stage, educator is

clear what to assess in learners in terms of behavioural objectives relating to different subjects.

#### **4.4.4. Finding situations in which achievements of objectives can be observed**

The objectives that were targeted, can be reflected in different situations. Only then we can assume that the objectives have been achieved. As the skills and competencies achieved through different situations of learning have specific ways of expression. For instance, a vocational learner may be expected to perform the skills and competencies he/she has achieved through different learning experiences. The goal of educational planners is to find such situations where these skills can be displayed. The situation can be in the classroom, laboratory, playground, library etc. It can be in terms of oral, written, performance, demonstration, group discussion, conversation etc. The teacher must decide the situations in which these objectives or actions can be expressed by the learners.

#### **4.4.5. Developing or selecting measurement techniques**

The present processes of education rely highly on evaluation and measurement and it is considered as an integral part of the teaching learning process. Tyler then suggested this step looking at the importance it holds for rectifying or improving learning instructions. Different methods of measuring learning outcomes are interview, written examinations, tests, inventories, observations, checklists, rating scales, attitude scale, projective techniques. These techniques are useful for measuring the student's learning progress. In this step, teachers either develop measuring tools and techniques on their own or can select readymade tools/standardised tools available. Here care must be taken to develop or select relevant tools suitable for measuring learning outcomes in different situations and subjects.

#### **4.4.6. Collecting students' performance data**

Data is very important for assessing learning outcomes. It is information relating to performance of students in different subjects. Data on students' performances can be collected through different tools and record keeping instruments. Some of those instruments are tests, assignments, portfolio, anecdotal records, progress report cards etc. Teachers must be very careful in collecting genuine and valid data on students' performance. These data/records can be stored, retrieved and analysed to identify the best works, interests, abilities and can be utilised as an evaluation technique to improvise learning instructions and suggest career choices to students. A comprehensive information about all students must be collected in this stage so that proper decisions can be taken.

#### **4.4.7. Comparing data with behavioural stated objectives**

The last step in the assessment process is to compare the performance data with the behavioural objectives decided earlier. The data collected can be compared with the extent to which the objectives for behavioural changes have been achieved. The behavioural changes are some of the skills and competencies such as effective communication, mathematical operations, socio- emotional balanced behaviour, situational awareness and reasoning and logical thinking. Here the assessor compares the information gathered from the students with the prefixed objectives and determines the effectiveness of teaching learning, curriculum, assessment procedures etc.

#### **4.4.8. Uses and Limitations**

Tyler's model of assessment is the first objective model of assessment in the field of curriculum and instructions. It can be used by the curriculum developer to evaluate the school curriculum by providing credible feedback from students. It can be very much helpful for teachers as well as school heads to assess the instructional process and achievement of learning outcomes. Many instructional decisions like suitability of curriculum, learning strategy, learning materials, assessment practices etc. can be taken on the basis of this model of assessment. This model follows a linear approach and proper sequence, hence systematic and organised way of assessment. The model suggested by Tyler is a lengthy and time-consuming process as it starts from board goals to making decisions. The model gives more stress on evaluation at the end of the implementation state forgetting the formative stage of teaching learning. Further, the model is silent about the process of evaluating objectives itself. To make this model useful, the assessor must be well versed with the rules and regulations of assessment.

#### ***Self-check exercise-2.1***

What are behavioural objectives and what are its needs in the Tyler model of assessment?
-----
-----
-----
-----
-----

---

#### **4.5. SUMMARY/KEY POINTS**

---

- **Model of educational assessment** gives a systematic procedure for conducting assessment to practitioners.
- **Ralph W. Tyler (1902-1994)** was an educator from America who contributed for the quality improvement of assessment and curriculum in his book *Basic Principles of Curriculum and Instruction*.
- The next step of Tyler's model is categorising the objective into dimension. Objectives may be directed towards developing different dimensions of development, such as **communication skill, demonstration skill, activity skills**.
- Different methods of measuring learning outcomes are interview, written examinations, tests, inventories, observations, checklists, rating scales, attitude scale, projective technique

---

#### 4.6. UNIT END EXERCISE

---

- Explain the Tylers Models of Assessment briefly
- Discuss the different steps followed in the Tylers Model of Assessment.
- Explain its relevance, uses and limitations of Tylers Model of Assessment .

---

#### 4.7. FURTHER READING

---

Goldie, J. (2006). AMEE Education Guide no. 29: Evaluating educational programmes. *Medical Teacher*, 28(3), 210–224. <https://doi.org/10.1080/01421590500271282>

Lonigan, C. J., Farver, J. M., Phillips, B. M., & Clancy-Menchetti, J. (2009). Promoting the development of preschool children's emergent literacy skills: a randomized evaluation of a literacy-focused curriculum and two professional development models. *Reading and Writing*, 24(3), 305–337. <https://doi.org/10.1007/s11145-009-9214-6>

Woods, J. A. (1988). Curriculum Evaluation Models: Practical Applications for teachers. *Australian Journal of Teacher Education*, 13(1). <https://doi.org/10.14221/ajte.1988v13n2>

## UNIT – 5

### STUFFLEBEAM’S CIPP MODEL FOR EDUCATIONAL ASSESSMENT

#### STRUCTURE

- 5.1. Learning Objectives
- 5.2. Introduction
- 5.3. Stufflebeam’s CIPP Model for Educational Assessment:
  - 5.3.1. Context
  - 5.3.2. Input:
  - 5.3.3. Process:
  - 5.3.4. Product:
  - 5.3.5. Uses and Limitations
- 5.4. Summary/Key Points
- 5.5. Unit End Exercise
- 5.6. Further Reading

---

#### 5.1. LEARNING OBJECTIVES

---

After reading the unit, students shall be able to;

- Explain the essence of Stufflebeam’s CIPP Model of Educational Assessment.
- Elaborate the Stufflebeam’s CIPP of Educational Assessment, its uses and implication
- Define the practical applicability of Stufflebeam’s CIPP of Educational Assessment in education

---

#### 5.2. INTRODUCTION

---

Assessment is an integral part of teaching learning process and curriculum development as it helps in the qualitative improvement of different elements of education. Assessment needs to be based on some models suggested by pedagogue and educationist. There are four models of assessment which are very relevant for assessing curriculum, programme and students learning. The model of assessment suggested by Ralph Tyler follows seven steps starting from deciding board goals to comparing data with the learning objectives. Metfessel and Michael extended Tyler's model of assessment and made it a goal oriented model of assessment. The main purpose of assessment is to ascertain the extent to which the learning

objectives are achieved by the learners. CIPP model of assessment advocated by Stufflebeam is very relevant and useful for evaluating any programmes like BEd, MA, Diploma. Stufflebeam has suggested assessment in four components such as context, input, process and product. Assessment is continuous and should be held at different times and levels. All the elements give feedback for improvement.

### 5.3. STUFFLEBEAM'S CIPP MODEL FOR EDUCATIONAL ASSESSMENT

Context, Input, Process, and Product (CIPP) are the four components that make up the comprehensive instrument known as the CIPP model of evaluation. This model of assessment was developed by *Phi Delta Kappa Committee Chairman Daniel Stufflebeam* (1936-2017), a Professor of Western Michigan University, US. The model can be used to study each of the four components of the curriculum/education system and is used to assess the quality of education at schools. This model is popularly known as the CIPP model of educational assessment, useful for teachers as well as curriculum developers for taking instructional decisions. Let's discuss the main four components of the CIPP model of assessment.

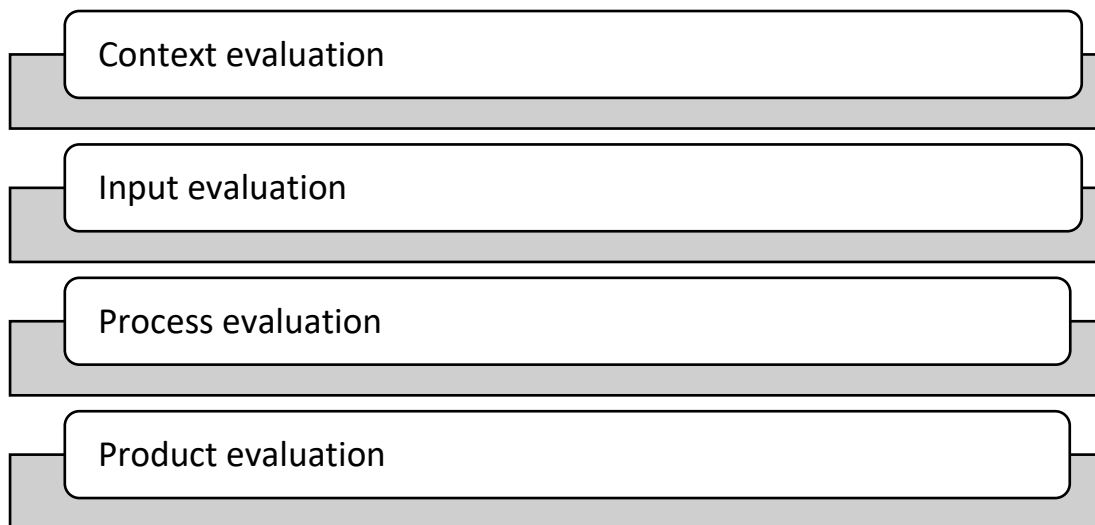


Figure-5.1: Infographic of CIPP model of assessment

### 5.4. Context

Assessing context is the first step in the educational assessment. Context refers to the situation or the environment in which the curriculum and teaching and learning will happen. It also includes the skills, competencies, values, and attitude that the education process wants to develop among learners. In fact, this step determines the goal and objectives of the assessment process.

This step of curriculum evaluation considers evaluating contexts of school curricular development that affects learning outcome and overall development of students. This may

include the directed competencies, skills and values that are intended to be developed among learners. The historical theories, citizenship duties, cultural heritage, scientific temper, environmental studies such as climate change issues and global citizenship can be considered as contexts for curriculum evaluation.

### **5.5. Input:**

"Input" may refer to the resources, materials, and individuals that are used to deliver the curriculum in the context of educational assessment. The input stage of the evaluation process is crucial because it gives assessors an insight into how well the curriculum is being applied. The utilization of resources and materials can be evaluated by evaluators to see whether they are effective and appropriate for the curriculum. They can assess whether teachers possess the abilities and information required to properly deliver the curriculum. Several techniques, including surveys, interviews, and observations, can be used by evaluators to assess input.

### **5.6. Process:**

The practice of assessing and evaluating how well the intended curriculum achieves the desired goals is known as curriculum evaluation. It is a crucial step in the adoption and implementation of any new curriculum. The process entails determining the evaluation's main audiences, issues, data sources, methodology, and criteria. It also entails assessing the needs of the pupils and contrasting the current curriculum with different programs. It's crucial to concentrate on how learning outcomes are assessed, making use of both qualitative and quantitative techniques. balancing formative and summative evaluations, keeping an eye on the program's usage of technology, reporting, and using the data to inform decisions.

### **5.7. Product:**

The product part of the CIPP model is concerned with evaluating the goals or outcomes of a program. It aids in determining a program's content or delivery's advantages and disadvantages, allowing for program improvement or future planning. The objectives decided at the beginning of the course are to be achieved by the learners in terms of knowledge, skills, competencies, attitude and values. In this stage of assessment, the test results are analysed and compared with the prefixed objectives to find out the level of the achievement. It indicates the quality of the program, course, syllabus, textbook, resources, learning outcomes.

#### ***Self-check- exercise-2.2***

Write down the different context that to be kept in mind when assessing a curriculum?

-----

-----
-----
-----
-----

### 5.8. Uses and Limitations

The CIPP framework was created as a way to integrate program decision-making with assessment. With a cycle of planning, structuring, implementing, reviewing, and changing decisions, it tries to give an analytical and reasonable basis for program decision-making. Each decision is analysed through a separate evaluation component, including context, input, process, and product evaluation. The CIPP approach aims to make assessment directly relevant to decision-makers' requirements throughout a program's stages and activities. As a framework for methodically directing the conception, design, implementation, and assessment of service-learning projects, as well as for providing feedback and judging the project's effectiveness for ongoing improvement, Stufflebeam's context, input, process, and product (CIPP) evaluation model is very relevant. The CIPP model provides a thorough instrument for assessing the current curricula. Context, Input, Process, and Product are the four components of the educational system that the model assists in analysing. The product portion of the model focuses on assessing the results or objectives of a program and identifies the program's content or delivery's advantages and disadvantages. This model suggested guidelines for taking decisions at different levels of education on the basis of the assessment following the CIPP model. But in reality, the decision making process is very complex which requires involvement of stakeholders and financial approval.

---

### 5.9. SUMMARY/KEY POINTS

---

- Context, Input, Process, and Product (CIPP) are the four components that make up the comprehensive instrument known as the CIPP model of evaluation.
- This model of assessment was developed by *Phi Delta Kappa Committee Chairman Daniel Stufflebeam* (1936-2017), a Professor of Western Michigan University, US.
- Context refers to the situation or the environment in which the curriculum and teaching and learning will happen.
- "Input" may refer to the resources, materials, and individuals that are used to deliver the curriculum in the context of educational assessment.



- The practice of assessing and evaluating how well the intended curriculum achieves the desired goals is known as curriculum evaluation.
- The product part of the CIPP model is concerned with evaluating the goals or outcomes of a program. It aids in determining a program's content or delivery's advantages and disadvantages, allowing for program improvement or future planning/
- The CIPP approach aims to make assessment directly relevant to decision-makers' requirements throughout a program's stages and activities

---

#### **5.10. UNIT END EXERCISE**

---

- Select any programme (BA/ MA/ Diploma) and evaluate it on the basis of the CIPP model of assessment
- Elaborate different dimensions of Stufflebeam's CIPP model for educational assessment/
- Explain its practical implications on educational measurement and evaluation

---

#### **5.11. FURTHER READING**

---

Goldie, J. (2006). AMEE Education Guide no. 29: Evaluating educational programmes. *Medical Teacher*, 28(3), 210–224. <https://doi.org/10.1080/01421590500271282>

Lonigan, C. J., Farver, J. M., Phillips, B. M., & Clancy-Menchetti, J. (2009). Promoting the development of preschool children's emergent literacy skills: a randomized evaluation of a literacy-focused curriculum and two professional development models. *Reading and Writing*, 24(3), 305–337. <https://doi.org/10.1007/s11145-009-9214-6>

Woods, J. A. (1988). Curriculum Evaluation Models: Practical Applications for teachers. *Australian Journal of Teacher Education*, 13(1). <https://doi.org/10.14221/ajte.1988v13n2>

---

## **BLOCK 02:**

### **INSTRUCTIONAL LEARNING OBJECTIVES**

---

**Unit 06:** Taxonomy of instructional learning objectives with special reference to cognitive domain

**Unit 07:** Criteria of selecting appropriate learning objectives

**Unit 08:** stating of general and specific instructional learning objectives

**Unit 09:** Relationship of evaluation procedure with learning objectives

**Unit 10:** Difference between objective based objective type test and objective based essay type test

### **UNIT-6:**

## **METFESSEL – MICHAEL MODEL FOR EDUCATIONAL ASSESSMENT**

### **Structure**

#### **6.1. Learning Objectives**

#### **6.2. Introduction**

#### **6.3. Metfessel – Michael Model for Educational Assessment**

- 6.3.1. Get the participants involved in determining the curriculum's objectives or define its goals
- 6.3.2. Transform the objectives into precise, quantifiable goals, as well as content and experience:
- 6.3.4. Create evaluation tools and technique before doing the evaluation:
- 6.3.5. Make observations and gather information on the goals that students are achieving:
- 6.3.6. Analyze the information and evaluate student's performance in relation to the goals:
- 6.3.7. Interpret the data to make judgments about the extent to which the curriculum is achieving its goals.
- 6.3.8. Use the results of the evaluation to make decisions about the curriculum and make recommendations.
- 6.3.9. Repeat the evaluation process on a regular basis:
- 6.3.10. Uses and Limitations

#### **6.4. Summary**

#### **6.5. Unit End Exercises**

#### **6.6. Further Reading**

---

### **6.1. LEARNING OBJECTIVES**

---

After reading the unit, students shall be able to;

- Explain stages of Metfessel-Michael Model of educational assessment.
- Describe steps and uses of Metfessel-Michael Model of educational assessment.
- Illustrate the stages of Metfessel-Michael Model of educational assessment
- Discuss the process of Metfessel-Michael Model of educational assessment.

---

### **6.2. INTRODUCTION**

---

Assessment is the integral part of the teaching and learning process. It determines the effectiveness of teaching of teachers, leadership of school heads, quality of textbooks, assessment practices and teacher education programmes. The achievement of learning outcomes by learners is ascertained by using assessment data. Hence, all the stakeholders of education from school to university are very much interested about the modalities and practices of assessment followed. So, the assessment must provide valid and reliable data to the policy makers and stakeholders as well. It must contribute to the quality improvement of

education. To make the assessment effective in helping the education process, it must follow models of assessment advocated by many psychometricians and educationists. Model of educational assessment gives a systematic procedure for conducting assessment to practitioners. The practitioners may be a teacher, school heads, curriculum developers, textbook writers, policy makers and educational administrators. Let's discuss Metfessel-Michael models of assessment and its uses in educational settings.

### 6.3. METFESSEL – MICHAEL MODEL FOR EDUCATIONAL ASSESSMENT

Norman Metfessel and James Michael developed the goal-oriented Metfessel-Michael Model of Curriculum Evaluation in 1967. This model is based on the idea that curriculum assessment should be focused on figuring out whether or not the program is accomplishing its objectives. Metfessel and Michael worked on the idea of Tyler's model of educational assessment and made it more relevant and useful for the education community.

#### Eight Steps of Assessment

This model suggested eight steps of assessment which are presented in the following pages.

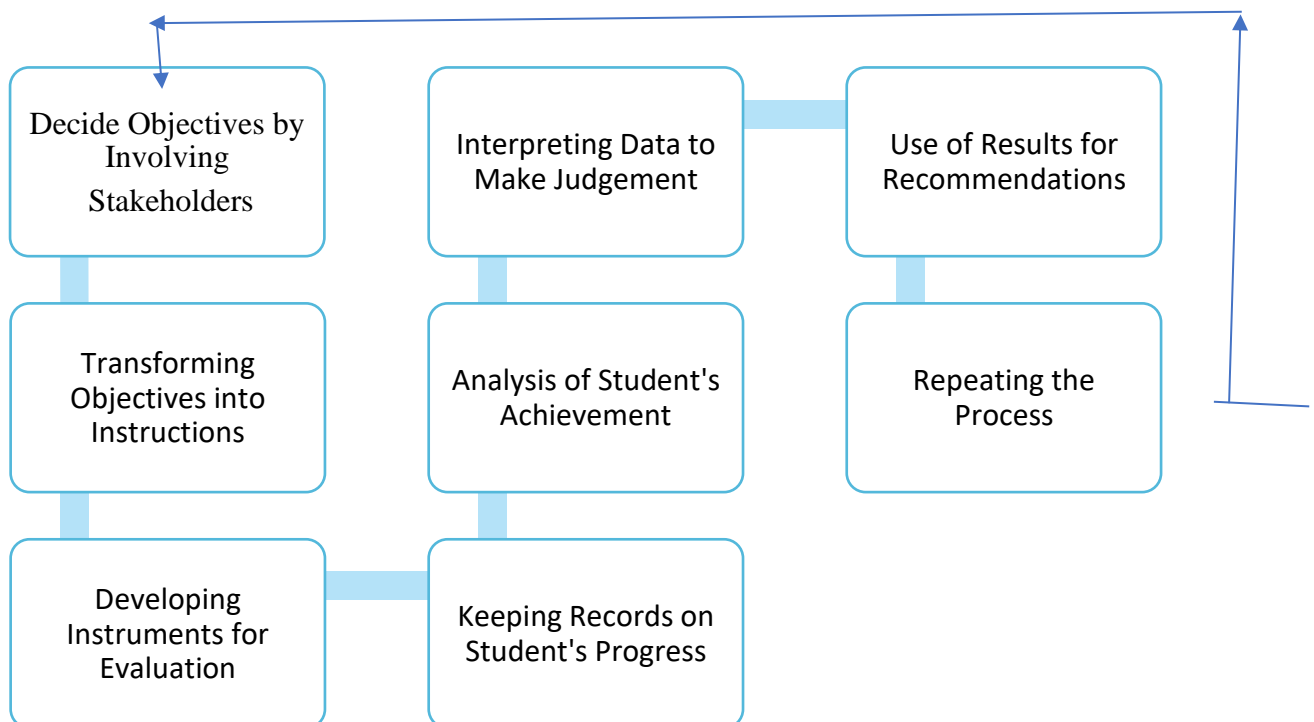


Figure-6.1: Infographic of Metfessel and Michael model of assessment

#### 6.3.1. Get the participants involved in determining the curriculum's objectives or define its goals:

The major stakeholders in the teaching learning process including the teachers, students, parents, heads of institutions, administrators, society and curriculum developers are to be involved in determining the objectives of the curriculum. This step talks about involving all of them in determining the objectives of the curriculum so that the curriculum can be relevant and useful for the learners and society. Hence, the first step in the assessment is to determine the objectives of the program or curriculum.

**6.3.2. Transform the objectives into precise, quantifiable goals, as well as content and experience:**

The broad goals should be converted into achievable learning outcomes and instructions through different teaching methods. The objectives decided by the stakeholders need to be stated precisely and objectively so that it can be observable and quantifiable. The objectives must include all round development domains of individuals such as social, physical, intellectual, psychic and emotional aspects. After defining the objectives in precise and quantifiable terms, it must be converted into content and learning experiences in different subjects. The assessor must see that objectives and content/learning experiences must be related to each other.

**6.3.3. Create evaluation tools and technique before doing the evaluation:**

Once objectives, content and learning experiences are decided and provided to the learners, the assessor must create tools and techniques of evaluation. These are the means of collecting information about learning from students. The desired learning outcomes can be aligned with specific instruments for evaluation such as interviews, tests, inventories, portfolios, anecdotes, and rating scales. These evaluation tools can be developed by the teacher as per the learning outcomes and content.

**6.3.4. Make observations and gather information on the goals that students are achieving:**

After getting the tools and techniques ready for assessment, the assessor can go on collecting valid and reliable information about the performance of students in different subjects. This step of assessment is very important as it provides the information for making decisions in the educational context. The recording and restoration of a student's progress should be kept in mind by the evaluators. The use of ICT can be helpful in keeping and tracking student's progress. ICT tools can be utilised to gather and store the collected information for future use.

**6.3.5. Analyze the information and evaluate student's performance in relation to the goals:**

Gathering the information about the student's learning has no use until it is analysed and interpreted to derive a certain conclusion. Analysis of the students performance is the process of breaking the information in pieces and comparing it with different criteria (objectives). Different statistical techniques such as measures of central tendency, dispersion and inferential statistics etc. can be used for analysing students' performance. Assessor must analyse the students' performance keeping in mind the learning outcomes/ objectives.

**6.3.6. Interpret the data to make judgments about the extent to which the curriculum is achieving its goals.**

The goals that were set in the first steps should be compared to the learning outcomes and analysed to the extent till which they have been achieved. The analysis of the collected information will lead to interpretation. Interpretation is the process of drawing conclusions on the basis of the evidence collected. Variety of interpretation about curriculum and teaching learning process can be made on the basis of the students performance. The interpretation like curriculum needs to be renewed or redesigned, teachers require training in new methodology etc. can be made.

**6.3.7. Use the results of the evaluation to make decisions about the curriculum and make recommendations.**

The findings of the results can be used to come up with recommendations and suggestions in terms of policies and reforms in improving the standards of the current teaching learning process. The main purpose of the assessment is to provide credible information for making instructional decisions like change of curriculum, teacher training, assessment process, teaching learning resources etc. On the basis of the interpretation, the assessor can use the result for changing the curriculum, developing a new textbook, innovative pedagogy etc.

**6.3.8. Repeat the evaluation process on a regular basis:**

The process for improvement is continuous and hence there is always scope for improvising the quality of evaluation. It must be done on a regular basis. In fact, assessment is the integral part of the curriculum making and teaching learning process. To make the education qualitative, assessment can be conducted on a regular basis by following the steps discussed above.

**6.3.9. Uses and Limitations**

A comprehensive and methodical approach to curriculum review is the Metfessel-Michael Model. For instructors who want to make sure that their curriculum is meeting the needs of

their students, it is a beneficial instrument. It is a goal-oriented paradigm , therefore its attention is on figuring out whether the curriculum is succeeding in accomplishing those goals. It is a systematic technique, which implies it adheres to a sequential methodology. Because it is comprehensive, it takes into account every component of the curriculum. It has certain limitations like, steps proposed are very lengthy and require involvement of all stakeholders of education.

**Self-check Exercise: 2.3**

What are the benefits of involving stakeholders in deciding objectives of the program as per the Metfessel-Michael Model of evaluation?

---

---

---

---

---

---

---

---

---

## 6.4. SUMMARY

---

- Norman Metfessel and James Michael developed the goal-oriented Metfessel-Michael Model of Curriculum Evaluation in 1967. This model is based on the idea that curriculum assessment should be focused on figuring out whether or not the program is accomplishing its objectives
- Once objectives, content and learning experiences are decided and provided to the learners, the assessor must create tools and techniques of evaluation.
- The desired learning outcomes can be aligned with specific instruments for evaluation such as interviews, tests, inventories, portfolios, anecdotes, and rating scales. These evaluation tools can be developed by the teacher as per the learning outcomes and content.
- The goals that were set in the first steps should be compared to the learning outcomes and analysed to the extent till which they have been achieved.
- Analysis of the students performance is the process of breaking the information in pieces and comparing it with different criteria (objectives).

- The main purpose of the assessment is to provide credible information for making instructional decisions like change of curriculum, teacher training, assessment process, teaching learning resources etc.

---

## **6.5. UNIT END EXERCISES**

---

1. Design a framework for developing and assessing the curriculum at secondary school level of your state by following Metfessel – Michale.
2. Select any programme (BA/ MA/ Diploma) and evaluate it on the basis of Metfessel – Michale of assessment.
3. Describe the differences between Tyler model and Metfessel model of assessment.

---

## **6.6. FURTHER READING**

---

Goldie, J. (2006). AMEE Education Guide no. 29: Evaluating educational programmes. *Medical Teacher*, 28(3), 210–224. <https://doi.org/10.1080/01421590500271282>

Lonigan, C. J., Farver, J. M., Phillips, B. M., & Clancy-Menchetti, J. (2009). Promoting the development of preschool children's emergent literacy skills: a randomized evaluation of a literacy-focused curriculum and two professional development models. *Reading and Writing*, 24(3), 305–337. <https://doi.org/10.1007/s11145-009-9214-6>

Woods, J. A. (1988). Curriculum Evaluation Models: Practical Applications for teachers. *Australian Journal of Teacher Education*, 13(1). <https://doi.org/10.14221/ajte.1988v13n2>



## UNIT-07

# PROVU'S DISCREPANCY MODEL FOR CURRICULUM EVALUATION

## Structure

- 7.1. Learning Objectives
- 7.2. Introduction
- 7.3. Provu's Models of Educational Assessment
  - 7.3.1. Program design
  - 7.3.2. Installation:
  - 7.3.3. Process
  - 7.3.4. Product
  - 7.3.5. Cost Benefit Analysis
  - 7.3.6. Uses and Limitations
- 7.4. Summary
- 7.5. Unit End Exercises
- 7.5. Further Reading

---

## 7. 1. LEARNING OBJECTIVES

---

After reading the unit, students shall be able to;

- Explain stages of Provu's model of educational assessment.
- Describe steps and uses Provu's of educational assessment.
- Illustrate the stages of Provu's model of educational assessment
- Discuss the process of using Provu's Discrepancy model of educational assessment.

---

## 7.2. INTRODUCTION

---

Assessment is the integral part of the teaching and learning process. It determines the effectiveness of teaching of teachers, leadership of school heads, quality of textbooks, assessment practices and teacher education programmes. The achievement of learning outcomes by learners is ascertained by using assessment data. Hence, all the stakeholders of education from school to university are very much interested about the modalities and practices of assessment followed. So, the assessment must provide valid and reliable data to

the policy makers and stakeholders as well. It must contribute to the quality improvement of education. To make the assessment effective in helping the education process, it must follow models of assessment advocated by many psychometricians and educationists. Model gives a framework and step by step process of conducting assessment so that it gives a valid and reliable result. Model makes assessment organised, systematic, usable, and objective. All the teachers and educators must understand the nitty gritty of different educational models so that they can conduct assessment. There are many models of assessment proposed by different authors. In this unit, Provu's model of assessment are discussed in detail with its uses and limitations.

Model of educational assessment gives a systematic procedure for conducting assessment to practitioners. The practitioners may be a teacher, school heads, curriculum developers, textbook writers, policy makers and educational administrators. Let's discuss Provu's model of assessment and its uses in educational settings.

---

### **7.3. PROVU'S MODELS OF EDUCATIONAL ASSESSMENT**

---

It is common to compare things/programmes/curriculum with prescribed standards to find out its worth. Assessment can be done by finding discrepancies in performance at different levels so that corrective measures can be taken to improve. This is the basic idea on which Provu's model of assessment is based. Malcolm Provu created the Discrepancy Evaluation Model (DEM) in 1956, which offers data for program evaluation and enhancement. The DEM defines evaluation as a comparison between a performance's (P) actual results and a desired standard (S). The difference between performance and standard is called discrepancy (D). Program design, installation, process, product, and cost benefit analysis are the five stages of evaluation that the DEM represents based on a program's natural development. The DEM's evaluation data collection aids in career planning and placement. These choices can be classified into three categories:

- Choices relating to program design or analysis
- Choices relating to the accomplishment of both intermediate and ultimate goals
- Choices relating to the program's current operations.

This model offers data for program evaluation and enhancement. The steps of DEM are discussed here with the assumption that it improves the quality of curriculum evaluation. This model is graphically presented in figure-2.4.



Figure: 7.1.: Graphic presentation of discrepancy model of evaluation

### 7.3.1. Program design:

The first step is to design the programme or curriculum. Design of the programme for curriculum aims at developing learning outcomes, competency, and knowledge which ultimately lead to good career choices and its role in preparing the students through the programme. Programmes/curriculum must be based on certain standards so that it is relevant for learners.

### 7.3.2. Installation:

Once the programme is designed, it must be installed to see its effect on learners. The installation stage provides an opportunity to identify and collect the resources required to properly and faithfully deliver a parenting curriculum. Further, during this phase it is made sure that the instruments and resources required for successful development are available.

### 7.3.3. Process

The curriculum that will be assessed must first go through several processes of preparation steps. The steps can be;

- Specification; it includes defining whom the curriculum is intended for, who will conduct the assessment and what the primary goals of the curriculum are that will be assessed. When the specification is in place, the curriculum has a foundation upon which it can be assessed.
- Resource mobilisation; it includes proper channelization of available resources required for best possible output.
- Operationalisation; it includes implementation of the developed process and formula aimed at optimum use of curriculum evaluation.

#### 7.3.4. **Product**

This step outlines the learning objectives for the cognitive, emotional and psycho-motor domains. This fragmentation of human abilities required specific and individualised programme curriculum for attaining holistic development. It is necessary to determine the product of the programme in terms of the students' learning and development.

#### 7.3.5. **Cost Benefit Analysis**

This step helps to understand whether a project decision makes sense from a business standpoint. Cost benefit analysis compares the expected or estimated cost and benefits connected with the evaluation model. This step is very important as it helps to see the product in terms of cost used.

Based on a program's natural development Malcolm Provu created the Discrepancy Evaluation Model (DEM). The DEM's evaluation data collection aids career planning and placement. It helps counsellors in making thoughtful decisions.

#### ***Self-check-Exercise-2.4***

What is the discrepancy according to Provu's Model of assessment?

---

---

---

---

#### 7.3.6. **Uses and Limitations**

The model created by Malcolm Provu deviates from the presumption that assessors can compare standards with performance, or what should happen and what actually happened, to establish the viability of a program. By contrasting these two items, a discrepancy between the performance expected and the actual performance can be seen. This model is designed to evaluate programs and determine whether they should be stopped, improved, or continued. The terms standard, performance, detail discrepancy, and measurability are also used in this concept. In this instance, the evaluator calculates the size of each gap in the program evaluation. Corrective actions can be carried out clearly by identifying the deficiencies in each program component. This model is used to determine the degree of conformance

between the performance that results from the program and the standards stated in the program. This model of assessment is not economical in time and money as it involves a lengthy process to determine discrepancy. Further, it is not easy to determine standards for each step to compare with actual results.

---

## 7.4. SUMMARY/KEY POINTS

---

- Malcolm Provu created the Discrepancy Evaluation Model (DEM) in 1956, which offers data for program evaluation and enhancement.
- The DEM defines evaluation as a comparison between a performance's (P) actual results and a desired standard (S).
- Design of the programme for curriculum aims at developing learning outcomes, competency, and knowledge which ultimately lead to good career choices and its role in preparing the students through the programme.
- Cost benefit analysis compares the expected or estimated cost and benefits connected with the evaluation model

---

## 7.5. UNIT END EXERCISES

---

4. Design a framework for developing and assessing the curriculum at secondary school level of your state by following Provu's model of assessment.
5. Select any programme (BA/ MA/ Diploma) and evaluate it on the basis of the Provu's model of assessment
6. Describe the differences between Tyler model and Provu's model of assessment
7. Provu's model of assessment is known as discrepancy model. Describe different types of discrepancy the model suggested and how it is helpful for the improvement of education.

---

## 7.6. FURTHER READING

---

Goldie, J. (2006). AMEE Education Guide no. 29: Evaluating educational programmes. *Medical Teacher*, 28(3), 210–224. <https://doi.org/10.1080/01421590500271282>

Lonigan, C. J., Farver, J. M., Phillips, B. M., & Clancy-Menchetti, J. (2009). Promoting the development of preschool children's emergent literacy skills: a randomized evaluation of a literacy-focused curriculum and two professional development models. *Reading and Writing*, 24(3), 305–337. <https://doi.org/10.1007/s11145-009-9214-6>

Woods, J. A. (1988). Curriculum Evaluation Models: Practical Applications for teachers. *Australian Journal of Teacher Education*, 13(1). <https://doi.org/10.14221/ajte.1988v13n2>

## **UNIT-08**

### **MEANING OF TOOLS AND TECHNIQUES IN ASSESSMENT**

#### **STRUCTURE**

- 8. 1. Learning Objectives
- 8.2.Introduction
- 8.3.Meaning of Tools and Techniques in assessment
  - 8.3.1. Difference between Tools and Techniques
- 8.4.Subjective and Objective tools of assessment
- 8.5 Essay and Objective tests.
  - 8.5.1. Essay Tests
  - 8.5.2. Characteristics of Essay Test
  - 8.5.3. Limitations of Essay Test
  - 8.5.4. Objective Tests
  - 8.5.5. Characteristics of Objective Tests
  - 8.5.6. Limitations of Objective Tests
- 8.6. Summary/Key Points
- 8.7. Unit End Exercise
- 8.8. Suggestions for Further Reading

---

#### **8.1. LEARNING OBJECTIVES**

---

After reading this unit, learners shall be able to;

- define and differentiate tools and techniques used in educational assessment.
- distinguish between subjective and objective assessment tools.
- Explain the uses various types of objectives and subjective test

---

#### **8.2. INTRODUCTION**

---

Assessment is an integral part of the teaching learning process which contributes greatly for gathering ideas about the knowledge, abilities, and academic needs of students so that better planning can be done for bringing improvement in the instructional strategies. In the realm of assessment, the role of various tools and techniques is paramount to help students learn more effectively. Teachers are the main assessors of students in the classroom; they create assessment tools with two main objectives in mind: gathering data to guide classroom instruction and tracking students' progress toward meeting year-end learning objectives. Teachers also help pupils learn techniques for self-monitoring and self-evaluation so they should make sure that students participate in setting learning objectives, creating action plans, and employing assessment procedures to track their accomplishment of

objectives to accomplish this effectively. Some commonly used tools and techniques for assessment are questionnaire, interview schedule, observation, rating scale etc. In this unit the concept and uses of different tools & techniques used in the educational assessment are discussed.

---

### 8.3.MEANING OF TOOLS AND TECHNIQUES IN ASSESSMENT

---

Tools and techniques in assessment are the instruments which help in determining the learning interventions needed for enhancing the academic proficiency of students. Several tools and techniques are used by the teachers to measure the academic abilities, skills, behavioral patterns, personality, and various other factors of a student's development. Tools in measurement mean the instruments employed for data collection, while techniques involve the methods and approaches applied to extract meaningful insights from the collected information. In educational context, tests or questions are examples of tools and administering either in group or individually is the example of techniques.

Traditional tools include written tests, quizzes, and examinations, providing a structured approach to evaluating knowledge and understanding. Performance assessments, such as presentations, portfolios, and practical demonstrations, offer a more holistic view of skills and abilities. Surveys and questionnaires serve as valuable tools for gathering subjective data, allowing for the exploration of attitudes, opinions, and experiences. Rubrics, scoring guides, and checklists provide standardized criteria for evaluation, enhancing objectivity and consistency in assessment.

The synergy between tools and techniques is crucial for effective assessment. The selection of appropriate tools depends on the nature of what is being assessed and the desired outcomes. Techniques guide the application of these tools, ensuring that the assessment process is fair, reliable, and valid.

#### **Difference between Tools and Techniques**

Tools are the physical or digital resources that facilitate the assessment process. They are the tangible devices or materials which teachers use to gather information to evaluate performance of students. For example, a written test, a survey, an interview protocol, or even a rubric can be considered tools.

On the other hand, techniques refer to the methods or approaches you employ to implement these tools effectively. It is about how you use the tools to collect, analyze, and interpret data. Techniques involve the skills, procedures, and strategies applied during assessment. This could include things like observation skills, data analysis methods, or even specific interviewing techniques.

In essence, tools are the concrete items you use, and techniques are the methods you apply to make the assessment process meaningful and reliable. They work hand in hand to ensure a comprehensive and accurate assessment.

<b><i>Self-check Exercise-8.1</i></b>
Write down three tools and two techniques that can be used in school by teachers.
-----
-----
-----
-----

---

#### 8.4.SUBJECTIVE AND OBJECTIVE TOOLS OF ASSESSMENT

---

As we have discussed earlier, Tools in the assessment process play a very significant role as they are the means to measure several aspects of development in a child. Normally in a classroom scenario, teachers use two common methods of assessment such as subjective and objective assessment. Objective assessment method is comparatively quicker than subjective assessment and it provides accurate information and with clear and precise evaluation. Subjective assessment on the other hand is time consuming as it provides more comprehensive information about the knowledge and skills of students. The objective tools bring structure and clarity, while the subjective ones add depth and personal insight to the assessment.

Essays, portfolios, oral presentations, projects, open-ended questions, or performance evaluations fall into the subjective category. They focus on the quality of student's work, their creativity level, analyzing potential and divergent thinking capacity rather than specific correct answers. The objective category evaluates the student's knowledge and understanding of specific facts or concepts; this category includes multiple-choice exams, numerical scores, rubrics, true-false exercises promoting fair evaluation for all the students. The difference between objective and subjective tools are presented in table-3.1.

**Table-8.2: Difference between objective and subjective tools of assessment**

<b>Objective Tools</b>	<b>Subjective Tools</b>
The assessment tools which focus to measure the knowledge of students on particular facts and figures are categorized under objective assessment tools.	The assessment tools which are used to measure the performance of students that are qualitative in nature are categorized as subjective tools.
It takes more time to prepare objective tools whereas less time is required for evaluating the students.	It is easy to prepare subjective tools but more time is required for evaluating the students.
Objective tools can measure both simple and complex learning outcomes	Subjective tools generally include items from understanding, application domains or other higher order skills.
Items in the objective tools can be scored as either right or wrong.	Items in subjective tools can be partially right or wrong.
Examples: Multiple choice questions, True/False items, Fill in the blanks etc.	Examples: essay, application, letter, long or open-ended questions.
Objective tests can cover many items and content sampling is adequate.	Subjective tools can cover a small number of test items and some contents may mis in the test.
Objective tools are usually highly valid and reliable.	Subject tools are difficult to ensure validity and reliability
Machine/software can be used for scoring objective tools	Subjective tools are manually scored and time consuming
Objective tools can be standardized across the Globe	Subjective tools are difficult to standardized



Objective tools are suitable for national and state level assessments like NAS/SAS	Only subjective tools are rarely used in national and state level assessment.
--	---

**Self-check Exercise-8.2**

List three subjective and three objective tools of assessment used in school education.

---

---

---

---

---

---

**8.5.ESSAY AND OBJECTIVE TESTS.**

There are several methods used by the teachers to assess the knowledge, understanding and skills of students. Two widely used methods among them are essay tests and objective tests. Each method has its unique characteristics, strength and applications contributing towards effective assessment. By incorporating both the assessment methods in a balanced way, we can obtain a holistic understanding of students' abilities resulting in a meaningful assessment practice.

**8.5.1. Essay Tests**

Essay test refers to written tests where students are required to compose their answers usually in the form of sentences, paragraphs, or passages that measure their understanding, thinking skills and other complex learning objectives. Essay type tests have intensive use in higher education and are very popular in classroom testing. In essay tests, learners are allowed to select, organize their ideas, and express their thoughts which they consider appropriate. When teachers want the learner to articulate their thoughts and demonstrate a deeper comprehension of the subject matter, they use this kind of test. It is characterized by open-ended questions, demand a subjective judgment about the quality of the answer. While answering, students get the freedom of response to think diversely in their own way which reflects their creativity level, ability to organize thoughts meaningfully, integrate ideas and other thinking skills. A student should have thorough knowledge and understanding of the concept to answer essay type items.

**8.5.2. Characteristics of Essay Test**

- Essay type test items enable students to plan their answer, organize, and express in their own words.
- In essay tests, questions are generally broad and open ended.
- Students get much freedom to express their individuality in essay tests.
- The answer may be partially right or wrong; there are provisions for partial credit for answers.
- Students who have good language skills are at an advantage in essay tests.
- Scoring in essay test is subjective in nature and inconsistent from teacher to teacher.

- Essay tests require subjective judgment of the skilled or informed person in that area to judge the quality and accuracy of the response.
- Essay type test items are easier to construct than multiple choice items as there is no need to create effective distractors; so, they are less time consuming in preparation.

### Self-check exercise-8.3

Prepare five essay questions for assessing school level subjects.
-----
-----
-----
-----
-----
-----
-----

#### 8.5.3. Limitations of Essay Test

- In essay tests, there are chances of subjectivity in scoring i.e., personal factors of the evaluator influence scoring.
- The scoring is unreliable in essay tests, it differs from teacher to teacher.
- Essay type test items require much time for answering and evaluating.
- It is greatly influenced by the handwriting and language abilities of students.
- It may encourage bluffing and selective study habits among students.

To minimize the weakness of subjectivity in scoring, a model answer sheet should be prepared including the important key points and other considerations. Scoring of all responses to a single item should be done at a time to avoid subjective bias of the evaluator. Essay tests if prepared carefully can measure complex learning outcomes such as critical thinking, creative thinking, problem solving, innovative thinking etc. among students. It is essential for teachers to use essay tests only when learning outcomes cannot be measured by objective tests.

#### 8.5.4. Objective Tests

Objective tests are highly structured tests that are designed to get clear and specific answers for a problem/question. Unlike subjective tests, it aims to test students' ability to recall information and other complex learning outcomes. Objective tests are considered a popular choice for large scale assessments such as National Achievement Survey/ State Achievement Survey because they are reliable in grading and free from the examiner's subjective opinion. According to Gilbert, "Any test having clear and unambiguous scoring criteria is an objective test." However, students' ability to organize ideas, problem-solving skills and creativity level cannot be measured using objective tests. Objective type tests can cover a wide range of syllabus in comparison to the subjective tests as it includes larger and more representative samples having high content validity. One may feel that it is easy to prepare objective test items, but it is not so; to prepare objective test items it requires mastery of knowledge in that

area, creativity ability as well as clear knowledge about the objectives and content of the course.

While preparing objective test items, it should be kept in mind that the items should be independent and one item does not give any clue to other items. Some common examples of objective type questions are multiple choice items, fill in the blank, true-false type, matching type etc. In current time, objective tests are being broadly used in all kinds of entrance examinations due to their characteristics of high reliability and validity.

#### **8.5.5. Characteristics of Objective Tests**

- Objective tests can measure both simple and complex learning outcomes like analytical or reasoning skills.
- Answering objective type test items takes less time in comparison to essay type items, so it is time-efficient.
- It allows teachers to cover a broad area of content for assessing large numbers of students with reduced marking workload.
- Objective tests are easy to administer and score.
- Answers in objective tests are either wrong or right; there is no scope for partial credit.
- Scoring possesses objectivity, hence reliable.
- Nature of scoring is generally quick(rapid), easy and consistent.
- Machine/software can be used for objective test administration and scoring.

#### **8.5.6. Limitations of Objective Tests**

- Objective tests are not suitable for testing certain skills like organization, writing abilities and abilities to present matter systematically etc.
- It requires expertise to construct objective type test items; constructing objective items are difficult as well as time consuming.
- There is less freedom given to students for expressing and explaining their views in objective tests.
- While answering multiple choice or true/false type items, it's possible to blindly guess the answer without having any idea about it. It may encourage guessing among students.

---

### **8.6. SUMMARY/KEY POINTS**

---

- Tools and techniques in assessment are the instruments which help in determining the learning interventions needed for enhancing the academic proficiency of students.
- Tools in measurement mean the instruments employed for data collection, while techniques involve the methods and approaches applied to extract meaningful insights from the collected information.
- Surveys and questionnaires serve as valuable tools for gathering subjective data, allowing for the exploration of attitudes, opinions, and experiences
- Tools are the physical or digital resources that facilitate the assessment process

- The objective tools bring structure and clarity, while the subjective ones add depth and personal insight to the assessment.
- Essay test refers to written tests where students are required to compose their answers usually in the form of sentences, paragraphs, or passages that measure their understanding, thinking skills and other complex learning objectives.
- Objective tests are highly structured tests that are designed to get clear and specific answers for a problem/question.
- Objective type tests can cover a wide range of syllabus in comparison to the subjective tests as it includes larger and more representative samples having high content validity.

---

### 8.7. UNIT END EXERCISE

---

1. Compare and contrast formative and summative assessment tools, providing examples of each.
2. Design a comprehensive assessment plan for a specific educational scenario, incorporating a mix of subjective and objective tools. Justify your choices based on the learning objectives and context.
3. Compare and contrast essay tests and objective tests. In what situations might one type of test be more appropriate than the other?

---

### 8.8. SUGGESTIONS FOR FURTHER READING

---

Gronlund, N. E. (1965). *Measurement and evaluation in teaching*.  
<http://ci.nii.ac.jp/ncid/BA12623208>

Goswami, M. (2013). *Measurement and Evaluation in Psychology and Education*  
Lee, W. Y. (2010). *Assessment and evaluation in higher education*.  
<http://ci.nii.ac.jp/ncid/BB11596810>

Patel, R. N. (2014). *Educational evaluation theory and practice*.  
<http://ci.nii.ac.jp/ncid/BA29677030>

## UNIT -09

### CONCEPT OF SCALES AND QUESTIONNAIRES

#### STRUCTURE

- 9. 1. Learning Objectives
- 9.2.Introduction
- 9.3.Scales
  - 9.3.1. Advantages of Rating Scale
  - 9.3.2. Descriptive graphic rating scale
  - 9.3.3. Disadvantages of Rating Scale
  - 9.3.4. Graphic rating scale
  - 9.3.5. Numerical rating scale
  - 9.3.6. Rating Scale
  - 9.3.7. Types of Rating Scales
  - 9.3.8. Uses of Rating Scale
- 9.4.Questionnaire
  - 9.4.1. Structured Questionnaire
  - 9.4.2. Unstructured Questionnaire
  - 9.4.3. Principles of Preparing Questionnaire
  - 9.4.4. Advantages of Questionnaire
  - 9.4.5. Disadvantages of Questionnaire
- 9.5. Summary/Key Points
- 9.6. Unit End Exercise
- 9.7. Suggestions for Further Reading

---

#### 9.1. LEARNING OBJECTIVES

---

After reading this unit, learners shall be able to;

- Define and differentiate tools and techniques used in educational assessment.
- Distinguish between subjective and objective assessment tools.
- Explain the uses various types of objectives and subjective test

---

#### 9.2. INTRODUCTION

---

Level of measurement or scale of measure is a classification that describes the nature of information within the numbers assigned to variables. Psychologist Stanley Smith Stevens developed the best known classification with four levels or scales of measurement: nominal, ordinal, interval, and ratio. In this article, Stevens claimed that all measurement in science was conducted using four different types of scales that he called "nominal" "ordinal" "interval" and "ratio" unifying both "qualitative" (which are described by his "nominal" type) and "quantitative" (all the rest of his scales).

---

#### 9.3. SCALES

---

Aim of education is to bring the holistic development of learners through teaching and assessing holistically. For holistic assessment, teachers need to use both testing and non-testing devices. Because tests are not suitable for assessing all kinds of learning outcomes and competencies. The learning outcomes related to affective and psychomotor domains such as attitude, values, appreciation, perception, performance, skills, interests, competencies etc. are hard to measure through tests. These non-cognitive abilities of learners are to be assessed through scales, inventories, schedules, checklists, questionnaires etc. The test gives the information about students' performance in quantitative terms which is objective and correctly interpretable. Scale gives information about students' development in non-cognitive qualities in ratings or presence/absence, agree/disagree etc. The results of the scale can be utilized for inculcating non-cognitive abilities of students through assistance, mentoring and guidance/counseling.

### **9.3.1. Rating Scale**

Rating scale is one of the widely used non-testing tools which consists of a set of characteristics or attributes to be judged and rating points/descriptors. It indicates the degree to which a certain trait or characteristic is present in the behavioral pattern. It is one of the methods of recording observation objectively and systematically. Rating scale is a standardized method of recording behavior with which individuals can be rated on a scale from low to high with respect to a particular trait. While scoring using rating scale, the 'Rater' or 'Observer' or 'Teacher' assigns a value to each characteristic according to pre-fixed criteria. Odd number pointed rating scales are used in normal settings like 3-, 5- & 7-point rating scales. According to Gronlund, "The value of rating scale in appraising the learning and development of pupils depends largely upon the care with which it is prepared and appropriateness with which it is used."

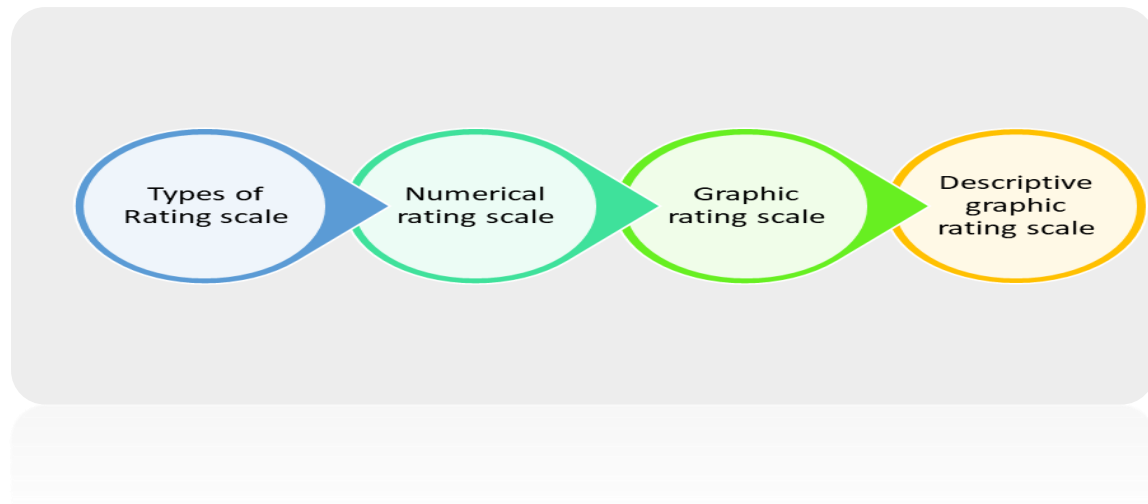
Instead of merely indicating the presence or absence of a specific attribute like in the checklist, rating scale provides more comprehensive information. A rater or observer should be fully instructed about the purpose and right use of the measuring tool. They should fill the rating scale during observation, or immediately after observation to ensure objectivity of the tool. The qualities like skills of communication, skills in demonstration, performance in art and drama, laboratory experiments, attitudes, values, etc. can be assessed by rating scales.

### **9.3.2. Types of Rating Scales**

#### **Types of Rating Scales**

There are three mostly used rating scales in education setting are;

- Numerical rating scale
- Graphic rating scale
- Descriptive graphic rating scale



### 9.3.3. Numerical rating scale

Like the name reflects, there are numbers indicating the degree of present characteristics in a numerical rating scale. The rater needs to assign some value to a particular trait. One has to mark or encircle the numbers each representing their own verbal descriptions. The attribute to be measured is presented in the form of a sentence and numbers representing the key values are to be chosen. For example;

*Direction:* Encircle the appropriate number indicating the performance of students in a dance competition. The number 1 represents least and the number 5 represents best.

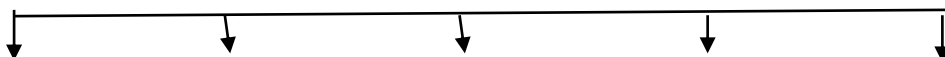
Sl No.	Criteria	1	2	3	4	5
1	Dance moves were up to the mark					
2	Facial expressions used were correct					
3	Follows the music					
4	Demonstrate proper mudras					
5	Costume is appropriate					

### 9.3.4. Graphic rating scale

In graphic rating scale, ratings are made in graphic form instead of using scale values. In this type of rating scale, each trait is presented in sentences followed by a horizontal line. The observer must put a mark on the line indicating the extent to which a certain trait is present. For Example;

*Direction:* Indicate the extent to which the student is present in co-curricular activities class by putting a tick mark ( ) along the horizontal line under each item.

1. To what extent the student is present in the co-curricular activities classes?



Never                  Seldom                  Occasionally                  Frequently                  Always

2. To what extent the student is eager to participate in the co-curricular activities?

Never                  Seldom                  Occasionally                  Frequently                  Always

### 9.3.5. Descriptive graphic rating scale

As the name suggests, descriptive phrases are used to identify points on the graphic scale. The description helps to clarify a particular dimension and it makes the descriptive graphic rating scale the most desirable type of scale to give rating. In some scales, only center and end points are described whereas in some, description is given under each point. A space for comments is also provided where needed.

1. To what extent the student is willing to take part in science lab experiments?

Never participates, participate normally very eagerly  
 participate  
 Not active as other group members with interest

### **Self-check Exercise-3.4**

Write down five items for assessing laboratory skills of students by using numeric rating scales.

---

---

---

---

---

---

---

### 9.3.6. Uses of Rating Scale

- Rating scale helps the teacher to rate their students on several personality traits like honesty, leadership quality, cooperativeness; and other skills like song, dance, debate, drawing, model, handwriting etc.
- Teachers can use rating scales to evaluate their teaching strategies, teaching learning methods and materials they use and their instructional procedures.
- It can be used to measure the attitude or perception of students towards the teaching learning process.
- Students of the higher classes can rate themselves using this tool which improves judgmental skills in them.



- This comprehensive tool provides the views and opinions of individuals on certain characteristics.
- They measure specific learning outcomes which are significant for the teacher.

#### **9.3.7. Advantages of Rating Scale**

- It works as a comprehensive platform for comparing learners on the basis of the same set of characteristics.
- Rating scale directs observation towards specific aspects of behavior and is widely used in educational fields.
- Students can also use a rating scale to rate their behavior in the process of self-assessment and peer assessment.
- It can be used for taking opinions from many students.
- Rating scale is easy to construct, economical in use and flexible in nature.

#### **9.3.8. Disadvantages of Rating Scale,**

- Rating scales tend to be less reliable because there are chances of subjectivity in scoring
- Examiner's value system, belief or pre-conception may affect the result of students. To avoid this situation and ensure reliability, multiple observers can rate the same sample with the same scale.
- The halo effect is a common type of error of rating scale that occurs when an observer's general perception about a person influences his rating.

Rating is a very popular tool in measuring a series of educational and personality traits in educational settings. Teachers should know the process of developing rating scales as per the requirement of class. Of course, there are certain standardized rating scales available, but these rating scales may not serve the purpose of a particular teacher. Utmost care must be taken by the rater/ teacher when marking on the scale. Because rating is badly influenced by the preconception and beliefs of the rater. Hence, rater/ teacher must be unbiased and objective in observation and recording the ratings.

---

### **9.4.QUESTIONNAIRE**

---

Questionnaire is one of the most extensively used assessment tools in teaching, learning and research purposes. It consists of a set of questions/statements to gather information from respondents/students on a variety of problems, issues, and lessons. In order to gather firsthand information from the learners, questionnaires are a very effective tool. Sir Francis Galton is known as the inventor of questionnaire. It is generally used to measure the attitudes, thought, beliefs, behavior, preference, and perception of a relatively large number of people. Questionnaire is considered as a self-report data collection instrument in which space is provided under each question for writing answers. It is widely used in research and assessment because it is comparatively cheap, less time consuming and convenient to use than other tools. According to John. W. Best, “ a questionnaire is used when factual

information is desired, when opinion rather than facts is required, an opinionnaire or attitude scale is used.” Questionnaire is broadly classified into two types such as structured questionnaire and unstructured questionnaire.

#### 9.4.1. Structured Questionnaire

In structured questionnaires, pre-determined and definite questions are used for accurate communication and responses. Each question is followed by options and students are required to select one or more options. The questions are restricted and respondents get less freedom to answer the questions. Nature of scoring in a structured questionnaire is objective, accurate and less time consuming.

#### 9.4.2. Unstructured Questionnaire

Unstructured questionnaires are those which include open and unrestricted questions and allow learners to express their thoughts in a free-flowing manner. Students are given freedom to share their ideas with depth. These types of questions record more data as they do not have a predetermined set of responses. The nature of scoring is subjective in nature and it sometimes becomes difficult to organize and interpret the collected data through unstructured questionnaires.

#### 9.4.3. Principles of Preparing Questionnaire

- Questions/items should match the objectives of the lesson and purposes of the testing.
- Clear and comprehensible wording should be used to avoid confusion on the part of the students.
- Do not use leading or loading questions. The level of question should be appropriate to the mental level of students.
- Negative or double negative statements should be avoided while preparing questions like Do not you think that we should not protect our environment?
- Correct spelling, grammar and punctuation marks should be used in items
- The sequence of questions should be from simple to complex
- Similar types of questions should be grouped together as section A & B for proper organization.

#### Self-check Exercise-3.5

Write down five questionnaire items for assessing opinion of students towards health & wellbeing.

---

---

---

---

---

---

---

#### 9.4.4. Advantages of Questionnaire

- With minimum expense and effort, questionnaires can collect information about learning from many students in less time.
- With the help of questionnaires, both qualitative and quantitative data can be collected from the students about learning.
- By using advanced technology like google form & Kobo tool, questionnaires can be helpful to collect data with minimum time, and effort removing geographical barriers.
- Written questions give space and time to the students so that they feel free to express their views which is not possible through the interview method.
- There is less pressure on the students to respond as time limits are not imposed.

#### 9.4.5. Disadvantages of Questionnaire

- Questionnaires sometimes confuse students to understand the actual meaning of the questions. A rigid questionnaire does not draw accurate responses from participants.
- Sometimes the participants hide true answers and are dishonest while answering the questions.
- Written questionnaires cannot be administered to students with low literacy competency.

Questionnaires are a very important tool for assessing opinion, views, perceptions, values, attitudes from students. The social and personal traits of learners can easily be measured with the help of questionnaires. Hence, teachers must know the process of questionnaire development and use it for educational and research purposes.

---

### 9.5.SUMMARY/KEY POINTS

---

- **Rating scale** is a standardized method of recording behavior with which individuals can be rated on a scale from low to high with respect to a particular trait.
- **Like the name reflects**, there are numbers indicating the degree of present characteristics in a numerical rating scale. The rater needs to assign some value to a particular trait.
- **In graphic rating scale**, ratings are made in graphic form instead of using scale values. In this type of rating scale, each trait is presented in sentences followed by a horizontal line.
- **Questionnaire** is one of the most extensively used assessment tools in teaching, learning and research purposes. It consists of a set of questions/statements to gather information from respondents/students on a variety of problems, issues, and lessons. In order to gather firsthand information from the learners, questionnaires are a very effective tool.
- **In structured questionnaires**, pre-determined and definite questions are used for accurate communication and responses. Each question is followed by options and students are required to select one or more options.

- **Unstructured questionnaires** are those which include open and unrestricted questions and allow learners to express their thoughts in a free-flowing manner. Students are given freedom to share their ideas with depth.

---

#### 9.6.UNIT END EXERCISE

---

- Define the meaning of scale and its different categories.
- In which way scale is different from questionnaire. discuss its practical implications
- What is the meaning of questionnaire and its various categories.
- How does questionnaire prepare? Discuss its merits and demerits.
- What principles should be taken care of in the preparation of scales?

---

#### 9.7. SUGGESTION FOR FURTHER READING

---

Gronlund, N. E. (1965). *Measurement and evaluation in teaching*.  
<http://ci.nii.ac.jp/ncid/BA12623208>

Goswami, M. (2013). *Measurement and Evaluation in Psychology and Education*  
Lee, W. Y. (2010). *Assessment and evaluation in higher education*.  
<http://ci.nii.ac.jp/ncid/BB11596810>

Patel, R. N. (2014). *Educational evaluation theory and practice*.  
<http://ci.nii.ac.jp/ncid/BA29677030>

## UNIT - 10

### SCHEDULES, OBSERVATION, INTERVIEW, INTEREST INVENTORIES AND PERFORMANCE TEST

#### STRUCTURE

- 10.1. Learning Objectives
- 10.2. Introduction
- 10.3. Schedule
  - 10.3.1. Observation
  - 10.3.2. Structured observation
  - 10.3.3. Unstructured observation
  - 10.3.4. Participant observation
  - 10.3.5. Non-Participant Observation
  - 10.3.6. Advantages of observation
  - 10.3.7. Disadvantages of observation
- 10.4. Interview
  - 10.4.1. Uses of interview
  - 10.4.2. Disadvantages
- 10.5. Interest Inventories
- 10.6. Performance Test
  - 10.6.1. Characteristics of Performance Tests
  - 10.6.2. Advantages of Performance Test
  - 10.6.3. Disadvantages of Performance Test
- 10.7. Summary/Key Points
- 10.8. Unit End Exercise
- 10.9. Suggestions for Further Reading

---

#### 10.1 LEARNING OBJECTIVES

---

After reading this unit, learners shall be able to;

- Define schedules and its various categories used in educational assessment.
- distinguish between subjective and objective assessment tools.
- Explain the uses various types of objectives and subjective test

---

#### 10.5. INTRODUCTION

---

Evaluation methods are used to judge student learning and understanding of the material for purposes of grading and reporting. Tools and techniques of evaluation are critically examines a subject and then assigns a grade or some other type of formal result based on how well they performed. Here we are going to learn all the tools and techniques of evaluation which will helps us understand evaluation.

Evaluation methods are used to judge student learning and understanding of the material for purposes of grading and reporting. Tools and techniques of evaluation are critically examines a subject and then assigns a grade or some other type of formal result based on how well they performed. Here we are going to learn all the tools and techniques of evaluation which will helps us understand evaluation.

---

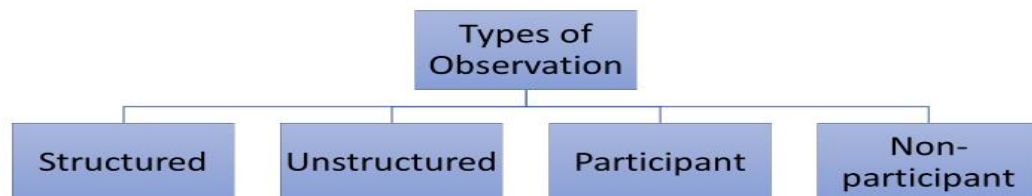
#### 10.3. SCHEDULE

---

Schedule is an assessment tool which is mainly used to guide interviews and observations for collecting data from the learners about learning. A well-structured schedule acts as a reminder which keeps the memory of the interviewer/observer/teacher refreshed and attentive of all the relevant aspects to be covered or observed within the allocated time. In fact, schedule systematizes and organized the observation and interviews. It is very much useful for assessing non-cognitive learning outcomes such as leadership, honesty, laboratory & library skills etc. Let us discuss about the observation schedule and interview schedule in following sections.

### 10.3.1. Observation

People do not always do what they usually say. So, to have the real know the real behaviour, observation as a tool plays the key role. The process of watching behavioral patterns of students to obtain information for understanding and assessment is calling as observational technique; especially to study the behavior of young children. Observation is considered as a very effective tool as no communication is needed for gathering information. Sense organs play a crucial role in the observation process. There is more use of eyes than ear and voice in the observation process. Observer/teacher only believes what he/she has observed with their own eyes. As the information is collected from primary sources through observation, it is reliable and valid in nature. According to C. A. Mourse, “Observation employs relatively more visual sense than audio or video organs.” Observation can be classified into following types based on the mode and nature of observation.



### 10.3.2. Structured observation

Structured observation is systematic, organized, and pre-planned. In case of structured observation, every detail of observation like qualities to be observed, time to observe, place of observation, criteria of observation, recording observation are pre-decided. The teacher/observer cannot change the plan of observation. To make observation structured, teacher must develop observation schedule and guide. The observation schedule is the list of items/ criteria to be observed during the observation. It is based on the learning outcomes to be assessed and level of students. In this method, the observer decides the answer to the following questions beforehand to make it effective:

- Who is the observer?
- What learning outcomes to be observed?
- In which time the observation must take place?
- How and where the observations are to be carried out?

### **10.3.3. Unstructured observation**

Unstructured observations are those where all relevant phenomena are observed and recorded extensively without planning or specifying in advance. This method is usually adopted for exploratory purposes or understanding the learners' inherent abilities or problems. Here, the teacher/observer has great role to play as observation scheduled was decided. Many learning outcomes such as adjustment, critical thinking, adoptability etc. can be successfully assessed through the unstructured observation.

### **10.3.4. Participant observation**

The role of teacher/ observer is paramount important in observation process. When the observer directly observes the activities and phenomena being a member of the group, is called as participant observation. Through this method, the natural behavior of the children can be studied in a holistic manner as the observer spends a great deal of time with the group. The skills in library, laboratory, acceptance, relationship, leadership etc. are suitably assessed through the participant observation.

### **10.3.5. Non-Participant Observation**

The limitation of participant observation is that students may hide the original behavior if came to know that they are being observed. So, teacher can use non-participant observation to assess the original and natural behavior of the learners. Here the observer experiences activities from outside and is not a part of the group. Non- participant observation is preferred when it is felt that an observer's presence may affect the natural behavior of the group.

### **10.3.6. Advantages of observation**

- One of the main advantages of observation is that through this the behavior of children can be observed in a natural setting as there is direct access to social phenomena.
- Information collected through observation schedules can provide both qualitative and quantitative information about the students in a holistic manner.
- Observation is helpful for collecting information from infants and people who are from other language backgrounds, as there is no need for communication.
- Personality is better assessed using this technique as there are less chances of hiding facts on the part of the students.

### **10.3.7. Disadvantages of observation**

- Observation as a tool of assessment is of time and resource consuming.
- Covert behavior of children cannot be studied through observation.
- There are chances of subjectivity or biasness on the part of the observer/teacher.

- This method is not useful to study the perception and opinion of students.

---

#### **10.4. INTERVIEW**

---

Interview is a form of oral communication between the interviewer (examiner) and interviewee (students) in which required information is directly obtained from the respondent verbally. An interview can be conducted in face-to face mode or over telephone or video conference. In this method, the interviewer elicits a response from the interviewee by asking questions related to subjects. It requires the active involvement of both for a productive conversation and result. Interview can be of structured or un-structured, individual or group. In case of structured interview, interview schedule and guide to be developed beforehand and interview to be conducted as planned. In case of un-structured interview, the teacher converse with learners and elicit the desired information. The following things are to be kept in mind while conducting an interview.

- Duration of the interview, place and mode of conducting should be decided beforehand.
- Questions in the interview schedule should be objective based/ outcome based.
- Interviewer/ teacher should be sociable in nature so that respondents will feel free to share their thoughts and ideas.
- Teacher must use appropriate language for conducting interview as understanding of questions is very important.
- Teacher must use positive body language while conducting interviews with students.
- Patient listening on the part of the interviewer is necessary for successful interview.
- No irrelevant discussions should take place at the time of the interview.
- Structured interview schedules should be framed to objectively assess the learners.

##### **10.4.1. Uses of interview**

- Teacher can use interview as a tool to assess many factual contents related to all school subjects.
- It is very much useful for foundation, preparatory and middle stage students as it can promote confidence in oral skills.
- Interview can be useful for identifying simple learning difficulties in pupils' behavior.
- While answering a questionnaire, sometimes the respondents are unable to answer properly due to understanding problems which do not happen in case of an interview as the doubts and misconceptions can be largely cleared.
- Interview questions elicit information about the attitude, perception, and perspectives of the respondents to know their personality.

##### **10.4.2. Disadvantages**

- During the interview student get less time for thinking to answer the questions. So, they get confused to answer correctly.
- Interview is time consuming and it is not possible to assess large number of students in a limited time frame.
- The quality of interview depends on the objectivity of the teacher/interviewer. Biasness of interviewer may affect the result.



- In some cases, respondents may not feel free to express their real thoughts in front of the other.
- It requires a high level of expertise from the teacher to use interview as assessment tool.

Interview is one of the unconventional tools of educational assessment which can be used by teachers across the school levels from foundation to secondary. The simple cognitive learning outcomes like remembering and understanding can be assessed through interview. The oral language skills like communication, confidence, presentation etc. can also be assessed by interview.

---

## 10.5. INTEREST INVENTORIES

---

Having knowledge about the interest areas of students is necessary for the teachers to design the teaching learning process accordingly or to guide them effectively for career. There are several methods or tools used in the assessment process to understand learners' interest, strengths, preferences, and potential areas of growth. Through informal methods like observation and direct questioning, teachers can note the things that the child likes to do, the type of videos she watches or the kind books she reads ; but these methods are usually restricted to school activities and we cannot compare an individual's interest with others. To overcome these limitations, standardized interest inventories have been designed for the purpose of measuring an individual's preferences, likes, dislikes across a range of activities and domains.

Unlike traditional assessment processes that focus on achievement tests and skills, inventories offer a more comprehensive understanding of learner's holistic development. Basically, in the educational settings, interest inventories are the structured items that guide students related to the areas like career counseling, academic choices, and personal development. Inventories also help them in decision making about career, future vocation, and various aspects of life by fostering self-awareness in them.

Inventory refers to the list of items related one area of learning or subjects or vocations followed by options in three- or five-point scales. It helps teacher in identifying students' choices, preferences and interest in scholastic, non-scholastic, and vocations. Examples of some of the popularly used interest inventories are: 'Kuder Preference Record-Vocational' (For high school students), 'Strong Vocational Interest Blank'(for high school and college level,) and 'What I Like to Do-An Inventory of children's interest'(for elementary level). Despite being a valuable tool to measure interest, inventory have certain limitations; as personal interests can evolve over time, and these inventories capture a snapshot of preferences at a particular moment. Additionally, cultural, and societal influences may shape individuals' responses. Therefore, interest inventories should be used in conjunction with other assessment methods to paint a more complete picture of an individual's capabilities and potential.

### Self-check Exercise: 3.7

Prepare five items for assessing students' interest towards curricular objectives.

---



---



---



---



---

---

**10.6.PERFORMANCE TEST**

---

Various techniques of assessment are used by the teachers to track the needs of students, their knowledge, and capabilities. Before choosing a relevant technique to assess students an examiner must determine the purpose of assessment and how the results are to be reported. Sometimes the teachers may wish to know more about students' understanding level or how they are able to reflect their skills in relation to the content taught rather than just what they know. In this context, one of the ways to accomplish this purpose is by using performance tests.

Performance tests require the students to perform an activity or a task, rather than just answering the asked questions. The students need to show or demonstrate their skills and competencies through their performance. In this kind of assessment, students are the active participants and they also have a chance to learn during the assessment process. Through this, students get motivation to learn more and increase their proficiency level. Performance test is concerned with how well the learners can apply their knowledge practically.

Some of the examples of performance tests are athletic competition, reading map, dance, drawing angles of certain measures, musical recital, identifying various coins or currencies, dramatic reading, wood work, demonstration in science laboratory and many more. Teachers are free to design activities for students for their performance test in accordance with the content.

**10.6.1. Characteristics of Performance Tests**

- Performance tests are both standardized and a non-standardized test and is flexible in nature.
- It can be used to measure the proficiency of students, motivate them and to provide them necessary feedback.
- Performance test does not require one right answer like typical tests rather indicates the degree of skilling.
- It promotes independent thinking, problem solving skills, reasoning power and knowledge construction.
- Teachers can evaluate by observing the activities of learners and scoring can be done with the help of rubrics.
- It focuses on both the product and process of learning. The required instruments or materials are prepared beforehand by the teacher and teachers have the liberty to design activities and programs relevant to the content taught.

**10.6.2. Advantages of Performance Test**

- Through performance tests teachers can observe students' performance directly in a real-world setting.
- Performance test provides desired information which cannot be extracted through verbal or written type traditional settings.

- Assessment process becomes interesting, promoting active involvement of students.
- This kind of assessment is very useful for young children, mentally retarded students or the students who have language disabilities.
- During the participation of the performance test, the teacher can study the student's personality by observing his/her gesture, posture, expression, and attitude.
- Any kind of cheating or malpractices like copying is not possible in performance test.
- Through a single activity, multiple concepts and objectives can be measured.

#### **10.06.3. Disadvantages of Performance Test**

- Assessment process requires comparatively more time than the assessment in a traditional setting.
- There are chances of subjectivity or biasness; the observer's perception, attitude, beliefs may affect the results.
- Performance test requires various materials or equipment which may be costly or not available everywhere.
- It may not be appropriate for every type of content, focusing only on certain skills or abilities.
- Teachers need to be in active state throughout the performance test to observe minute details.

Skills are the integral part of the schooling process which needs to be assessed suitably. Skills consists of micro skills which are observable and measurable. Skills can only be assessed through observation and performance test. The laboratory skills, communication skills, writing skills, dramatic skills etc. can be effectively assessed through the performance test.

---

### **3.12. SUMMARY**

---

- Schedule is an assessment tool which is mainly used to guide interviews and observations for collecting data from the learners about learning.
- A well-structured schedule acts as a reminder which keeps the memory of the interviewer/observer/teacher refreshed and attentive of all the relevant aspects to be covered or observed within the allocated time.
- The process of watching behavioural patterns of students to obtain information for understanding and assessment is calling as observational technique.
- Structured observation is systematic, organized, and pre-planned. In case of structured observation, every detail of observation like qualities to be observed, time to observe, place of observation, criteria of observation, recording observation are pre-decided.
- Unstructured observations are those where all relevant phenomena are observed and recorded extensively without planning or specifying in advance. This method is usually adopted for exploratory purposes or understanding the learners' inherent abilities or problems.
- The role of teacher/ observer is paramount important in observation process. When the observer directly observes the activities and phenomena being a member of the group, is called as participant observation

- The limitation of participant observation is that students may hide the original behaviour if came to know that they are being observed. So, teacher can use non-participant observation to assess the original and natural behaviour of the learners.
- Interview is a form of oral communication between the interviewer (examiner) and interviewee (students) in which required information is directly obtained from the respondent verbally.
- Inventory refers to the list of items related one area of learning or subjects or vocations followed by options in three- or five-point scales. It helps teacher in identifying students' choices, preferences and interest in scholastic, non-scholastic, and vocations.
- Performance tests require the students to perform an activity or a task, rather than just answering the asked questions. The students need to show or demonstrate their skills and competencies through their performance.

---

### 3.13. UNIT END EXERCISE

---

- Define schedules, its various categories, advantages and disadvantages?
- What is performance test? Discuss its purposes, merits and demerits?
- Discuss the meaning, nature, advantages and disadvantages of interview?
- Differentiate between schedule and interview ? discuss its uses critically?

---

### 3.14. SUGGESTION FOR FURTHER READINGS

---

Gronlund, N. E. (1965). *Measurement and evaluation in teaching*.  
<http://ci.nii.ac.jp/ncid/BA12623208>

Goswami, M. (2013). *Measurement and Evaluation in Psychology and Education*  
Lee, W. Y. (2010). *Assessment and evaluation in higher education*.  
<http://ci.nii.ac.jp/ncid/BB11596810>

Patel, R. N. (2014). *Educational evaluation theory and practice*.  
<http://ci.nii.ac.jp/ncid/BA29677030>

---

## **BLOCK 03:**

### **TOOLS AND TECHNIQUES OF ASSESSMENT AND CONSTRUCTION OF TEST**

---

**Unit 11:** Steps of test construction: planning, preparing, trying out and evaluation

**Unit 12:** Principles of construction of objective type test items, matching, multiple choice, completion and true – false

**Unit 13:** Principles of construction of essay type test

**Unit 14:** Non- standardized tools: Observation schedule, interview schedule, , check list,

**Unit 15:** Non- standardized tools: portfolio and rubrics, rating scale

## UNIT -11

### GENERAL PRINCIPLES OF TEST CONSTRUCTION AND ITS STANDARDIZATION

#### STRUCTURE

11.1. Learning Objectives

11.2. Introduction

11.3. Concept of General Principles of Test Construction and its standardisation

11.3.1. Planning the Test

11.3.2. Development of Test Blueprint

11.3.3. Preparing the Test

11.3.4. Trying-out the Test

11.3.5. Evaluating the Test

11.3.6. Preparation of the Test Manual

11.4. Summary/Key Points

11.5. Unit End Exercise

11.6. Suggestion for Further Reading

---

#### 11.1. LEARNING OBJECTIVES

---

After reading this unit, learners shall be able to;

- Define the general principles of Test Construction and its standardisation.
- Explain the concepts of planning the test, test blueprint, preparing the test, trying out the test, evaluating the test
- Distinguish between trying out the test and evaluating the test

---

#### 11.2 INTRODUCTION

---

Attention must be given to the below mentioned points while constructing a potent, constructive and relevant questionnaire/schedule:

• The researcher must first define the problem that s/he wants to examine, as it will lay the foundation of the questionnaire. There must be a complete clarity about the various facets of the research problem that will be encountered as the research progresses.

• The correct formulation of questions is dependent on the kind of information the researcher seeks, the objective of analysis and the respondents of the schedule/questionnaire. Whether to use open ended or close ended questions should be decided by the researcher.

They should be uncomplicated and made with such a view that there will be an objective part of a calculated tabulation plan.

- A researcher must prepare a rough draft of the schedule while giving ample thought to the sequence in which s/he wants to place the questions. Previous examples of such questionnaires can also be observed at this stage.
- A researcher by default should recheck and if required make changes in the rough draft to improve the same. Technical discrepancies should be examined in detail and changed accordingly.
- There should be a pre-testing done through a pilot study and changes should be made to the questionnaire if required.
- The questions should be easy to understand the directions to fill up the questionnaire clearly mentioned; this should be done to avoid any confusion.

The primary objective of developing a tool is obtaining a set of data that is accurate, trustworthy and authentic so as to enable the researcher in gauging the current situation correctly and reaching conclusions that can provide executable suggestions. But, no tool is absolutely accurate and valid, thus, it should carry a declaration that clearly mentions its reliability and validity.

Standardization refers to the consistency of processes and procedures that are used for conducting and scoring of a test. To compare the scores of different individuals the conditions should be the same. In case of a new step the first and major step in standardization is formulating the 122 directions. This also includes the type of materials to be used, verbal instructions, time to be taken, the way to handle questions by test takers and all other minute details of a testing environment. Establishing the norms is also a key step for standardization. Norm refers to the average performance. To standardize a test, we administer it to a big, representative sample of the kind of individuals it was designed for. The aforementioned group sets the norms and is called the standardization sample. The norms for personality tests are set in the same way as those set for aptitude tests. For both, the norm would refer to the performance of average individuals. To construct and administer a test, standardization is a very important. The test is administered on a large set number of the people (the conditions and guidelines need to be the same for all). After which the scores are modified using Percentile rank, Z-score, T-score and Stanine, etc. The standardization of a

test can be established from this modified score. Hence, “standardization is a process of ensuring that a test is standardized, (Osadebe, 2001)”. There are lots of advantages when a test is standardized. A standard test is usually produced by experts and it is better than teacher made test. The standardized test is highly valid, reliable and normalized with Percentile rank, Z-score, T-score among scores derived from others to produce age norm, sex norm, location norm and school-type norm. Generally, a standardized test could be used to assess, and compare students in the same norming group. The normal process for administering standardization includes: 1) A calm, quiet and disturbance free setting 2) Accurately understanding the written instructions, and 3) Provisioning of required stimuli. This makes the normative data applicable to the individuals being evaluated.

---

### **11.3. GENERAL PRINCIPLES OF TEST CONSTRUCTION AND ITS STANDARDIZATION**

---

An instrument or procedure used to gauge an individual's knowledge, skills, abilities, or other traits is called a test. In essence, it is a methodical approach to learning about someone's abilities or knowledge. Tests are designed to measure a variety of things, including job performance, personality traits, aptitude, academic achievement, and personality traits. It can be non-standardized, suited to circumstances or people, or standardized, with uniform scoring and administration protocols. To ensure accurate and significant results, tests must be carefully constructed and used, considering factors like validity, reliability, and fairness. There are various types of tests, each serving a different purpose in the assessment process. Achievement tests measure what a person has learned or accomplished in a specific area whereas aptitude tests assess a person's potential to develop certain skills or abilities. Likewise, personality tests aim to evaluate aspects of an individual's personality, such as traits, behaviours, or preferences and diagnostic tests help identify strengths and weaknesses in a person's skills or knowledge. Along with these, there are two widely used approaches of tests such as subjective and objective tests. Subjective tests provide an in-depth understanding of a person's knowledge and critical thinking skills whereas objective tests have a specific answer or set of answers focusing on the memorization aspect. When we talk about construction of tests, they are either constructed by teachers or by experts. Classroom tests that are constructed by teachers are teacher-made tests whereas standardized tests are developed by the experts in the field. All the types of tests help to make informed decisions about individuals' abilities and characteristics. It is important to choose the right type of test based on the goals of the assessment and the information needed.



It may not be always possible for the examiner to get a ready-made test for the needed purpose so new tests are developed and designed by the experts or teachers. The development and utilization of tests are integral components of the fields of psychology, education, and various other disciplines. Test construction is one of the most important aspects of the assessment and measurement process as the effectiveness of the test design determines the accuracy of the result to be obtained. Tests serve as valuable tools for assessing individuals' knowledge, skills, abilities, and other characteristics. However, to ensure that test results are valid, reliable, and fair, it is essential to adhere to general principles of test construction and standardization. Construction of a test and standardization is a very systematic process which includes several steps as follows.

### **11.3.1. Planning the Test**

A skill-full planning results in better test construction. This is the first step in the test construction process. In this step, overall planning of framework is done related to the test; like which content is to be measured, what will be the format for the questions, definition of construct to be measured, nature of the population for whom the test is being constructed, language of the test and how the test will be administered and scored etc. Objectives or purpose of the test, time and cost should also be taken into consideration by the examiner. After deciding the above points, the teacher needs to develop a Test Blueprint or Table of Specification. It is a table that describes the weight ages given to different content, objectives, and types of items. It helps teachers to prepare a balanced test-giving proper weight age to content and objectives. Otherwise sometimes the teacher may set more questions from one unit than other, and set more questions from the knowledge domain than understanding & application. It is the framework that guides the test constructor for the preparation of questions.

The constructor of the test must consider the following sequence of steps while planning the test.

- Determine the purpose for which the test is being constructed.
- Find out the characteristics of the group for whom the test is intended.
- Defining the objectives or translating the purpose of test in operational terms
- Specifying the content to be covered along with the skills such as knowledge, understanding, application etc.

- Specification of the test format or the blueprint preparation showing weight age to the list of content areas, instructional objectives, and types of test items.

### 11.3.2. Development of Test Blueprint

One of the most important components in the test construction process is the Blueprint. It is also known as test specification or test plan. Blueprint is a well-designed three-dimensional chart or framework which includes the format of the test and the weightage of the topics that are to be covered. Weightage given to the content to be assessed, instructional objectives and types of items should be decided before. Blueprints should be prepared well in advance before the test is constructed to assist the teacher in organising her teaching-learning plan and to be on track. It serves as a guide for test developers to ensure that the test aligns with its intended purpose and objectives. Where feasible, students should be encouraged to contribute or suggest to the process of blueprint development; this way they will feel involved in planning and curious about the further assessment process. By contributing to the planning step, students will develop skills like critical thinking, decision making; however, the final decisions are to be made by the teacher. Sample of a three-dimensional blueprint is given below.

**Table 4.1: Test Blueprint**

Objectives	Knowledge	Understanding	Application	Skill	Total
Content	E S O	E S O	E S O	E S O	
concept of Cell	1 - -	- - 2	- - 2	- - -	5
Animal Cell	1 - -	- 1 -	- 2 1	- - -	5
Plant Cell	1 - -	1 1 -	- 2 -	- - -	5
Cell Organelles	1 - -	- - 1	1 1 1	- - 1	6
Function of cell organelles	1 - -	1 1 -	- 1 -	- - -	4
Cell Division	- - -	2 - 4	- 2 4	- - 3	15
Cell Diagram	- - -	- - 1	- 2 1	- 2 4	10

Total	5    -    -	4    3    8	1    10 9	-    2    8	50
	5	15	20	10	
E- Essay type		S- Short-answer type	O- Objective type		

### 11.3.3. Preparing the Test

After planning, the test is prepared based on the blueprint or framework. This step is all about writing test items. Different test items are written by the test constructor keeping in mind the format of the test, content area to be covered, expected learning outcome and most importantly, the objectives. In the first draft of the test, a greater number of items are kept since many of them can get eliminated while trying out and revising. Test developers need to write instructions to students about how to answer and where to answer; along with a model answer script for short answer and essay type items and answer key for objective type. While writing the test items, the researcher or test constructor should refer to existing standardized tests in the concerned area for reliability checking. First editing and review is done by the test constructor and then checked by the experts where inappropriate and ambiguous items are corrected or removed. The items which contribute for the attainment of the test objectives are kept only.

At the time of developing the test items, following rules and regulations should be kept in mind;

- Preliminary draft of the test should have more than double the items required for the test.
- Items should be written as per the test blueprint.
- Each item must be based on a specific learning outcome.
- Items should contain proper language and vocabulary at the level of the learners.
- Comprehensiveness and adequacy of the test should be maintained.
- Test items should be attainable with in time allotted
- Difficulty level of items should be taken care of (majority of items must be at average level)
- Objectivity should be ensured in the time of test preparation
- Clues or hints in the questions to help examinees answer should be avoided.

### 11.3.4. Trying-out the Test

After the prepared items are checked and tested thoroughly, trying out is done on a small representative sample to select the best items. The purpose of this step is to identify defective, complex, and ambiguous test items. It helps to know how students react to the items and to

determine the difficulty level and discriminating power of the test item. Ideal test items reflect normality in the result.

Trying out is done in three phases. First phase is pre-try-out or individual try-out. In this phase, the test constructor checks the grammatical mistakes, ambiguity in language and complexity of the items from the result of the achievement test. Second phase of piloting is group try out or proper try out. In a group try out, the size of sample should be large and representative of the population concerned and time limit should be generous. Answer sheets are then scored with the help of a pre-designed scoring key and item-analysis is done. Lastly, in the final try-out, the final draft is prepared by revising, reviewing, editing, and eliminating multiple test items again and again. During pilot testing, the test constructor must keep the following points in mind.

- The environmental condition should be free from distraction or noise
- Establishing rapport with the subjects is an important aspect while administering the test
- The mental condition of the subjects should be taken into consideration
- Sitting facility and space should be provided and the group should not be too large.

#### **11.3.5. Evaluating the Test**

Evaluating the test refers to making judgement about the individual test items and test. This step is mostly statistical in nature. In this step, reliability determination is done after administering the test on a new sample and calculating the reliability coefficient. After determining the reliability, the test constructor compares the test with some external criteria from other independent data and determines the validity of the test. Karl Pearson's coefficient correlation and point- biserial coefficient correlation are some useful methods to find out the validity coefficient. This step helps in the finalisation of test items and development of norms for interpretation.

#### **11.3.6. Preparation of the Test Manual**

The last step of the test construction is preparation of the test manual. Test manual contains the brief details of the test like number of items, target group, technical quality, norms etc. Test constructor describes the psychometric properties of the test in this manual and lists the references and norms. The test developer describes the details of the administration process, way of scoring and time limitations of the test.

**Self-check Exercise-4.1**

What is a test blueprint and how does it help in test development?

---

---

---

---

---



---

**11.7. SUMMARY/KEY POINTS**

- Standardization refers to the consistency of processes and procedures that are used for conducting and scoring of a test
- Generally, a standardized test could be used to assess, and compare students in the same norming group.
- Blueprint is a well-designed three-dimensional chart or framework which includes the format of the test and the weightage of the topics that are to be covered. Weightage given to the content to be assessed, instructional objectives and types of items should be decided before.
- Evaluating the test refers to making judgement about the individual test items and test.
- Karl Pearson's coefficient correlation and point- biserial coefficient correlation are some useful methods to find out the validity coefficient.
- Test manual contains the brief details of the test like number of items, target group, technical quality, norms etc.
- Test constructor describes the psychometric properties of the test in this manual and lists the references and norms.
- After the prepared items are checked and tested thoroughly, trying out is done on a small representative sample to select the best items. The purpose of this step is to identify defective, complex, and ambiguous test items.

---

**11.8. UNIT END EXERCISE**

- What are the general principles of test construction and its standardisation?
- What are the processes of standardisation? Explain it
- How does the test Blueprint develop? Explain it
- How does test manual develops?

---

**11.9. SUGGESTION FOR FURTHER READING**

Gronlund, N. E. (1965). *Measurement and evaluation in teaching*.  
<http://ci.nii.ac.jp/ncid/BA12623208>

Goswami, M. (2013). *Measurement and Evaluation in Psychology and Education*  
 Lee, W. Y. (2010). *Assessment and evaluation in higher education*.  
<http://ci.nii.ac.jp/ncid/BB11596810>

Patel, R. N. (2014). *Educational evaluation theory and practice*.  
<http://ci.nii.ac.jp/ncid/BA29677030>

## **UNIT -12**

### **WRITING TEST ITEMS**

#### **STRUCTURE**

12.1. Learning Objectives

12.2. Introduction

12.3. Writing Test Items

12.3.1. Objective Type Test Items

12.3.2. Supply type test items

12.3.2. True & false type of test items

12.3.3. Selection Type Test Items

12.3.4. Matching type test items

12.3.5. Multiple-choice test items

12.3.6. Essay Type Test Items

12.3.7. Extended response type

12.3.8. Restricted response type

12.3.9. Interpretive type test items

12.3.10. Selection of the most appropriate item types

12.4. Summary

12.5. Unit End Exercise

12.6. Suggestion for Further Reading

---

#### **12.1 LEARNING OBJECTIVES**

---

After reading this unit, learners shall be able to;

- Explain the uses of objective test and its various types.
- Explain the uses of essay type of test and its various types.
- Differentiate the specific uses of objective and essay type test

---

#### **12.2. INTRODUCTION**

---

The test items used for assessment are typically divided into three broad categories: i.e., the objective type, essay type and interpretive type. Objective type is highly structured and asks students to provide one or two words or choose the best response from a small number of options. Essay questions allow students to choose, organize, and present their response in the form of an essay. Interpretive test items require students to interpret a given stimulus

(table/graph/paragraph) to respond to the question. Each type should be utilized when it makes the most sense, with appropriateness being determined by the learning outcomes to be measured as well as by the distinct benefits and constraints of each item type.

---

### **12.3. WRITING TEST ITEMS**

---

#### **12.3.1. Objective Type Test Items**

One of the widely used test items in recent times is the objective type of test. In an objective type of test, students are required to write answers either from options or supply in words. It ensures the objectivity in items and scoring for which it is popular. The objective test items are of several types. It can be broadly classified into *supply type and selection type*. Supply type test items are those that require the students to supply the answer in word phrase or number whereas selection types require them to select/choose the answer from a certain number of alternatives given. These two general categories are further divided into the following types commonly:

#### **12.3.2. Supply type test items:**

This is an objective type test in which students are required to provide answers in words/sentences to a direct question or complete the sentence by filling the blank with a suitable answer. It demands students to recall the answer from their memory and supply as per the question. This kind of test is mainly used to assess lower-level cognitive capacities like memory. There are two kinds of supply type test; short answer and completion type.

*Short answer type: Example of item*

Answer the following questions in one word.

- a. Which is the highest peak in the world?
- b. What is the capital of Japan?
- c. Who is the father of biology?
- d. What is the chemical formula for sulfuric acid?

*Completion type: Example of item*

Complete the sentences by using suitable answers.

- a. The animals who can live both in water and land are called\_\_\_\_\_.
- b. A person who studies about birds is called\_\_\_\_\_.
- c. Child moving on a swing is an example of \_\_\_\_\_ motion.
- d. The largest organ of the body is\_\_\_\_\_.

#### **12.3.3. Suggestions for constructing supply type test items**

- Words of the question should be written in such a way that the answer will be in one or two words or numbers.
- Word questions in such a manner so that it can be similarly interpreted by all students taking the test.
- More than two blanks in a single item should be avoided which may create confusion or ambiguity.
- Blanks near the end of the sentence are preferable which gives proper sense of the question to the reader.
- Uniformity in length of blanks maintained otherwise it might misguide students towards irrelevant clues about answers.

**Self-check Exercise-4.2**

Write three short answer types and three completion type test items from any school subject.

---

---

---

---

---

---

---

**12.3.4. Selection Type Test Items**

The supply type test items cannot assess the complex learning outcomes like understanding, applying, evaluating, and creating. Hence, selection type test items are required for assessing such complex learning outcomes. Here, students are provided both questions and possible answers in form of alternatives, asked to select or recognise the most appropriate answer. These types of test items are useful to assess learning competencies and learning outcomes in all subject areas and at all levels of school education. There are a variety of selection type test items such as true & false, matching, and multiple-choice items.

**12.3.5. True & false type of test items**

The true & false type test items are used to assess knowledge of facts, dates, rules, and principles. Here, students are provided a meaningful sentence and asked to mark true or false. The example like;

***Put a tick mark on the True/False at the end of the sentence.***

- a. Diarrhea patients should be advised to drink ORS. (T/F)
- b. We should promote deforestation to reduce global warming (T/F)
- c. Rice plant has tap root (T/F)
- d. We can separate butter from cream through filtration method (T/F)

**Suggestions for constructing true and false items**

- Use simple clear language to avoid confusion
- Approximately equal number of true and false items should be kept in the test.
- Using a mix of true and false statements in a random manner is advisable to prevent the unintentional disclosure of irrelevant information.
- The items measuring important learning outcomes and content only should be included.
- Try to avoid double negative statements which create confusion among students.

**12.3.6. Matching type test items**

One of the important types of test items is matching type which can assess remembering, understanding, and applying level of learning outcomes. Here, students are provided two columns; column-A and column-B and students are asked to match the columns. The content that demands association between one with another like name of book and authors, scientific invention and scientist etc. can be assessed through matching type items. The example like;



*Match column 'A' with column 'B'*

Column 'A'	Column 'B'
a) Climbing birds	Penguin
b) Perching birds	Hen
c) Scratching birds	Woodpecker
d) Swimming birds	Pigeon
e) Wedding birds	Crane
	Vulture

#### **Suggestions for constructing Matching test items**

- Ensure that each matching test comprises items that are focused on a single concept.
- Avoid direct one-to-one matches and include extra response options to encourage thoughtful consideration for the final match.
- Organize the responses systematically, such as arranging words alphabetically or ordering numbers in ascending or descending sequences.
- Keep all items on the same page to streamline the process, save time, and prevent any potential confusion for the test-taker.

#### **12.3.7. Multiple-choice test items**

Multiple choice test items are widely used in educational and recruitment settings because of its ability to measure all types of learning outcomes in all subjects. The complex learning outcomes like applying, analysing, evaluating, creating can also be measured through multiple choice test items. It consists of two parts; stem and alternatives. Stem can be a question or incomplete sentence which presents a meaningful problem. Alternatives are the possible answers out of which one is correct (Key) and others are distractors.

***Select the most appropriate answer from the alternatives provided. Tick the correct option.***

- The gas whose amount varies in weather changes is \_\_\_\_\_.
  - Carbon dioxide
  - Oxygen
  - Helium
  - Water vapor
- Which fuel is used for generating electricity in thermal power plants?
  - Water
  - Wind
  - Diesel
  - Coal
- Pole star is found in which hemisphere in the night sky?
  - Northern
  - Southern
  - Eastern
  - Western
- Which of the following is not a part of the digestive system?
  - Lungs
  - Liver

iii. Stomach

iv. Pancreas

### **Guidelines for constructing multiple choice items**

- Refrain from using overly complex distractions and unfamiliar terminology or symbols.
- Ensure that all possible responses are reasonable and consistent. Students should not easily dismiss any distractor as irrelevant or nonsensical.
- Maintain consistency in the length of correct answers and distractors.
- Keep multiple-choice items clear and directly related to the instructional objective, avoiding any suggestive wording.

All types of objective test items share a common characteristic that sets them apart from essay tests. They present students with a tightly structured task that confines the range of responses they can provide. To arrive at the correct answer, students must demonstrate the specific knowledge, comprehension, or skill requested by the item. They are not permitted to redefine the problem or to arrange and express their response in their own words. Instead, they must choose one from a set of possible answers or provide the accurate word, number, or symbol. This structured approach to problem-solving and the restriction on response methods contribute to a scoring process for objective tests that is swift, straightforward, and precise.

#### ***Self-check Exercise-4.3***

Write four multiple choice test items to assess applying learning outcomes.

---

---

---

---

---

---

---

---

---

---

However, on the downside, this same structure renders objective test items unsuitable for assessing a student's ability to select, organize, integrate ideas, and engage in independent thought processes. To evaluate such abilities, we must rely on essay questions.

### **12.3.8. Essay Type Test Items**

An essay test is a form of written assessment that requires the test-taker to compose a sentence, paragraph, or longer composition. It also entails a subjective evaluation of the quality and comprehensiveness of the response during scoring. Unlike other types of tests that primarily focus on identifying, interpreting, and applying data, the essay test is designed to measure more complex learning outcomes. These outcomes include the ability to effectively organize ideas, integrate concepts, express oneself in writing, and provide detailed information. One distinguishing feature of essay tests is the freedom they afford to test-takers in constructing their responses. Students have the liberty to choose, connect, and articulate their ideas in their own words. Depending on the extent of this freedom in organizing

thoughts and crafting responses, essay questions are generally categorized into two main groups:

### **Extended response type**

This type of essay item gives maximum liberty to students to organise their thoughts and express their own ideas in their own preferred way. There are no constraints or boundaries to limit the thought or words within and students can pen down or elaborate their ideas freely. To assess the creativity level and diverse thinking abilities of students, this is the most relevant method. The example is;

- Discuss the foreign policy of Manmohan Singh Government and Narendra Modi Government.
- Explain different factors that cause a threat to national integration in India.

### **Restricted response type**

The restricted response type imposes specific limitations on how students must structure their answers in a proper direction. These limitations can pertain to the format, length in words, or organization of the response, constraining the students' freedom to some degree. It helps the students to think in a systematic way and scoring becomes easy. The context is pre-specified that limits the answer minimising the freedom to express the own diverse thoughts of students. The examples are;

- Explain, in not more than 200 words, the chemical changes take place in our everyday life.
- Write an essay within 500 words on the rainy season highlighting its importance, scenario in villages, advantages, and disadvantages.

### **Guidelines for writing essay test**

- Take sufficient time and careful consideration when formulating questions, enabling the opportunity for revision, and editing before their application.
- A properly structured essay prompt should offer clear guidance to students, encouraging the desired type of response.
- Clearly specify the relative importance of different components within a question, allowing students to determine where to allocate focus during their writing.
- Provide an ample time limit to prevent the essay test from becoming a test of writing speed rather than an assessment of knowledge and skills.
- Indicate the anticipated answer length for each question within the question itself.

#### **Self-check Exercise-4.4**

Write two extended essays and two restricted essays from any school subjects.

---

---

---

---

---

---

---

---

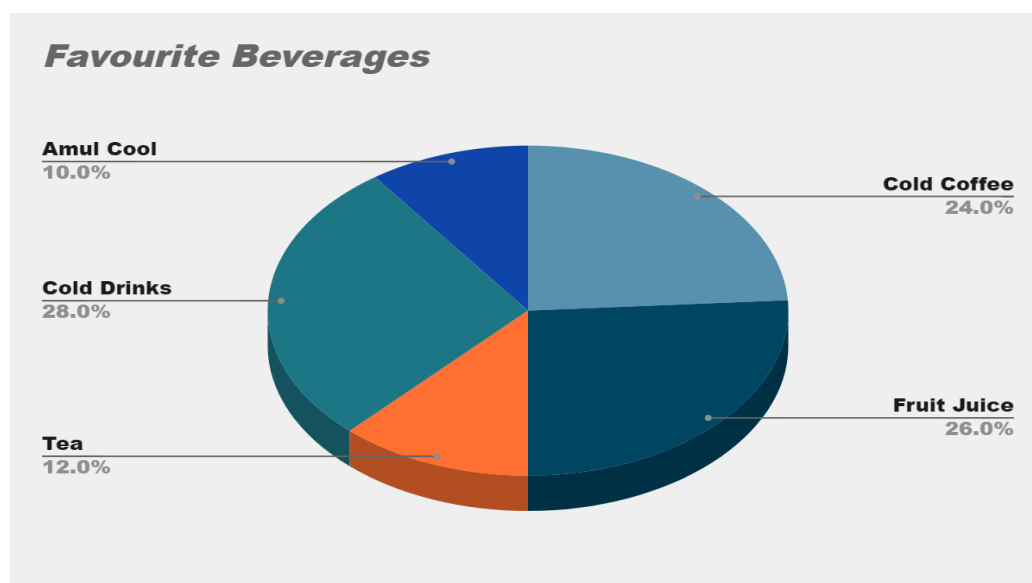
### 12.3.9. Interpretive type test items

The use of interpretive test items is a great way to assess a variety of cognitive abilities using the same assessment method. This type of test requires students to interpret a given item and respond according to the instructions; This can be accomplished in a variety of ways like using a graph, table, or other data set, reading a brief passage, and responding to the related questions. Using these items for assessment helps teachers find out how well students understood a particular concept and identify where they need more assistance. Students can evaluate, compare, solve problems, and much more in one set! With the help of these test items, critical thinking can be assessed and applied very easily. Because they are not just reading and recalling questions, Students must actively evaluate the information they are given, think back on prior understanding, and then form their answers from that.

This kind of material is adaptable to different grade levels and can be used in any subject area. For instance, a first-grade interpretive item may require students to look at a picture of a group of people and respond to questions about what they are doing, whereas a fifth-grade interpretive item may require students to read a brief passage and respond to questions about the main idea. Since they require students to do more than just recall facts, interpretive items are an effective assessment tool. To respond to the question, they must be able to comprehend and interpret the data. It is crucial to give students scaffolding, such as a list of questions to answer or a rubric, to help them learn. This will make it more likely that all students will finish the assignment and provide accurate responses.

For example, an interpretive item that might be used in 5<sup>th</sup> grade for math class is using a pie chart showing data on the percentage of boys and girls interested in different games. Students would be asked to answer related questions and interpret the data.

*In summer, a survey was conducted among 600 people to know their favourite beverage. The collected data is arranged in a pie chart.*



*Answer the following questions based on the above pie chart.*

- i) How Many people like to have cold coffee in summer?
- ii) Find the difference between the number of people who drink fruit juice and Amul cool?
- iii) Which of the above drinks is highest preferable among people as per the chart ?
- iv) How many people like cold drinks more than tea?
- v) Which is the lowest selling beverage according to the above pie chart?

In spite of being able to measure higher order thinking skills, there are certain limitations of interpretive test items. They are difficult to prepare and require more time for administering. Sometimes the big paragraphs have heavy demands on the reading skills of students.

#### **12.3.10. Selection of the most appropriate item types**

Different types of test items have varying degrees of effectiveness in measuring different types of learning outcomes. For instance, short-answer questions are well-suited for evaluating the recall of specific facts but may not be suitable for assessing comprehension, application, interpretation, or other intricate learning objectives. True-false or alternative-response items are most valuable when the goal is to determine the accuracy of a statement, differentiate between fact and opinion, or discern appropriate from inappropriate responses. However, like short-answer questions, they may not adequately gauge more complex learning outcomes.

Matching items share a similar limitation, as they are primarily geared towards assessing learning outcomes involving the identification of straightforward relationships and the classification of items into predefined categories. On the other hand, multiple-choice questions are highly versatile and can effectively measure a wide range of learning outcomes, from basic to more intricate ones. Nonetheless, even multiple-choice items have their limitations and may not be efficient for assessing the most complex learning objectives. Essay questions and other open-ended formats are particularly effective for evaluating skills such as data organization, the presentation of original ideas, and certain types of problem-solving scenarios.

#### **12.4. SUMMARY**

- In an objective type of test, students are required to write answers either from options or supply in words. It ensures the objectivity in items and scoring for which it is popular
- The objective test items are of several types. It can be broadly classified into *supply type and selection type*
- Multiple choice test items are widely used in educational and recruitment settings because of its ability to measure all types of learning outcomes in all subjects.
- The complex learning outcomes like applying, analysing, evaluating, creating can also be measured through multiple choice test items. It consists of two parts; stem and alternatives.
- An essay test is a form of written assessment that requires the test-taker to compose a sentence, paragraph, or longer composition. It also entails a subjective evaluation of the quality and comprehensiveness of the response during scoring.

- The use of interpretive test items is a great way to assess a variety of cognitive abilities using the same assessment method. This type of test requires students to interpret a given item and respond according to the instructions
- Validity of a test refers to the extent to which assessment test results serve the particular purpose for which it was intended.
- It refers to the extent to which a tool appears to measure what it claims to measure. Face validity is not related to what the test measures.

### 12.5. UNIT END EXERCISE

- Define the objective test? discuss its various types,
- Define the Essay type of Test? Discuss its various types and its significance.
- What are the basic characteristics of good measuring instrument? Elaborate
- What is the validity? Explain the different factors affecting validity?
- Define reliability? Explain the particular factors affecting reliability?
- Distinguish the difference between reliability and validity?

### 12.6. SUGGESTION FOR FURTHER READING

Gronlund, N. E. (1965). *Measurement and evaluation in teaching*.  
<http://ci.nii.ac.jp/ncid/BA12623208>

Goswami, M. (2013). *Measurement and Evaluation in Psychology and Education*  
Lee, W. Y. (2010). *Assessment and evaluation in higher education*.  
<http://ci.nii.ac.jp/ncid/BB11596810>

Patel, R. N. (2014). *Educational evaluation theory and practice*.  
<http://ci.nii.ac.jp/ncid/BA29677030>

# **UNIT 13**

## **GOOD MEASURING INSTRUMENTS**

### **STRUCTURE**

13.1. Learning Objectives

13.2. Introduction

13.3. Basic Characteristics of Good Measuring Instrument

13.3.1. Validity

13.3.2. Types of Validity: Face validity, Content validity, Criterion validity, Construct validity

13.3.3. Factors affecting validity of the test

- Difficult sentence construction and reading vocabulary:
- Unclear direction:
- Difficulty level of test items:
- Extraneous factors:
- Inadequate coverage:
- Ambiguity:

13.3.4. Reliability

13.3.5. Types of reliability (Test-retest reliability, Equivalent or parallel forms of reliability, Split half method of reliability, Kuder-Richardson reliability, Inter-rater reliability)

13.3.6. Factors affecting reliability of the test - Subjectivity in scoring, Ambiguous wording of the test items, Inconsistency in test administration, Length of the test, Difficulty level of the test items, Optional questions:

13.3.7. Relation between reliability and validity

13.3.8. Objectivity

13.3.9. Usability

13.3.10. Norms

13.3.11. Types of norms

13.3.12. Standard score norms

13.3.13. Z-Score

13.3.14. T-Score

13.3.15. Stanine Scores

13.3.16. C-Scale

13.4. Summary

13.5. Unit End Exercise

13.6. Suggestion for Further Reading

---

**13.1. Learning Objectives**

---

After reading this unit, learners shall be able to;

- Explain the basic characteristics of good measuring instruments.
- Elaborate the validity, its types of validity and its different factors
- Express the reliability, its different types of reliability and its factors.
- Differentiate between reliability and validity

---

**13.2. INTRODUCTION**

---

An educational test is not just that measures achievement in subjects of study, but it is also a psychological test that leads to an assessment of the overall development of a student. According to Anastasi, psychological tests are essentially an objective and standardised measure of a sample of behaviours. For Freeman it is a standardized instrument designed to measure objectively one or more aspects of a total personality by means of a sample of verbal or non-verbal responses or by means of other behaviours.

Test is a stimulus selected and originated to elicit responses which can reveal certain psychological traits in the person who deals with them. The diagnosis or redictive value of psychological traits in the person who deals with them. The diagnosis or predictive value of a psychological test depends upon the degree to which it serves as an indicator of a relatively broad and significant area of response. It is obvious that psychological test is the quantitative and qualitative measurement of the various aspects of behaviour of the individual for making generalised statements about his total performances.

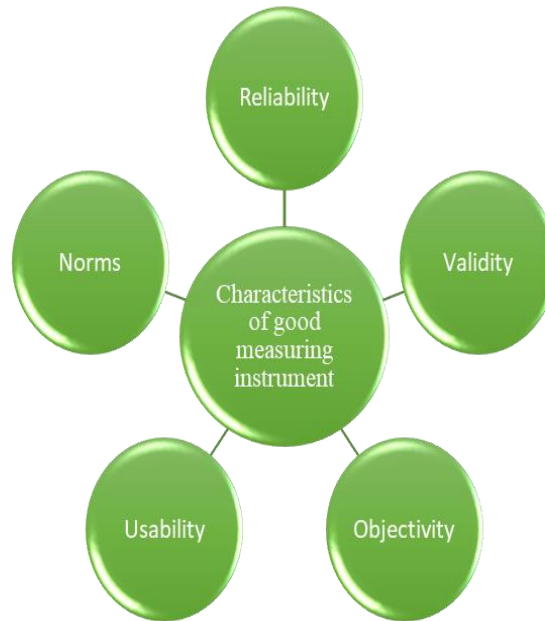
---

**13.3. BASIC CHARACTERISTICS OF GOOD MEASURING INSTRUMENTS**

---

Assessing students is an integral part of the teaching learning process. Therefore, teachers evaluate students' performance after every lesson and academic year. Testing aids teachers in determining their level of efficacy, the suitability of their teaching methods, and the appropriateness of their lesson plans. Students benefit from knowing what they have learned as well as their strengths and weaknesses. All these decisions require evidence about student performance. How to get this evidence? One way is to collect this evidence by using a good test. A good test is required to perform the function. What constitutes a good test then becomes a question. What qualities make a good test? When a test has the qualities of Validity, Reliability, Usability, and Norm and is created by adhering to the guidelines/steps of test construction, it is good. Test items should not be ambiguous i.e., clearly stated with one meaning. These principles guide test developers in creating tests that accurately measure the intended constructs/outcomes while minimizing bias and promoting ethical considerations. There are four essential technical qualities of all measuring instruments such as validity, reliability, usability, and norms.





## Validity

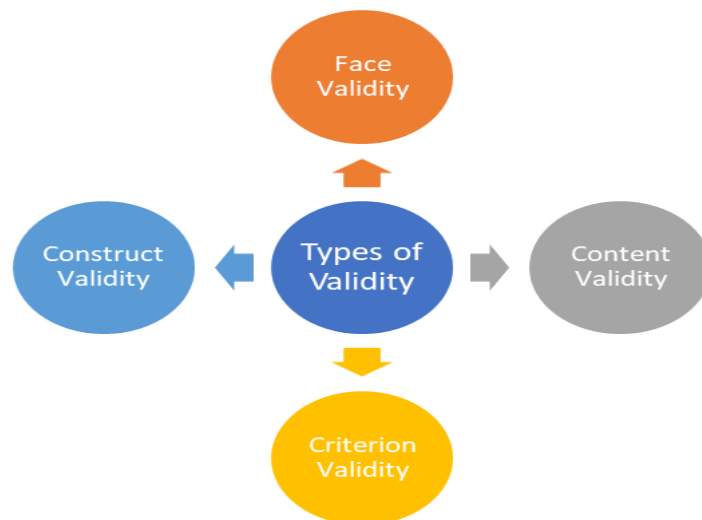
The most essential characteristics of a measurement instrument is validity. In layman's terms, validity means truthfulness, accuracy, correctness and worthiness. Validity of a test refers to the extent to which assessment test results serve the particular purpose for which it was intended. If the test is intended to measure the mathematical reasoning of the pupil, the results should describe about the same construct only i.e., mathematical reasoning, not scientific reasoning, or attitude. Moreover, when the instrument fulfils its purpose or measures what it originally intended to measure, then it is valid. Before teaching in a class, the teacher always has some specific objectives in mind which the teacher wants the students to achieve. Validity is not a general characteristic that is present or absent, it is situation specific; for a particular purpose of the group. No test is valid for all purposes. It is valid for a specific purpose. Hence, it is very essential to develop a test that is valid which can help in correct interpretation of the result. Validity can be of different types such as face, content, construct, and criterion validity.

## Types of Validity

- **Face validity:** It refers to the extent to which a tool appears to measure what it claims to measure. Face validity is not related to what the test measures, but is concerned with the test items that seem to be related with the variables being measured. Face validity is not validity in the true sense. Sometimes looking valid may not guarantee validity; also, sometimes the test showing low validity is genuinely valid in practice. A mathematics test is required to look like a mathematics test to have face validity.
- **Content validity:** Content validity is concerned with the extent to which items of the test are a good sample (representative) of the total content area to be measured. The key aspect of content validity is sampling of content so that the inference drawn over the total content domain is valid. It also describes the extent to which a test measures up against all the elements of the content. This type of validity is very essential for achievement tests. When developing an educational test, we must adhere to the test blueprint for ensuring its content validity. Further, comments and suggestions of content experts are also helpful for making the test content valid. For example, the test

developed for measuring mathematics of grade 9 should cover the representative sample of questions from the textbook of grade 9 to be content valid.

- **Criterion validity:** It is the kind of validity which is very much required for aptitude testing. This validity is statistical in nature which requires knowledge of correlation. There are two types of criterion validity; *concurrent* and *predictive*. In case of concurrent validity, test score is correlated with recent test results available for the same students. For example, a teacher wants to estimate the concurrent validity of a medical aptitude test that he/she developed. Teacher must administer the test to get the score and collect the score of the recent class test of the same students. The correlation coefficient of these two sets of scores will indicate the concurrent validity. For estimating the predictive validity of the same medical aptitude test, teachers need to wait for the completion of the medical course of the same group of students to get the predictive criterion. Then correlation coefficient between these two sets of scores will indicate the predictive validity.
- **Construct validity:** It refers to whether a test measures the intended construct (intelligence, aptitude, personality) adequately. This type of validity is very useful for validating psychological constructs. It measures how well test performance can be interpreted as a meaningful measure of some characteristics or quality. This is a very lengthy and technical process which requires knowledge of psychological theory and statistics.



### Factors affecting validity of the test

Validity of a test refers to the extent to which the test measures what it intended to measure. But there are several factors which affect the validity of a test or tend to make the test results invalid. For example, in the general science test, the teacher overloads questions related to environmental pollution and thus it is less valid a measurement of achievement in general science. So, a teacher should be aware of these subtle factors influencing validity while constructing a classroom test or using standardized test; some of the important factors are pointed out below:

- Difficult sentence construction and reading vocabulary:

In an achievement test if a student fails to understand the complex sentence structure or vocabulary of the question, he/she will not be able to reply even if she knows the answer. In this case, it becomes a reading comprehension test rather than an achievement test which invalidates the test for their intended use.

- **Unclear direction:**

If the direction given in the test items to be answered is not clear, students may get confused about how to respond to the items. This may lead to lessening the validity of the test.

- **Difficulty level of test items:**

If the level of questions in the test is too difficult or too easy as per the mental level of students, they cannot discriminate between the high and low achievers; thereby the validity of the test is lowered.

- **Extraneous factors:**

Several extraneous factors like length of the writing, organising answers, drawing diagrams and handwriting etc. influence the test validity. For example, in a science test if a student gets higher marks because she has drawn attractive diagrams, then the test is measuring drawing competence rather than her whole achievement in science; this reduces the validity of the test.

- **Inadequate coverage:**

Generally, the essay type tests are not able to cover a desired portion of content, because they do not provide a representative sample of content to be measured which lowers the validity.

- **Ambiguity:**

If there are ambiguous test items, they tend to confuse students; especially bright students are more confused in comparison to poorer students. This situation often distracts their mind from the focus area which reduces the validity of the test.

## **Reliability**

One of the important features of a measuring instrument is reliability. Reliability means consistency or dependability of test results. Reliability of a measuring instrument reflects its consistency or trustworthiness with which an instrument yields stable and accurate results. It is concerned with replicability or repeatability over time. A test is reliable when it gives consistently similar results in different occasions, time, and situations. A reliable test is free from biasness or chance error. It is the extent to which a particular measurement is consistent and reproducible. The reliability is statistical in nature and expressed in correlation coefficient. There are different kinds of reliability which are discussed in the following paragraphs.

### **Types of reliability**

- **Test-retest reliability:** This is the simplest and commonly used method of reliability test. It is obtained by administering the same test twice over a period to a group of

individuals. The gap between two tests can be one week to six months. The scores of both the test scores are tabulated and correlation is calculated. If the correlation is higher, then the reliability is more. This type of reliability test is used to access the external consistency (stability) of the tool over time. Disadvantage of the test-retest method is that it takes a long time to obtain the results because of gaps in two administrations.

- **Equivalent or parallel forms of reliability:** It is used to assess the consistency of the results of two tests constructed in the same way from the same content domain. Assessment tools of different versions are applied to the same group of individuals i.e., two tests (parallel forms) constructed in the same way from the same content domain that seeks to assess the same content and objectives. Parallel form questions are based on the same content, assess the same objectives, same difficulty level questions, same type of items etc. For estimating this kind of validity, one must prepare two sets of questions (parallel), administer it in the same group simultaneously or with a gap, score both the test, then calculate correlation between two sets of scores. This correlation is indicative of reliability. This kind of reliability is widely used for educational settings. The main disadvantage is to prepare exactly two parallel forms of the test.
- **Split half method of reliability:** This kind of reliability measures the internal consistency of the test scores. It measures to the extent to which all parts of the test contribute equally to what is being measured and tests the internal reliability. This is done by correlating the result of one half of a test with the result of another half. If the two halves of the test provide similar results, this would suggest that the test has internal reliability. This method is effective in case of large questionnaires which measure the same construct. For finding split-half reliability, one should prepare a single test and administer it on a group of students, then divide the test into two equal halves, score the two halves of the test separately. A test can be splitted in two halves by dividing tests on odd and even methods. Putting all odd questions such as 1, 3, 5, in one group and all even items such as 2, 4, 6 in another group. Here the test is divided for scoring purposes only, not for administration. Now we have two sets of scores for estimating correlation that will indicate *half reliability* of the test. To get full reliability of the test the Spearman Brown Prophecy formula is used.

$$r_t = 2(1/2r) / 1 + 1/2r$$

$r_t$  = Total reliability

$1/2r$  = Half reliability

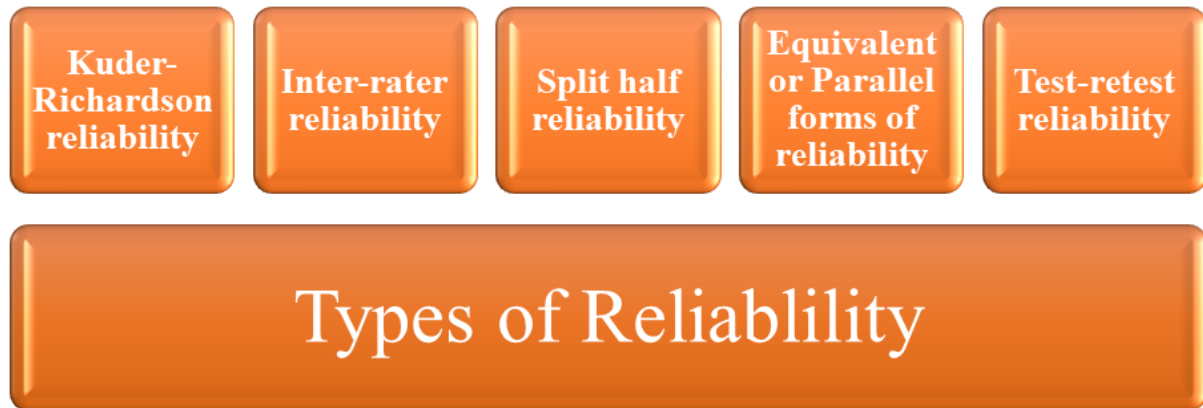
For example, if correlation between the two halves of a test is .65, the reliability of the full test will be;

$$r_t = 2 \times .65 / 1 + .65 = .78$$

This correlation indicates high reliability of the test.

- **Kuder-Richardson reliability:** This method of reliability determines the internal consistency of the test with single administration. According to Gronlund, its estimates of reliability provide information about the degree to which the items in the test measure similar characteristics. This reliability can be calculated by using Kuder-Richardson Formula 20 manually as well as with the help of computers.

- **Inter-rater reliability:** This reliability is the degree to which different observers give consistent estimates for the same subject or content. This is used to reduce the biases or observer effect on the result. When multiple observers (raters) observe the same thing all of them cannot be biased or have the same point of view towards the subject. This leads to increased objectivity of the test.



#### Factors affecting reliability of the test

Reliability in assessment refers to the consistency of the measure over time and across researchers. It is the extent to which test scores are not affected by the chance factors. When we measure the achievement level of a student, we expect that the test will be trustworthy or the score would be similar under different administrators at different times. Human behaviour fluctuates from time to time and situation to situation; So, in psychological and sociological measurements, there are some factors which tend to make the test low in reliability or affect the reliability of the test. Some of them may be due to the items in the test itself, some may be due to the test users and test takers. The teachers should be aware about those factors while designing the test to lessen their effects on the reliability of the test. Some of the important factors affecting the reliability of the test are explained below.

- **Subjectivity in scoring:**

Subjectivity in the process of scoring is indirectly proportional with the reliability of the test. As we know, subjectivity in assessment is when the personal bias, beliefs, perception, or value system of an examiner affect the result. If a measure is scored subjectively, it reduces the reliability of the test.

- **Ambiguous wording of the test items:**

Clear and concise test instructions tend to increase the reliability of the test whereas lack of clarity in instructions and complex wording of the test items lead to reduce it. If there is ambiguity in the wording of test items, the same student may interpret the same question in different ways at different times. This makes the test less reliable.

- **Inconsistency in test administration:**

Sometimes the examiner has insufficient knowledge about the test administration process; also, deviation is seen in the process of test administration such as deviation in procedure, timing etc. Fluctuations in the behaviour of the students such as attention, interest, illness, fatigue, or lack of motivation etc. also tend to reduce the reliability of the test.

- **Length of the test:**

Generally, the longer tests are more reliable. Length of the test has a positive correlation with its reliability; more the number of items in a test, greater will be its reliability. For example, a test of 50 items is more reliable than a test of 20 items, and less reliable than a test of 90 items. However, adding too many items out of limit may reduce the reliability of the test.

- **Difficulty level of the test items:**

When the items given in a test are too difficult or too easy for the participants, it can neither discriminate between low and high achiever nor contribute to the reliability of the test.

- **Optional questions:**

If there are optional questions given the test items, then one student may not appear the same question which she has attempted in the first administration; reliability of the test can get affected negatively by this.

***Self-check Exercise-4.5***

Split-half reliability of a test is .76. Calculate the full reliability of the same test.

-----

-----

-----

-----

-----

-----

-----

**Relation between reliability and validity**

The ultimate purpose of assessment is to check whether students can achieve the predetermined objectives or not. For this, the measuring instrument should be objective based i.e., it should be able to measure the learning outcome accurately based on the objectives (which was intended to measure). Reliability is a necessary but not sufficient condition for validity; validity includes reliability. If a test is valid, it is said to be reliable but a reliable test may or may not be valid. *For example*, a test is consistently measuring logical reasoning over time, situation, and form but the purpose of the testing is to measure the mathematical

reasoning. Here, the test is reliable because it consistently gives similar results but it is valid as it does not measure mathematical reasoning.

### **Objectivity**

Objective means free from biases. In case of test, objectivity refers to objectivity in test items and scoring procedure. A test must be objective in nature. A measuring instrument is objective when the examiners/examinee's opinion and perception does not affect the scoring. Fairness of a test to the subject is called objectivity. Generally, the test items that can be readily scored like true-false type, multiple choice type, and alternative response type are highly objective whereas long type or essay type items are highly subjective. As a test developer, we must strive to make the test objective as possible by following rules of writing test items. A test having objectivity has the following characteristics;

- Test items are so worded that it can be interpreted similarly by all students/ test takers.
- Test items must be so formulated that it can have only one answer agreed by experts.
- Test is not affected by the scorer's own value, judgement, attitude, and beliefs.
- Being free from all the biases of language, gender, culture etc.

Objectivity of the test can be ensured by; providing specific instructions, making essay type tests more unambiguous and well-constructed, preparation of scoring or marking key/scheme and using objective type of test items wherever possible.

### **Usability**

One of the most important criteria for the quality of measurement is the usability of the measuring instrument. Usability of an assessment tool refers to practical applicability of the test; how far the test is suitable and usable in a classroom situation. If a test has high validity and reliability but lacks usability, it will not be useful for the educational purpose. Following points below describe several criteria of usability of a measuring instrument.

- The test should be economical in terms of cost and time i.e., it should be affordable and the test items can be attainable within allotted time.
- Administering the test, scoring (assigning quantitative value to the test result) and interpreting the result should not be difficult.
- Usability of a test also shows how comprehensible the test items are to the participants. Test items that cover much of the content or subject matter are comprehensive and capable of fulfilling purpose.

### **Norms**

Teacher gets the score after administering and scoring the test on students. This score is called the Raw score. Raw scores are not meaningful and comparable for making any instructional decisions. Because it gives only numerical descriptions about students' performance. For example, the score 45 indicates a number 45, it does not indicate 45 out of what? If one student secured 45 in English and 65 in social science, it is difficult to say in which subject student has done better? Hence, raw scores are not interpretable and meaningful for stakeholders of education. To make the raw score meaningful and usable, norms are required.

Norm is one of the characteristics of the measuring instrument which is helpful for interpreting test results. Test norms are prepared to properly interpret the results. Norms are the average performance of the representative group of individuals on any test. Common norms are; age norms, grade norms, percentile norms, percentile, and standard norms. A norm group or reference group to be appropriate must have features of recency, relevance, and representativeness.

### **Types of norms**

Norms are useful to compare the performance of an individual with that of a group. There are different types of norms used in educational and psychological tests. Each type of norm serves a specific purpose and helps in differentiating and interpreting assessment results in various contexts. They are discussed below:

#### ***Age norms:***

Age norms, in the context of assessment, evaluate an individual's development and performance in comparison to the expected/ average performance of the pupils of his/her age. These norms are useful for teachers and parents to track developmental progress of their children and identify potential areas of concern.

#### ***Grade norms:***

Grade norms describe the test performance of individuals in comparison to the average performance of others in the same grade. This type of norming is useful for the educators to assess a student's academic progress within the context of their grade in relation to their peers so that it will be easier to make decisions related to required educational support or interventions for the students.

#### ***Percentile norms:***

Percentile norms express an individual's performance relative to a larger group in terms of the percentage of pupils scoring below her. For instance, a student at the 90th percentile performed better than 90% of her batch mates. Percentile norms offer a standardized way of understanding where an individual stands in comparison to a larger population.

### **Standard Score Norms**

Standard scores are a way of expressing the relative position of a pupil to a standardization group. It shows how far the raw score is above or below the average, overcoming the difficulties of unequal units on which percentile norms are based. The standard scores are expressed in terms of standard deviation units. Some of the widely used standard scores are Z-score, T-Score, Stanines etc.

#### **Z-Score**

Z-score is one of the simplest ways to convert raw score into a standard score. it is the most used standard score which shows how far a particular score deviates from the mean in a distribution. A positive Z score indicates the score above the mean while a negative Z score



indicates below the mean. Z score is symbolized by the letter 'Z' and can be derived through the following formula:

$$Z = \frac{X-M}{\sigma}$$

Here, Z= Standard Score

X= Raw score

M= Mean Score

$\sigma$  = Standard Deviation

### Example:

Ravi, a student of grade VII has secured 80 marks in mathematics and 75 marks in English subject. The mean score is 70 in mathematics with a standard deviation of 15 and in English, the mean score is 64 with a standard deviation of 12. Find Out in which subject Ravi performed better?

Ans- Z score for mathematics is

$$Z = \frac{80-70}{15} = 0.667$$

Z score for English is

$$Z = \frac{75-64}{12} = 0.916$$

Since the Z score is more for English, we can conclude that Ravi performed better in the English test than mathematics.

### T-Score

T- score is another type of standard score used in psychological and educational testing. In Z-score, when the raw score is smaller than the mean, the result comes with a minus sign which tends to create difficulty while interpreting the result. To overcome this difficulty of Z-score, we use a modified version of it, which is T-score. The formula used to compute the T score is as follows.

$$T \text{ score} = 50 + (10Z)$$

Example:

From our earlier example, we have a Z score of 0.66 in mathematics and 0.91 in English. Let us convert these two into T scores.

$$T \text{ score in mathematics} = 50 + (10 \times 0.66) = 56.6$$

$$T \text{ score in English} = 50 + (10 \times 0.91) = 59.1$$

From here also, we can conclude the same as above that in English, the performance is better than mathematics. One notable characteristic in T score is that results are produced in positive integers making the interpretation process much easier and simpler.

### Stanine Scores

Another quick way to understand where a student's performance falls in comparison to a larger group is represented by the stanine scale. The term stanine is derived from the word's 'standard' and 'nine.' This is also a normalized standard score which assumes that there is normal distribution in the measurement.

Total distribution in a stanine scale is divided into equal standard units ranging from 1 to 9 with mean or average of 5. Each stanine represents a specific range of percentile ranks. For example, a stanine of 9 indicates a percentile rank in the top 4%, while a stanine of 1 represents a percentile rank in the bottom 4%. Through this scale, raw scores can be converted into stanine scores by arranging the original scores in order and then assigning stanines according to the normal curve percentage as shown in the following table.

Stanine	1	2	3	4	5	6	7	8	9
% of case	4	7	12	17	20	17	12	7	4

### C-Scale

A small extended form of stanine scale is C-scale which is developed by Guilford. In this scale, it has 11 units of standard scores instead of 9. Here also, the mean or average score is 5 with extreme scores 0 and 10. In C-scale, normalized standard scores are assigned to each position in terms of percentage as given in the following table.

C-scale score	0	1	2	3	4	5	6	7	8	9	10
% of cases for each unit	1	3	7	12	17	20	17	12	7	3	1

### **Self-check Exercise-4.6**

Teacher administered an achievement test to 40 students. The mean of the group is 32 and standard deviation is 4.56. Find out the standard score for a student who secured 38 marks and interpret the result.

---

---

---

---

---

---

---

---

---

---

---

### 13.4. SUMMARY

---

- In layman's terms, validity means truthfulness, accuracy, correctness and worthiness. Validity of a test refers to the extent to which assessment test results serve the particular purpose for which it was intended
- **Face validity:** It refers to the extent to which a tool appears to measure what it claims to measure.
- **Content validity:** Content validity is concerned with the extent to which items of the test are a good sample (representative) of the total content area to be measured.
- **Criterion validity:** It is the kind of validity which is very much required for aptitude testing. This validity is statistical in nature which requires knowledge of correlation
- There are two types of criterion validity; *concurrent and predictive*. In case of concurrent validity, test score is correlated with recent test results available for the same students.
- **Construct validity:** It refers to whether a test measures the intended construct (intelligence, aptitude, personality) adequately. This type of validity is very useful for validating psychological constructs. It measures how well test performance can be interpreted as a meaningful measure of some characteristics or quality
- Reliability means consistency or dependability of test results. Reliability of a measuring instrument reflects its consistency or trustworthiness with which an instrument yields stable and accurate results.
- **Test-retest reliability:** This is the simplest and commonly used method of reliability test. It is obtained by administering the same test twice over a period to a group of individuals.
- **Equivalent or parallel forms of reliability:** It is used to assess the consistency of the results of two tests constructed in the same way from the same content domain.
- **Split half method of reliability:** This kind of reliability measures the internal consistency of the test scores. It measures to the extent to which all parts of the test contribute equally to what is being measured and tests the internal reliability.
- **Kuder-Richardson reliability:** This method of reliability determines the internal consistency of the test with single administration. According to Gronlund, its estimates of reliability provide information about the degree to which the items in the test measure similar characteristics.
- **Inter-rater reliability:** This reliability is the degree to which different observers give consistent estimates for the same subject or content. This is used to reduce the biases or observer effect on the result.
- In case of test, objectivity refers to objectivity in test items and scoring procedure. A test must be objective in nature. A measuring instrument is objective when the examiners/examinee's opinion and perception does not affect the scoring. Fairness of a test to the subject is called objectivity.
- Usability of an assessment tool refers to practical applicability of the test; how far the test is suitable and usable in a classroom situation
- Norms are the average performance of the representative group of individuals on any test. Common norms are; age norms, grade norms, percentile norms, percentile, and standard norms.

- Age norms, in the context of assessment, evaluate an individual's development and performance in comparison to the expected/ average performance of the pupils of his/her age.
- Grade norms describe the test performance of individuals in comparison to the average performance of others in the same grade.
- Percentile norms express an individual's performance relative to a larger group in terms of the percentage of pupils scoring below her.
- Standard scores are a way of expressing the relative position of a pupil to a standardization group. It shows how far the raw score is above or below the average, overcoming the difficulties of unequal units on which percentile norms are based.
- Z-score is one of the simplest ways to convert raw score into a standard score. it is the most used standard score which shows how far a particular score deviates from the mean in a distribution.
- T- score is another type of standard score used in psychological and educational testing. In Z-score, when the raw score is smaller than the mean, the result comes with a minus sign which tends to create difficulty while interpreting the result.

---

### 13.5. Unit End Exercise

---

- What are the basic characteristics of good measuring instruments?
- What is the meaning of validity? What are the types of reliability? Which factors do affect the validity?
- What is reliability? What are the types of reliability? Which factors do affect the reliability?

---

### 14.8. Suggestion for Further Reading

---

Gronlund, N. E. (1965). *Measurement and evaluation in teaching*.  
<http://ci.nii.ac.jp/ncid/BA12623208>

Goswami, M. (2013). *Measurement and Evaluation in Psychology and Education*  
Lee, W. Y. (2010). *Assessment and evaluation in higher education*.  
<http://ci.nii.ac.jp/ncid/BB11596810>

Patel, R. N. (2014). *Educational evaluation theory and practice*.  
<http://ci.nii.ac.jp/ncid/BA29677030>

# UNIT 14

## STANDARDISATION OF MEASURING INSTRUMENTS AND ITEM ANALYSIS

### STRUCTURE

- 14.1. Learning Objectives
- 14.2. Introduction
- 14.3. Standardisation of Measuring Instruments and Item analysis
- 14.4. Methods of standardization in assessment
  - 14.4.1. Test Development
  - 14.4.2. Pilot Testing
  - 14.4.3. Administration Protocols
  - 14.4.3. Scoring Rubrics
  - 14.4.4. Norming and Calibration
  - 14.4.5. Ongoing Monitoring
  - 14.4.6. Developing Manual
- 14.5. Item Analysis
  - 14.5.1. Steps of item analysis
  - 14.5.2. Item analysis procedures for Norm- Referenced and criterion referenced tests.
  - 14.5.3. Item analysis procedures for Norm- Referenced Classroom Tests
  - 14.5.4. Estimating difficulty level of test items
  - 14.5.5 Discrimination index
  - 14.5.6. Distractor Analysis
  - 14.5.6. Item analysis procedures for Criterion- Referenced Mastery Tests
- 14.6. Summary
- 14.7. Unit End Exercise
- 14.8. Suggestion for Further Reading

## 14.1. Learning Objectives

## 14.2. Introduction

---

**14.1. LEARNING OBJECTIVES**

---

After reading this unit, learners shall be able to;

- Define the Standardisation of Measuring Instruments and Item analysis.
- Explain the Methods of standardization in assessment
- Elaborate the concept of item analysis, steps, item analysis procedur

---

**14.2. INTRODUCTION**

---

A standardized measurement instrument is a rigorously developed tool that measures a concept (or an indicator) in an objective, standardized manner. It can be defined as a series of self-reported questions or items used to measure a concept. The response categories to an item are usually in the same format, often in the form of a numbered scale. The items measuring a concept form a scale in respect of which a quantified score is obtained, often by adding the results or with a more or less complex weighting system. The scores on the scale can subsequently be converted into a norm in order to facilitate interpretation. The instruments must undergo rigorous validation stages and the information on psychometric properties such as test-retest reliability, internal coherence, specificity and sensitivity must be determined and made available.

Item analysis is a process which examines student responses to individual test items (questions) in order to assess the quality of those items and of the test as a whole. Item analysis is especially valuable in improving items which will be used again in later tests, but it can also be used to eliminate ambiguous or misleading items in a single test administration. In addition, item analysis is valuable for increasing instructors' skills in test construction, and identifying specific areas of course content which need greater emphasis or clarity. Separate item analyses can be requested for each raw score

---

**14.3. STANDARDISATION OF MEASURING INSTRUMENTS AND ITEM ANALYSIS**

---

Standardization of measuring instruments is fundamental to the accuracy and consistency of measurements in various fields. It ensures that measurements are meaningful, comparable, and traceable to recognized standards. Standardization in assessment refers to the systematic process of developing, administering, scoring, and interpreting tests or assessments in a consistent and uniform manner. This process aims to ensure that all test-takers are subjected to the same conditions and that their scores can be meaningfully compared, regardless of when or where the test is administered. Standardisation process makes the measuring instrument relevant and usable for a larger population even across the Globe. The mental ability test like the Standard Progressive Matrix is widely used across the World because it is a standardized test. The method of standardisation includes the following steps.

---

**14.4 METHODS OF STANDARDIZATION IN ASSESSMENT**

---

- **Test Development:** During the test development phase, clear guidelines are established for item writing, test format, and scoring procedures. A detailed test blueprint or test plan is often created to ensure that the assessment measures the

intended construct or capacities. One needs to write test items and scoring keys/model answers along with instructions.

- **Pilot Testing:** Before full-scale administration, assessments are pilot-tested on a small group of participants to identify potential issues with item wording, difficulty, or scoring procedures. Feedback from pilot testing helps to refine the items that can be modified. The piloting data can be used to calculate the technical features of the test like validity, reliability etc.
- **Administration Protocols:** Standardized protocols for test administration are developed to ensure uniformity in the test-taking environment. These protocols specify instructions given to test-takers, time limits, and procedures for handling administrations.
- **Scoring Rubrics:** Detailed scoring rubrics are created to guide consistent and objective scoring. These rubrics provide explicit criteria for assigning scores to open-ended questions or performance-based tasks.
- **Norming and Calibration:** After administering the assessment to a representative sample, the results are used to establish norms and calibrate scoring. This process ensures that scores are comparable and interpretable. Details of norms are developed for the standardised instrument for interpretation of results.
- **Ongoing Monitoring:** Standardized assessments are periodically reviewed and updated to ensure they remain relevant and valid. This may involve revising test items, norms, or scoring procedures to reflect changes in the field or population.
- **Developing Manual:** Test manual is a very essential part of the standardisation process. It gives a detailed idea about the test development process, process of administration, scoring, technical qualities, and norms for interpretation. It is like a key without which no measuring instrument can be used.



Standardisation of measuring instruments makes it valid, reliable, objective, and usable. Most of the psychological tests are standardized on a larger population by taking samples from different countries to make it wider applicability. The process of standardisation is lengthy and a time taking process. A Standardized test is required when important decisions are to be

taken based on the test result. A teacher can follow all the steps of standardisation to develop authentic measuring tools.

---

## 14.5. ITEM ANALYSIS

---

Item analysis is a statistical technique used for selecting and rejecting the test items based on their difficulty level and index of discrimination. A test should neither be too difficult or too easy, so the procedure used to judge the quality of the test items is called item analysis. Purpose of item analysis is to find out whether all the test items are contributing towards assessing the achievement of objectives or not. The defective or ambiguous items which are not serving the purpose are eliminated and good test items are constructed. It is desirable to evaluate the usefulness of the items after a test has been given to and scored by a representative group. This can be achieved by considering the subjects' responses to each item. Item analysis can provide three different types of information about the items in quantitative indices like difficulty level, discrimination power, and distraction analysis when done systematically. So basically, item analysis is done for the following purposes.

- Selecting appropriate items for preparation of final draft
- Bringing modification in items where needed
- Obtaining information about difficulty value, discriminatory power, and validity index of items.

### Steps of item analysis

- Once the sample group test has been administered and scored, the answer sheets are to be arranged in order from the highest score to the lowest score.
- Make two groups from the arranged answer sheets; one group having the highest 27% of the students' scores and one having the lowest 27% of the students' scores.
- For each item, note the number of students in each group who answered the items correctly.
- Estimate item difficulty, discrimination index of the test items using formula.

### Item analysis procedures for Norm- Referenced and criterion referenced tests.

The method for analysing effectiveness of test items differ in case of Norm referenced and criterion referenced tests as they serve two different functions. In norm-referenced tests, we compare or rank the students based on their performance whereas in criterion referenced tests, the mastery of students in particular subjects are measured based on their learning outcome.

### Item analysis procedures for Norm- Referenced Classroom Tests

Determining effectiveness of test items in case of Norm-referenced classroom tests include computing the index of item difficulty (Percentage of pupil who got the items right), discriminating power of each item (difference between high and low achievers), and effectiveness of distractors (degree to which the distractors attract more low achievers than high achievers). Although effectiveness of items can be revealed by item analysis through inspection, to obtain a precise estimate of difficulty level and discrimination of test items, some simple formulae are applied to the item analysis data as follows;



### Estimating difficulty level of test items

One can determine whether the items are too easy or difficult through the item analysis. The percentage of learners who correctly answered the test question gives an indication of how difficult the question was. The purpose of estimating item difficulty is to know the proportion of learners who know the answer of the item. The item difficulty index computes the percentage of students who answered the items correctly through the following formula.

$$\text{Difficulty index} = R/N \times 100$$

Where, R= number of students who answered the item correctly.

N= Total number of students who attempted the item.

*For example;* A teacher has administered a test on a sample of 40 students. 22 students corrected the item number-1 and 10 students answered wrongly the item number-1. The difficulty index of item number-1 can be calculated as follows.

$$\text{Difficulty index} = 22/40 \times 100 = 55$$

This indicates that the item number-1 is of average difficulty value.

Difficulty index can also be expressed in decimals rather than percentage. It may be noted that higher the difficulty index, the easier the item is. If an item is too difficult or too easy, it cannot discriminate between the high and low performers and an item having 0% or 100% difficulty index has no discriminating value. It is ideal to select the items having average difficulty value for classroom testing.

### Discrimination index

The item discrimination power is another crucial index of item. This index shows how well the item can distinguish between high achievers and low achievers. If an item is effective, it is anticipated that more people from the high scorers will get it right in comparison to the number of people in the lower group. For finding the discrimination index, one must divide the whole group into two groups; upper group and lower group. The two groups can be the students scoring highest 27% and students scoring lowest 27%. An item's discrimination index can be obtained by the following formula.

$$\text{Discrimination Index} = R_U - R_L / \frac{1}{2} N$$

Where  $R_U$  – Number of correct responses from the upper group

$R_L$  – Number of correct responses from the lower group

N – Total number of students who attempted the item.

*For example:* A teacher has administered a test on 200 students. The number-1 item is correctly answered by 48 students from the upper group and 34 students from lower group. The discrimination index of the item number-1 can be calculated as follows.

$$\text{Discrimination index} = 48 - 34 / 100 = .14$$

This indicates item number-1 is having low discrimination power.

A discrimination index is usually expressed as a decimal. The value may range from -1.00 to +1.00. We might want to add test items with higher discrimination index values because the items having high discrimination index are the better items. If discrimination index has a positive value, then it is said that a larger proportion of the more knowledgeable students than poor students got the item correct. The items having zero or negative discrimination index indicate that the items are too easy or too difficult or ambiguous. Those items should be revised or removed from the test.

### **Distractor Analysis**

Distractors are incorrect options or alternatives in multiple choice tests. If a distractor attracts a higher proportion of good students than poor students, it is said to be a poor distractor. It is reasonable to assume that for an effective item, more members of the high achiever's group will select the right answer than members of the low achievers group. Thus, by looking at the pattern of responses to the different alternatives or distractors, it is possible to assess the effectiveness of the distractors. Alternatives that have not been selected or are rarely selected need to be revised because they make no difference to how effective the item is.

*For example:* The response pattern of students to a multiple-choice question is tabulated in the following table.

Alternatives/ Group	A*	B	C	D
Upper group (20)	12	5	0	3
Lower group (20)	6	10	0	4

It can be said that option A is the key and attracts more students from the upper group than lower group. The option B and D are distractors that attract more students from the lower group than upper group. But no students from either group selected option C. The option C is a poor distractor as it does not attract any students.

#### ***Self-check Exercise-4.7***

Calculate the difficulty index of the following data. Total number of students attempting the test is 80, 60 students corrected and 20 students wrongly answered the item.

-----

-----

-----

-----

-----

-----

### **Item analysis procedures for Criterion- Referenced Mastery Tests**

Since Criterion- referenced tests are designed to describe the mastery of students or learning performance of students rather than to discriminate or rank them, it is not necessary to apply the above formulae used in norm referenced tests here. According to Gronlund, A major question to be asked to check the effectiveness of test items in criterion referenced tests is ‘To what extent did the test items measure the effects of the instructions?’ To address this question, the same test should be administered both before and after instruction (pretest and post-test), with the results compared.

---

#### **14.6. SUMMARY**

---

Constructing a test requires hard work and loads of effort on the part of the test developer or teacher. For constructing tests, the teacher should be properly oriented or trained about the principles and methods of test construction. Test development is a technical process which requires the teacher to be familiar with the principles and steps of test construction. Test development involves four important steps such as planning, preparing, trying out and evaluation. Planning is the stage of the decision-making process where the teacher decides about the content, learning outcomes, types of items, number of items, duration of test etc. Teacher develops a test blueprint during this stage which guides the further test development process. Teacher writes items as per the test blueprint. Usually, more items are written than required at this stage. Teachers must keep in mind the learning outcomes to be measured while writing the test items. Every item must measure a specific learning outcome. Tests must include both subjective and objective test items depending on the content and learning outcomes to be measured. Once test items are written, it must be reviewed and edited by the teacher, peer teachers and expert. This review can help the test developer in modifying poor items. The draft test is to be tried out on a small sample to see the effectiveness of each item, its language ambiguity, appropriateness of options, etc. Based on the try-out, some items can be dropped or modified to be included in the final draft of the test. Item analysis is one of the very important aspects of test development. It is the process of determining difficulty index, discrimination power and suitability of distractors. Item analysis gives feedback for selecting good and effective test items. There are different procedures for conducting item analysis for norm referenced and criterion referred tests. The last step is to evaluate the test and determine the technical quality of the test such as validity, reliability, norms etc.

The measuring instrument must have technical quality to make it appropriate for making any educational decisions. Validity is one of the essential qualities of the test which means truthfulness and purposiveness of test result. It is always relative in nature depending on the purpose of the testing. There are different types of validity; content validity, criterion validity and construct validity. Another important characteristic of the test is reliability. It refers to the consistency of the test results over time, different administration, and forms of the test. Reliability can be of test-retest, parallel forms, split-half and Kuder Richardson. Along with validity and reliability, a test must be objective and usable. Norms are very much required for making the raw score meaningful and comparable. The norms can be of age norms, grade norms, percentiles, standard score norms etc. These norms help the teacher to convey the test result to stakeholders of education.

---

#### **14.7. UNIT END EXERCISE**

---

- What is the meaning of standardisation of measuring instruments and item analysis?

- What are the methods of standardisation in assessment?
- What is item analysis? Explain its steps and procedure briefly?

---

**14.8. SUGGESTION FOR FURTHER READINGS:**

---

- Gronlund, N. E. (1965). *Measurement and evaluation in teaching*.  
<http://ci.nii.ac.jp/ncid/BA12623208>
- Goswami, M. (2013). *Measurement and Evaluation in Psychology and Education*
- Lee, W. Y. (2010). *Assessment and evaluation in higher education*.  
<http://ci.nii.ac.jp/ncid/BB11596810>
- Patel, R. N. (2014). *Educational evaluation theory and practice*.  
<http://ci.nii.ac.jp/ncid/BA29677030>

## **UNIT-15**

### **MEASUREMENT OF ACHIEVEMENT AND PSYCHOLOGICAL TRAITS**

#### **STRUCTURE**

- 15.1. Learning Objectives
- 15.2. Introduction
- 15.3. Measurement
- 15.4. Measurement of Achievement
- 15.5. Measurement of Aptitude
- 15.6. Measurement of Intelligence
- 15.7. Measurement of Attitude
- 15.8. Measurement of Interest and Skills
- 15.9. New Trends of Evaluation
- 15.10. Summary
- 15.11. Unit End Exercise
- 15.12. Further Reading

---

#### **15.1.LEARNING OBJECTIVES**

---

After reading this unit, the learners shall be able to

- Explain about different methods of measuring achievement.
- Discuss about different methods of measuring aptitude.
- Explain about different methods of measuring intelligence.
- Describe different methods of measuring attitudes.
- Discuss about different methods of measuring interest.
- Explain about different methods of measuring skills.
- Describe different new trends of evaluation.

---

#### **15. 2. INTRODUCTION**

---

In the multidimensional world of education, it is crucial to assess and measure different aspects of student's personality to better understand, guide, and improve the learning experience. This unit explores various aspects of educational evaluation, looking closely at student achievement, abilities, intelligence, attitudes, and the assessment of interests and skills. The discussion also covers new trends in educational evaluation, explaining the significant changes in grading methods, semester structures, continuous internal assessment

approaches, the development of question banks, and the use of computer technology in assessment methods.

---

### **15.3.MEASUREMENT**

---

**Measurement**, the process of associating numbers with physical quantities and phenomena. Measurement is fundamental to the sciences; to engineering, construction, and other technical fields; and to almost all everyday activities. For that reason the elements, conditions, limitations, and theoretical foundations of measurement have been much studied. See also measurement system for a comparison of different systems and the history of their development.

Measurements may be made by unaided human senses, in which case they are often called estimates, or, more commonly, by the use of instruments, which may range in complexity from simple rules for measuring lengths to highly sophisticated systems designed to detect and measure quantities entirely beyond the capabilities of the senses, such as radio waves from a distant star or the magnetic moment of a subatomic particle.

Measurement begins with a definition of the quantity that is to be measured, and it always involves a comparison with some known quantity of the same kind. If the object or quantity to be measured is not accessible for direct comparison, it is converted or “transduced” into an analogous measurement signal. Since measurement always involves some interaction between the object and the observer or observing instrument, there is always an exchange of energy, which, although in everyday applications is negligible, can become considerable in some types of measurement and thereby limit accuracy.

---

### **15.4. MEASUREMENT OF ACHIEVEMENT**

---

Achievement is very important for formal education as it determines the effectiveness of educational machinery. It refers to the level of knowledge, skills, and competencies that individuals have acquired through their educational experiences. It is a fundamental concept in education, serving as a key indicator of learning outcomes and educational success. Measurement of achievement is crucial for educators, policymakers, and students to assess progress, make informed decisions, and improve the quality of education. Let us understand the meaning of achievement before discussing the process of measuring achievement.

Achievement in education signifies the attainment of specific learning objectives, educational standards, and subject matter mastery. It reflects a student's ability to apply their knowledge, skills, and competencies in various academic, vocational, and non-academic domains. Achievement goes beyond memorization and encompasses the development of critical thinking, problem-solving, communication, and other skills essential for personal and professional growth. Achievement in education can be of two types.

**Educational Achievement:** Educational achievement refers to the successful attainment of educational goals and objectives, often measured through assessments and evaluations. It includes both academic and non-academic accomplishments. It is related to the mastery of academic subjects, including mathematics, science, language arts, and social studies. It is often measured through traditional assessment tools, such as tests and exams.

**Psychological Achievement:** It refers to the achievement of students in psychological attributes such as mental ability, aptitudes, interests, creativity, reasoning ability etc. This can be measured through different psychological tests, scales, checklists etc.

The measurement of achievement in education involves the systematic assessment of what individuals have learned or accomplished within a specific educational context. This process is crucial for evaluating the effectiveness of instruction, understanding individual progress, and making informed decisions about educational programs. Educational achievement is usually measured by an achievement test based on different subject areas.

**Achievement Test-**Achievement test is one of the most popular methods of measuring achievement. It is a commonly used instruments to measure an individual's knowledge, skills, or proficiency in a specific subject or set of subjects. The primary purpose of achievement tests is to evaluate the extent to which students have mastered the content outlined in the curriculum or specific learning outcomes. Two prominent forms of achievement tests employed in education are; teacher made achievement tests and standardized achievement tests. Let us discuss both types of achievement tests in detail.

### **Teacher-made Achievement Test**

Teacher-made achievement tests are created and used by educators for their classroom purpose. These assessments are tailored to align with the specific content covered during instruction, making it highly context-dependent. Teachers have the autonomy to design

questions, select formats, and determine the overall structure of the test. One of the primary characteristics of teacher-made tests is their local development and relevance. Teacher craft assessments that directly correspond to the curriculum and learning objectives they have established for their students. This local development aspect allows for a high degree of customization. Teachers can adapt the difficulty level of questions to match the unique needs of their students, ensuring that the assessment is both challenging and relevant. Teacher-made tests also facilitate immediate feedback. Since teachers are intimately involved in the test creation process, they can quickly assess student performance. This immediacy allows for prompt feedback, enabling teachers to identify areas of strength and weakness and make timely instructional adjustments.

### **Advantages of Teacher Made Test**

- Teacher-made tests closely mirror classroom instruction, reflecting the specific content and skills emphasized in the curriculum. This alignment ensures that assessments are relevant to the ongoing educational experience of students.
- Teachers have the flexibility to adapt assessments to the diverse needs of their students. They can modify questions, formats, and even time constraints based on their understanding of individual learning styles and preferences.
- Since teachers are intimately involved in the test implementation process, they can provide immediate feedback to students. Timely feedback enhances the learning process, allowing students to address misconceptions and improve their understanding before moving on to new material.
- These tests are developed by educators who understand the unique context of their classrooms and students. Teacher-made tests can be crafted to reflect the specific needs, interests, and experiences of the students, making the assessment more relevant and meaningful.

### **Limitations of Teacher Made Test**

- There is a potential risk of bias in teacher-made tests. The assessments may reflect the teacher's individual perspectives, teaching methods, and personal preferences, introducing a subjective element.



- Achieving consistency across different classrooms or schools may be challenging. Variations in test development, even within the same educational institution could compromise the standardized nature of assessments.
- Teacher-made tests are often context-specific. Results may not be easily generalizable to broader populations or comparable across different educational settings due to the localized nature of the assessments.
- Teachers may face time constraints in creating, administering, and grading assessments, especially in the context of heavy workloads. The practicality of teacher-made tests may be limited by time considerations, potentially impacting the quality of the assessment.

### **Standardized Achievement Test**

Standardized achievement tests are developed externally by testing organizations or educational experts. These assessments are designed to be uniform, administered, and scored consistently across a broad and diverse population of students. They are often norm-referenced, comparing an individual's performance to a predetermined normative sample. The standardization of these tests ensures that every student, regardless of geographic location or educational background, faces the same set of questions under comparable conditions. The development process involves rigorous testing and validation to enhance reliability and validity. The advantage of the standardized achievement test is that it is very high in standard which can be used across countries and Globe. But this test is not contextualized, hence cannot be useful for local schools or teachers. For this reason, standardized tests are not very popular among teachers and educators for assessing achievement of students.

Some popular standardized achievement tests are discussed below:

**California Achievement Test-**The California Achievement Test (CAT) is released by CTB/McGraw Hill and comprises five levels tailored to different grade levels. Two alternative forms of the test are accessible. The assessment delivers scores in reading (covering vocabulary and comprehension), language (encompassing mechanics, usage, structure, and spelling), and mathematics (evaluating computation, concepts, and problem-solving).

**Comprehensive Tests of Basic Skills (CTBS):** Like the California Achievement Test, the CTBS is produced by CTB/McGraw-Hill, but it is designed for students spanning grades K-

12. The test offers seven levels corresponding to different grade levels. Level A serves as a pre-instructional or readiness test, assessing scores for letter forms, letter names, and Mathematics. Level B, intended for students who have completed their initial year of instruction, provides scores for reading, language, Mathematics, and Total Battery. Levels C, 1, 2, 3, and 4 yield scores in reading, language, Mathematics, reference skills (excluding Level C), Science, and Social Studies. A cumulative total battery score is also presented, combining scores from reading, language, and Mathematics.

**Iowa Tests of Basic Skills (ITBS)-** The Iowa Tests of Basic Skills (ITBS), published by the Riverside Publishing Company, is suitable for students in grades K-8. This assessment was standardized using the same sample as the Cognitive Abilities Test (CogAT), an academic aptitude test. Consequently, employing these two tests facilitates the identification of aptitude-achievement disparities. The ITBS provides scores across various domains, including listening, word analysis, vocabulary, reading, comprehension, language (spelling, capitalization, punctuation, and usage), visual and reference materials, Mathematics (concepts, problem-solving, computation), Social Studies, Science, writing and listening supplements, as well as basic and total battery scores.

**The Metropolitan Achievement Tests (MAT)-** The Metropolitan Achievement Tests (MAT), published by Harcourt Brace Jovanovich, is designed for students in grades K-9. The battery consists of six levels spanning various grades, with two alternate forms for the Primary Level and three alternate forms for the other five levels. The primary level offers scores for listening for sounds, reading, and numbers. The subsequent level, Primary I, includes scores for word knowledge, word analysis, reading, Mathematics computation, and Mathematics concepts. Primary II incorporates these elements along with spelling and Mathematics problem-solving. The remaining levels all provide scores for word knowledge, reading, language, spelling, Mathematics computation, Mathematics concepts, and Mathematics problem-solving. Additionally, Science and Social Studies scores are available for the two higher levels.

**Standard Achievement Test Series-** The Standard Achievement Test Series, published by Harcourt Brace Jovanovich, shares similarities with the MAT. It offers six levels tailored for different grades, along with two alternate forms. At all levels, subtests for reading, Mathematics, and language arts are accessible. Except at the lowest level, scores are also provided for Science, Social Studies, and, except at the highest level, listening

comprehension. Notably, this test has a distinctive feature: it can be obtained as either a basic battery, encompassing only the reading, Mathematics, and language art subtests, or as a complete battery, encompassing all the subtests. Practice tests are additionally available for all levels except the highest.

Coimbatore Achievement Test (Social Science) authored by R.K. Mission, Mathematics Test (Tamil) by R.K. Mission, School Progress Record by L.N. Dubey, English Test by B.R.CAgra, Mathematics Test by B.R.CAgra are some examples of Indian Standardized Achievement tests that can be used by the teachers for assessment.

### **Advantages of Standardized Achievement Test**

- Standardized tests undergo rigorous development processes, ensuring consistent administration and scoring. This objectivity minimizes biases and discrepancies in assessment, leading to reliable and comparable results.
- Standardized tests allow for comparisons across different schools, regions, or populations. The standardized format ensures that the performance of individuals or groups can be compared objectively, providing insights into educational trends.
- These tests are efficient for large-scale assessments. Standardized tests can be administered to many students simultaneously, making them practical for assessing broad populations.
- Standardized tests are often developed by external experts or testing organizations. This external input enhances the validity and reliability of the assessment, aligning it with established educational standards.
- Standardized test results can be used by policymakers, administrators, and educators to identify areas for improvement, allocate resources effectively, and make data-driven instructional decisions.
- Standardized tests often demonstrate predictive validity for future academic success. High scores on standardized tests are often correlated with positive academic outcomes, providing valuable insights into students' potential for success.

### **Limitations of Standardized Achievement test**

- Standardized tests may lack contextual relevance. The rigid format of these assessments may not capture the specific nuances of individual classrooms or local curriculum variations.

- Standardized tests follow a one-size-fits-all approach. This approach may not accommodate diverse learning styles, preferences, or individual needs, potentially disadvantages certain students.
- The pressure to perform well on standardized tests may lead to "teaching to the test." Educators may focus narrowly on the content covered in the test, neglecting broader educational goals and essential skills.
- High-stakes nature of standardized tests may induce stress and anxiety in students, potentially affecting performance.

***Self-check Exercise-5.1***

Make a list of standardized achievement tests useful for assessing reading and writing ability by exploring internet sources.

---

---

---

---

---

---

---

---

**15.5. MEASUREMENT OF APTITUDES**

---

Aptitude refers to a person's inherent or natural ability to excel or perform well in a specific area or task. It is a quality or potential that individuals possess, often from birth or early in life, which enables them to acquire certain skills, knowledge, or talents more easily and effectively than others. Aptitude might seem like ability, interest, or intelligence but each of them is different in some aspect. To understand the term aptitude better following definitions by experts can be considered:

Bingham (1937) states, "Aptitude is a characteristic or set of conditions that are symptomatic to the individual."

According to Traxler (1966), "An aptitude is a present condition which is indicative of an individual's potential for the future."

Jones (1971) describes that attitude is not an ability rather it helps to predict the probable development of certain abilities.

Freeman (1971) defined aptitude as a combination of characteristics indicative of an individual's capacity to acquire (with training) some specific knowledge skill, or a set of organized responses, such as the ability to speak a language, to become a musician, or to do mechanical work.

In essence, aptitude represents a valuable foundation upon which individuals can build and develop their abilities and talents, ultimately enabling them to excel in their chosen pursuits.

### **Methods of Measuring Aptitude**

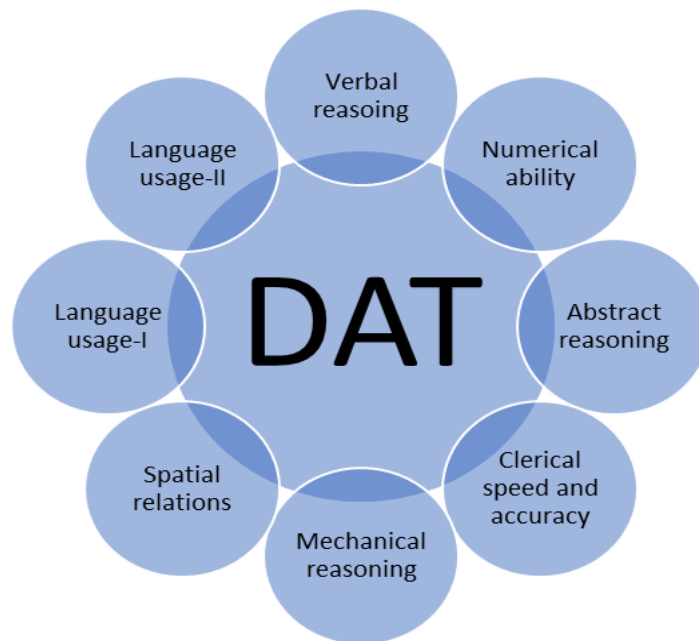
The measurement of aptitude involves assessing an individual's natural abilities, talents, or potential to acquire specific skills. Different aptitude tests have been designed to gauge an individual's capacity to excel in certain areas rather than measuring their existing knowledge. Some of the popular aptitude tests are discussed below:

#### **Differential Aptitude Test (DAT)**

The Differential Aptitude Tests were developed by the psychological cooperation in 1947 to measure the aptitude of high school students for educational and vocational guidance. At first the tests were used only for 8th to 12th grade students but later they were also used for educational and vocational guidance of young adults and employee selection. The tests were further revised in 1963 and 1972 to meet the demands of small and large guidance programmes.

The DAT includes eight types of subtests such as Verbal Reasoning, Numerical Ability, Abstract Reasoning, Clerical Speed and Accuracy, Mechanical Reasoning, Spatial Relations, Language Usage-I: Spelling, Language Usage-II: Sentences. Each of the tests has its unique advantages but combining 2 or more tests in a subgroup provides comprehensive understanding of a particular type of aptitude. The Verbal Reasoning, Numerical Ability and Abstract Reasoning tests are associated with general intelligence, the Mechanical Reasoning and Space Relations relate to the ability of visualizing objects and manipulating those visualization, the Clerical Speed Accuracy Test and Language Usage I : Spelling and

Language Usage:II: Sentence associate with the skills of doing office work and academic success. Let us discuss the 8 tests individually in a little detail.



**Verbal Reasoning:** These tests evaluate an individual's ability to understand and manipulate language, including tasks related to vocabulary, analogies, and reading comprehension.

*Example-* Pick out the words to fill the blanks, so that the sentence will be true and sensible. For the first blank pick out one of the numbered words and for the second blank pick out the lettered words.

I) \_\_\_\_\_ is to doctor as lawyer is to \_\_\_\_\_.

1. Diagnosis 2. Patient 3. Lawsuit 4. Judge

A. Heal B. Advocate C. Court D. Legal

II) \_\_\_\_\_ is to be as web is to \_\_\_\_\_.

1. Insect 2.silk 3. Honey 4. Hive

A. Buzz B. Spider C. Sting D. Honeycomb

**Numerical Ability:** These tests assess mathematical and quantitative reasoning skills, including arithmetic, algebra, and data interpretation.

*Example-* Choose the correct answer for each problem

I) Add 23 and 22

A. 14    C. 16

B. 45    D. 59    E. None of These

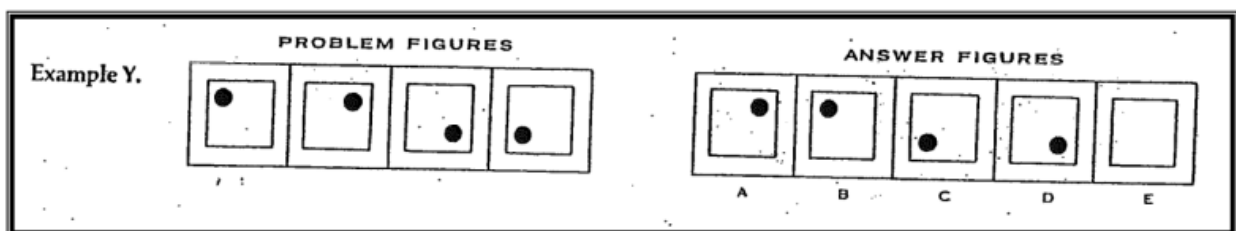
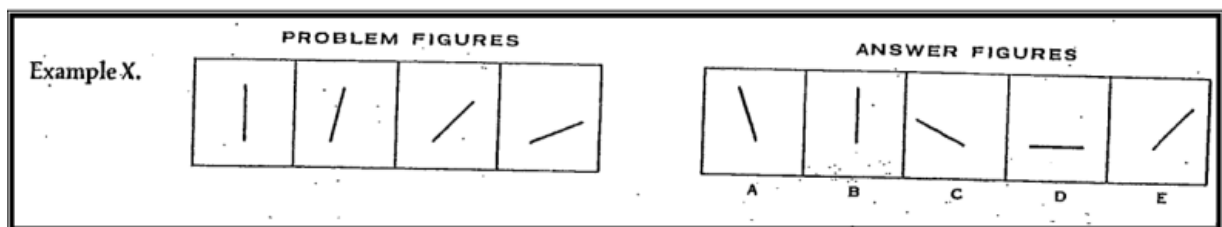
II) Subtract 40 from 50

A. 15    C. 26

B. 16    D. 8    E. None of These

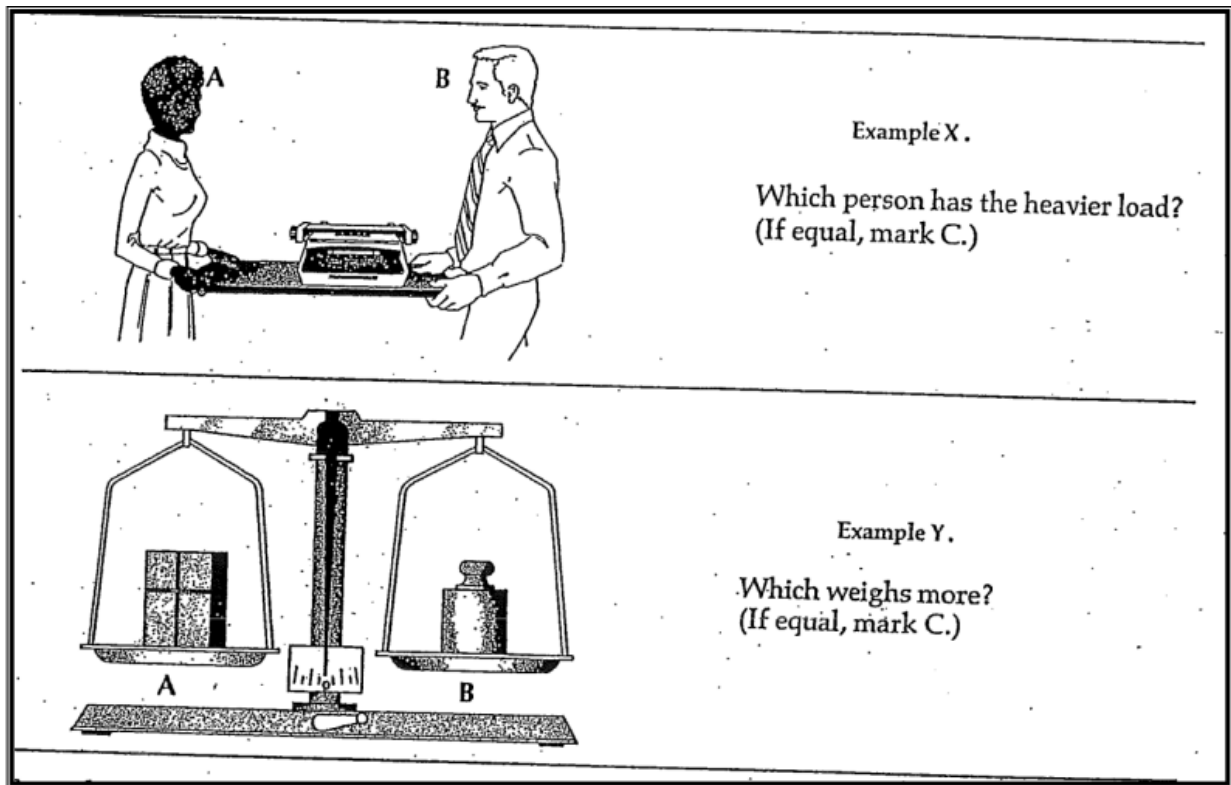
**Abstract Reasoning:** These tests evaluate the capacity to understand abstract concepts, patterns, and relationships without relying on specific prior knowledge.

*Example-* Select the answer figure that completes the series begun in the problem figures.



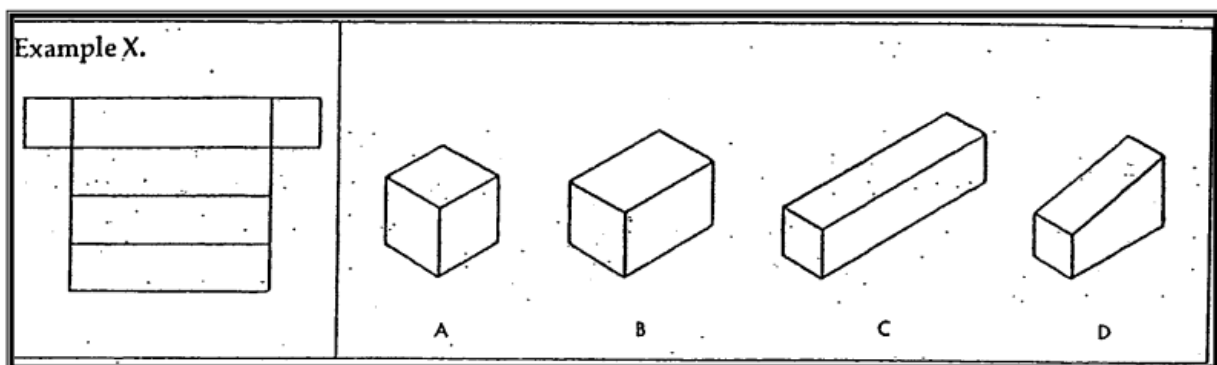
**Mechanical Reasoning:** These tests Assess the ability to understand and apply mechanical concepts and principles, often used in professions that involve working with machinery.

Example-

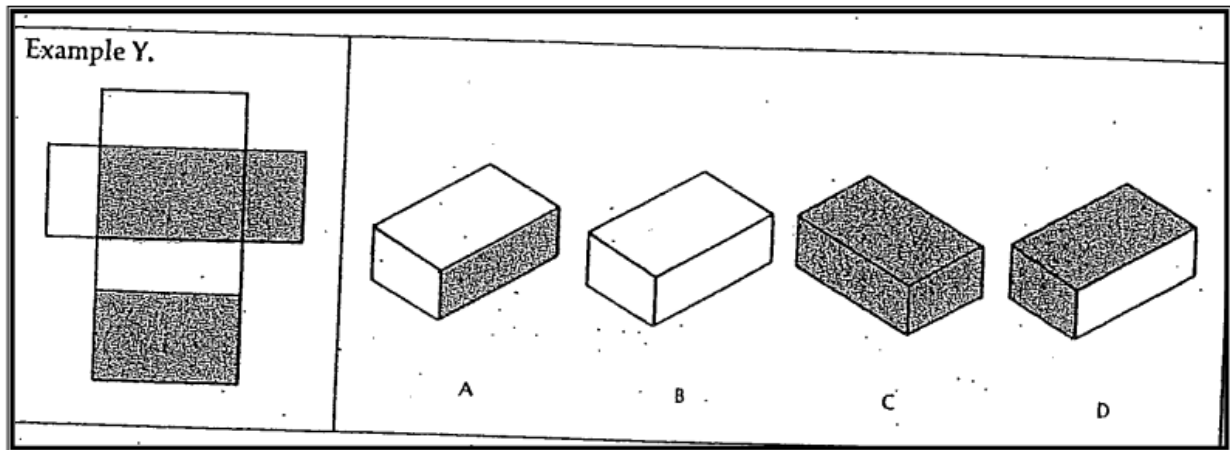


**Spatial Relation:** It measures an individual's ability to visualize and manipulate objects in space, often relevant in fields such as engineering and architecture.

*Example-* The test consists of 40 patterns that can be folded into figures. Five figures are shown for each pattern, and you must determine which of the figures can be made from the displayed pattern.







**Clerical Speed and Accuracy:** These tests intended to measure the speed of perception, memory retention and speed of response in simple perceptual tasks. The respondent must select the marked combination in the test booklet, then look for the same combination in a group of similar combinations in a separate answer sheet and underline it.

*Example-*In the test item, one of five combinations is underlined. Find and mark the corresponding combination on the answer sheet.

<u>AB</u>	AC	AD	AE	AF
aA	aB	BA	Ba	<u>Bb</u>
A7	7A	B7	<u>7B</u>	AB
Aa	Ba	<u>bA</u>	BA	Bb
3A	3B	<u>33</u>	B3	BB

(Test Items)

AC	AE	AF	<u>AB</u>	AD
BA	Ba	<u>Bb</u>	aA	Ab
<u>7B</u>	B7	AB	7A	A7
Aa	<u>bA</u>	bB	Ba	BA
BB	3B	B3	3A	<u>33</u>

(Sample of Answer Sheet)

**Language Usage Part I: Spelling and Language Usage Part II: Sentences-**These tests measure an individual's understanding and command of grammar, syntax, and other aspects of language structure. Both the forms spelling and sentences together provide a good estimate of an individual's ability to distinguish correct from incorrect language usage.

*Example- Spelling-* Indicate whether the spelling of each word is right or wrong.

- a. COW
- b. CALT
- c. PEN
- d. TEST

Sentences- Mark the one lettered part of the sentence contains an error and mark the corresponding letter. If there is no error mark N.

1. Was we / going to the / office/ next week / at all

A            B            C            D            E            N

2. Why/ are you / going /to/ the city?

A        B            C        D        E        N

All the tests of DAT except the Clerical Speed Accuracy test are power tests i.e. it provides respondents with sufficient time to attempt all items and express their true level of knowledge and ability. DAT Battery has been extensively used in India for several decades. It is considered useful for predicting academic success as well as for educational and vocational guidance, selection, and placement. The battery is ordinarily used for scholastic and industrial selection. The Indian reprint incorporates slight modifications, for a more appropriate application in Indian conditions.

***Self-check Exercise-5.2***

Describe three uses of DAT for education and guidance.

---

---

---

---

---

---

---

---

**General Aptitude Test Battery (GATB)**

Another popular aptitude test is GATB which was constructed by the United States of Employment Services to measure a wide range of occupational aptitude in 1930 and was again revised and modified in 1983. The GATB consists of 12 tests to measure nine types of aptitude.

The 12 tests of GATB are as follows.

- Name Comparison
- Computation
- Three-Dimensional Space
- Vocabulary
- Tool Matching
- Arithmetic Reasoning
- Form Matching
- Mark Making
- Place
- Turn
- Assemble
- Disassemble

The aptitudes tested under GATB, its purpose and the tests used to measure that aptitude are described below in the table.

**Table-5.1: Sub-tests of general aptitude test battery**

Symbol	Aptitude	Purpose	Test
<b>G</b>	General Intelligence	General learning ability, ability to grasp instructions and underlying principles	Vocabulary, Arithmetic Reasoning, Three-Dimensional Space
<b>V</b>	Verbal Aptitude	Ability to comprehend and interpret verbal communication, including words, paragraphs, concepts, and ideas.	Vocabulary
<b>N</b>	Numerical Aptitude	Ability to perform arithmetic operations with speed and precision	Computation, Arithmetic Reasoning

<b>S</b>	Spatial Aptitude	Ability to visualize and manipulate objects in space	Three-Dimensional Space
<b>P</b>	Form Perception	Ability to detect relevant detail in things and to perform visual comparisons and discriminations in pictorial or graphic content.	Tool Matching, Form Matching
<b>Q</b>	Clerical Perception	Ability to perceive detail, to observe differences in verbal and numerical material like tables, lists etc.	Name Comparison
<b>K</b>	Motor Coordination	Ability to make precise movements with speed and accuracy by coordinating eyes and hands or fingers rapidly and accurately	Mark Making
<b>F</b>	Finger Dexterity	Ability to move fingers, and manipulate small objects with fingers swiftly and precisely	Assemble, Disassemble
<b>M</b>	Manual Dexterity	Ability to move hands effortlessly and adeptly in placing and turning motions	Place, Turn

The abilities are generally divided into three categories: cognitive (G, V, N), perceptual (S, P, Q) and psychomotor (K, F, M). Approximately 2.5 hours is required to complete the whole test. This test is a great battery for vocational guidance of youth and adults.

---

## 15.6. MEASUREMENT OF INTELLIGENCE

---

Intelligence is a fundamental, complex, and multifaceted aspect of human cognition. It encompasses the ability to acquire, understand, and apply knowledge effectively. It goes beyond mere information retention, involving problem-solving skills, creativity, and adaptability to new situations. Intelligence is not limited to a single domain but can manifest in different ways, such as linguistic, logical-mathematical, spatial, musical, interpersonal, and intrapersonal intelligences etc. The meaning of intelligence has been explored and defined by various theorists, and the definition of intelligence has evolved over time, shaped by diverse theoretical perspectives. Some of the definitions of intelligence by different theorists are given below.

“Intelligence is the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment.” (Wechsler)

"Intelligence is the ability to learn from experience, adapt to new situations, understand and handle abstract concepts, and use knowledge to manipulate one's environment." (Robert Sternberg)

"Intelligence is the ability to solve problems, or to create products, that are valued within one or more cultural settings." (Howard Gardner)

"Intelligence, as the term is most widely used, refers to the general cognitive ability that includes reasoning, problem-solving, and the ability to learn from experience." (Charles Spearman)

"Intelligence is a mental capability that allows us to learn from experience, solve problems, and adapt to new situations." (John B. Carroll)

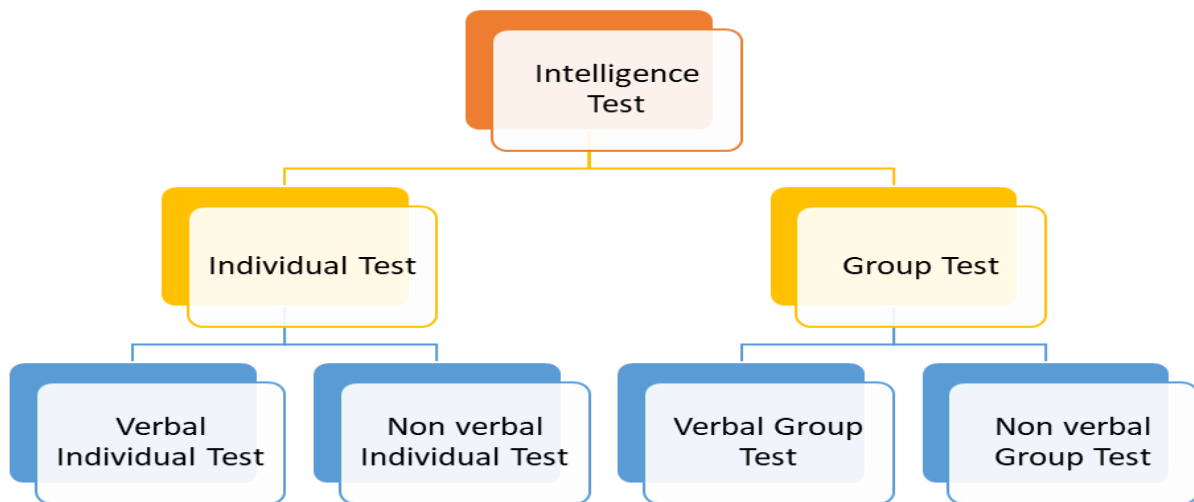
The nature of intelligence is a complex and multifaceted concept that has been the subject of extensive study and debate in psychology and cognitive science. However, some key aspects that contribute to the nature of intelligence are as follows:

- Intelligence is a multidimensional construct that can manifest in different ways.
- The nature of intelligence can be influenced by cultural factors. Different cultures may value and prioritize different cognitive skills, leading to variations in how intelligence is expressed and recognized.
- A fundamental aspect of intelligence is the ability to adapt to new situations, learn from experiences, and apply knowledge effectively
- Intelligence is dynamic in nature. Factors such as education, experiences, and continued learning can contribute to the development and enhancement of cognitive abilities.
- Individuals vary in their cognitive strengths and weaknesses, and the nature of intelligence recognizes and respects these individual differences.

### **Methods of Measuring Intelligence**

It is important to note that intelligence is a multidimensional construct, and no individual test can comprehensively encapsulate the diversity of human cognitive capabilities. The

measurement of intelligence involves the use of various tests designed to assess cognitive abilities. Intelligence tests aim to provide a standardized and objective measure of an individual's intellectual functioning. There are different types of intelligence tests, each focusing on various aspects of cognitive abilities. Some common approaches to measuring intelligence are described below.



**Verbal Intelligence Test-** Verbal intelligence tests assess an individual's ability to understand, analyze, and use language. They often involve tasks such as vocabulary, comprehension, and verbal reasoning. Some examples of verbal intelligence tests are given below.

**Wechsler Adult Intelligence Scale (WAIS)-** The Wechsler Adult Intelligence Scale is a test designed specifically for adults. It consists of 11 subtests, which are divided into two groups - verbal and performance. The verbal group includes six subtests: General Information, Similarities, Arithmetic Reasoning, Comprehension, Digit Span, and Vocabulary. The remaining five subtests are grouped under the performance scale.

**Binet-Simon intelligence scale-** The Binet-Simon scale, also referred to as the Binet-Simon Intelligence Scale, was created by French psychologists Alfred Binet and Théodore Simon during the early 1900s. It was developed to measure an individual's intellectual abilities and is recognized as the first modern intelligence test. This scale was designed to assist in identifying children who were experiencing difficulties in school so that they could receive additional support and interventions. The Binet-Simon tests are a set of 56 tasks and questions designed for children between the ages of 3 and 13 years, with the aim of creating a

scale to measure their intellectual abilities. The tests help to compare the child's performance with that of the average child of the same age.

**Army Alpha test-** The Army Alpha test was developed by Robert Yerkes and six other researchers during World War I to evaluate many US Army recruits. It was introduced in 1917 to provide a systematic method for assessing intellectual and emotional performance of soldiers. The test measures language ability, arithmetic ability, ability to follow instructions, and general knowledge. The scores of Army Alpha Test were used to ascertain a soldier's suitability for service, job classification, and potential for leadership positions.

**Non-Verbal Intelligence Test-** Non-verbal intelligence tests focus on assessing cognitive abilities without relying heavily on language. These tests measure spatial reasoning, pattern recognition, and problem-solving skills. **Some of the examples of Non-Verbal intelligence test are described in following paragraphs.**

**Raven's Progressive Matrices-** The Raven's Progressive Matrices is a nonverbal ability test that was initially developed by John C. Raven in 1936. This test is used to assess abstract reasoning and it features a progressive format in which questions get tougher as the test progresses. The objective is to identify the absent element in a pattern typically presented in the format of a matrix. Hence, it is called Raven's matrices. As the test is non-verbal, it is considered to reduce cultural bias. There are three versions designed for participants with varying skill levels, and the assessments can be conducted from the age of five to the elderly. These three tests are Raven's Standard Progressive Matrices, Raven's Colored Progressive Matrices, and Raven's Advanced Progressive Matrices.

**Army Beta Test-** Army Beta Test, the non-verbal equivalent of the Army Alpha Test was developed for the soldiers who are illiterate or speak a foreign language during World war I. Demonstration charts and pantomimes were used to convey instructions to test subjects. Another type of test in the beta used geometric designs, cut photos, etc., and required different principles for their construction.

**Individual Intelligence Test-** Individual intelligence tests are administered one-on-one by a trained examiner. This allows for personalized assessment, feedback, and observation of an individual's cognitive processes. The Stanford-Binet Intelligence Scales and the Wechsler Intelligence Scales (e.g., WAIS) are administered individually.

**Group Intelligence Test-** Group intelligence tests are designed to assess intelligence in a more time-efficient manner by testing multiple individuals simultaneously. They are typically administered to large groups. The Otis-Lennon School Ability Test (OLSAT) is an example of a group intelligence test commonly used in educational settings. Cognitive Abilities Test (CogAT): administered in group settings to assess cognitive abilities in children.

**Performance Intelligence Test-** Performance tests assess an individual's ability to apply knowledge and skills to real-world tasks. These tests often involve hands-on or practical components. Examples include the Performance Assessment of Contributions and Effectiveness (PACE) and the Assessment of Work Performance (AWP), Bhatia Battery test etc.

***Self-check Exercise-5.3***

Write the difference between individual and group intelligence tests.

---

---

---

---

---

---

---

---

---

**Concept of Mental Age (MA) and Intelligent Quotient (IQ)**

Interpretation of intelligence score requires understanding of mental age and chronological age. Mental age is a concept introduced by Alfred Binet to quantify a person's intellectual functioning in terms of the level typically associated with a specific chronological age. It represents the age at which an individual's performance on intelligence tests is comparable to the average performance of individuals in each age group. For example, if a 10-year-old child solves the problems meant for a 15-year child then he has a mental age of 15. Chronological age i.e. the real age of the child is 10.



**Intelligent Quotient (IQ)**- The concept of IQ was first developed by German psychologist William Stern in 1914. The Intelligence Quotient is a numerical measure to express an individual's intelligence relative to their chronological age.

*The formula is  $IQ = (Mental\ Age / Chronological\ Age) \times 100$*

Example: Lisa has a chronological age of 8 and a mental age of 10. Then what will be her IQ?

Answer:  $IQ = (Mental\ Age / Chronological\ Age) \times 100$

$$IQ = (8/10) \times 100 = 125$$

So, the IQ of Lisa will be 125.

---

## 15.7. MEASUREMENT OF ATTITUDES

---

Attitude refers to a psychological tendency or disposition to evaluate, react to, and behave towards people, objects, situations, or concepts in a particular way. It is a state of mental and emotional readiness to react to a situation, person, or thing because it encompasses a person's beliefs, feelings, and behavioral intentions concerning a specific target. For example, one's view regarding any sports, a subject or social issues, etc. Attitudes can be positive, negative, or neutral, and they play a crucial role in shaping an individual's thoughts, feelings, and actions. Following are some definitions of attitude by experts:

According to Frank Samuel Freeman, "An attitude is a dispositional readiness to respond to certain institutions, persons or objects in a consistent manner which has been learned and has become one's typical mode of response."

According to Louis Leon Thurstone, "An attitude denotes the total of man's inclinations and feelings, prejudice or bias, preconceived notions, ideas, fears, threats, and other any specific topic."

Anastasi defined attitude as "A tendency to react favorably or unfavorably towards a designated class of stimuli, such as a national or racial group, a custom or an institution."

According to Gordon Allport, "An attitude is a mental and neural state of readiness, organized through experience, exerting a directive or dynamic influence upon the individual's response to all objects and situations with which it is related."

It can be said that attitude is a state of opinion about some events/phenomenon/person in terms of favorable and unfavorable. It expresses the individual perception about issues and problems.

### **Method of Measuring Attitude**

Measuring attitudes in educational psychology is important for understanding students' beliefs, feelings, and opinions towards various aspects of education. Attitude scale is a very popular tool for measuring attitude. Mainly three dimensions are incorporated in an attitude scale:

- 1) Direction- Direction is an important aspect of an attitude that shows whether a person is favorable or unfavorable towards something. It shows positive or negative feelings about something.

- Example- I like mathematics. (Positive direction)

I do not like mathematics. (Negative Direction)

- 2) Degree- It shows the amount of liking or disliking attached to the feeling of a person toward something. A person may have different degrees of liking or disliking, which can be mild, moderate, strong, very strong, etc.
- 3) Intensity- It shows the strength of a feeling or the level of confidence of expression about something

Example- I am crazy about dogs. (This shows high intensity of an attitude.)

The two most frequently used scales for the measurement of social attitude are Thurstone scale or the method of equal appearing intervals developed by Thurstone and Likert scale or the method of summated ratings developed by Likert.

### **Thurstone Scale**

Thurstone's technique of scaling, also known as the Thurstone scaling method or the method of equal-appearing intervals, is a psychometric scaling method developed by Louis Leon Thurstone, a pioneer in the field of psychometrics in 1928. This technique is used to measure and assess the subjective intensity of attitudes, opinions, or characteristics that cannot be directly measured but can be indirectly assessed through a series of statements or items.


Here are the key steps involved in development of Thurstone scale:

- Initially, a pool of statements or items related to the construct of interest is generated. These items are designed to cover a wide range of attitudes, opinions, or characteristics that individuals may hold regarding the topic being studied.
- A panel of judges, typically experts in the field, is to be identified. The judges' role is to evaluate the statements or items and assign a numerical rating to each item generally from 1 to 11 based on their judgment of the item's intensity or relevance to the construct being measured. Here 1 represents the lowest intensity and 11 is the highest.
- After that, the median/mean and interquartile range of each item are calculated and arranged in ascending order of the median, and items having equal median are arranged in descending order of their quartile deviation. To understand it better let us take an example.
- Suppose there are a total of 90 items rated by experts.
- The median/mean of item no 42,65,88 is 1 and the interquartile range is 1.5,1.25 and 1 respectively.
- The median/mean of item no 10 and 28 is 2 and the interquartile range is 2.5 and 1 respectively.
- The median/mean of item no 06,31,14 and 57 is 3 and the interquartile range is 2,1.75,1.1 and 0 respectively.

The table below shows the arrangement of the above-mentioned items.

**Table-5.1 Arrangement of items in Thurston Scale**

Item No.	Median/ Mean	Interquartile Range
42	1	1.5
65	1	1.25
88	1	1
10	2	2.5
28	2	1
06	3	2



31	3	1.75
14	3	1.1
57	3	0

- After sorting all the items in a table, the final item selection process begins by choosing the item with the least interquartile range for each mean/median value. In the above table we have 3 mean/median values that are 1, 2 and 3. We must choose the item with the least Inter-quartile range for each mean/median value. For mean/median value 1, item no 88 is selected as its inter-quartile range is the lowest. For mean/median value 2 and 3 item no 28 and 57 can be chosen respectively for the same reason.
- Then the final statements are arranged in random order before administering the scale.
- The respondents tick marks the statements with which they agree and leave the rest.
- The weight of checked statements is added and divided by the number of statements checked.

### Likert Scale

Likert scale is the most popular type of attitude scale. It was developed by Rensis Likert in the year 1932. It is also known as summative rating scale. The scale contains several statements regarding the subject that one wants to study and the respondents must choose whether he/she is in favor of the given statement or not. It can be of 3-point, 5-point, 7 points etc.

Likert scale is the easiest to construct, the researcher constructs statements that reflect the main issue that is to be studied. Two types of statements appear on Likert scale. The first type of statements endorses a positive or favorable attitude towards the issue. The second type of statements endorse negative or unfavorable attitudes towards the issue. Equal number of favorable and unfavorable statements must be included in the scale. After developing the series of statements about the issue to be studied they are arranged randomly. No of options provided alongside each statement depends on how many pointer scales are being constructed. A five-point rating scale will have 5 options to choose from such as:

- Strongly agree (SA)
- Agree (A)

- Uncertain (U)
- Disagree (D)
- Strongly disagree (SD)

The numeric values of each statement are assigned differently for a positive statement and negative statement. For example, the point values for a positive statement are as typically as SA=5, A=4, U=3, D=2, SD=1 and for a negative statement the point values may be assigned as SA=1, A=2, U=3, D=4, SD=5.

*Example:*

- a) CWSN are entitled to equal educational opportunity as other students.

If a respondent strongly agrees (5) or agrees (4) for this statement then it indicates he has a positive attitude towards equal opportunity for Children with Special Needs (CWSN).

- b) CWSN are not entitled to equal educational opportunity as other students.

If a respondent strongly agrees (1) or agrees (2) for this statement then it indicates he has a negative attitude towards equal opportunity for CWSN.

After constructing the statements, they are administered to the students and collected back. The teacher then calculates each respondent's score by adding the values for each response. The scores chosen by the respondent for all the questions are then summed up to determine their attitude towards the question being studied. This rating scale is also known as the summative rating scale because it sums up the responses to all the questions.

### **Advantages of Attitude Scales**

- Attitude scales allow teachers/researchers to quantify subjective constructs such as attitudes and opinions. This quantification makes it easier to analyze and compare data, facilitating empirical research.
- Scales with a standardized set of questions or items, ensuring that all respondents are asked the same questions in the same way. This reduces the potential for bias in data collection.
- Well-designed attitude scales can be highly reliable, meaning they consistently measure the same construct across different administrations.
- With careful construction and validation, attitude scales can be valid, meaning they accurately measure the construct they are intended to measure.

- Scales are efficient data collection tools. They allow researchers to gather a large amount of information from respondents in a relatively short amount of time.

**Disadvantages of Attitude Scales:**

- Respondents may provide socially desirable or biased responses, especially when sensitive topics are involved. This can lead to inaccurate measurements if respondents are not completely honest.
- Scales provide a structured framework, which may limit the depth of understanding. They may not capture the richness and complexity of attitudes in real-life contexts.
- Constructing a valid scale can be challenging. If the scale does not measure the intended construct accurately, the results will be invalid. This can happen due to poor item selection, wording, or inappropriate response options.
- Some scales use fixed response options that might not capture the full range of attitudes. Respondents may feel constrained by the available options.

In conclusion, attitude scales are valuable tools for quantifying and comparing attitudes and opinions, but they have limitations and potential pitfalls that teachers/researchers must consider. Careful design, validation, and interpretation are essential to ensure the accuracy and reliability of measurements obtained from attitude scales. Teacher can prepare own attitude scale to study the opinion of students towards different issues which can be used for extension programmes in school.

***Self-check Exercise-5.4***

Develop an attitude scale to measure the attitude of students towards the use of Mobile for learning.

---

---

---

---

---

---

---

---

**15.8. MEASUREMENT OF INTEREST AND SKILLS**

---

**Interest**

Interest is a dynamic and multifaceted construct that plays a fundamental role in education. Interest can be defined as an individual's psychological state characterized by a positive emotional response to a particular subject, activity, or domain. It is a multifaceted construct that encompasses both affective and cognitive aspects. When students are interested in a topic, they are more likely to be engaged, motivated, and eager to learn, which can lead to improved academic performance and overall educational experiences.

Jones defines interest as “a feeling of liking, associated with a reaction either actual or imaginary to a specific thing or situation.”

Super writes “interest is the product of interaction between inherited aptitude and endocrine factors on the one hand and opportunity and social evaluation on the other.”

According to Bingham “an interest is a tendency to become absorbed in an experience and to continue it.”

### **Methods of Measuring Interest**

Understanding and measuring students' interests is crucial for educators and researchers in the field of education. It provides valuable insights into what motivates and engages learners, guiding the design of effective instruction, curriculum, and assessments. Some popular methods of measuring interest are discussed below:

- **Self-Report Surveys and Questionnaires:** Self-report surveys and questionnaires are one of the most widely used methods for measuring interest. Students are asked to respond to a series of questions or statements about their preferences, likes, and dislikes in specific areas. Responses are typically collected on a Likert scale, where individuals rate their level of interest. These surveys are easy to administer and provide insights into individual interests, but they may be subject to response bias.
- **Observation:** Observational methods involve watching individuals in real-life situations to assess their interest and engagement. This can be particularly valuable in educational settings where educators or researchers observe students' behavior, facial expressions, and interactions during lessons or activities to gauge their level of interest.
- **Interest Inventory:** Interest inventories are structured assessments that evaluate individuals' interests across various domains, such as careers, hobbies, or academic

subjects. Respondents typically rate their interest in different areas, and the results help identify potential areas of interest.

Interest inventories are structured assessments that evaluate individuals' interests across various domains, such as careers, hobbies, or academic subjects. Respondents typically rate their interest in different areas, and the results help identify potential areas of interest. These assessments can provide valuable insights into potential career paths and academic pursuits. Some very popular interest inventories are discussed below:

**Strong Vocational Interest Blank (SVIB)**-The "Strong Vocational Interest Blank" (SVIB) is a career assessment tool created by Edward Kellogg Strong, Jr. in 1927. Its purpose is to help individuals identify their interests and make informed career choices. The test has undergone revisions and expansions over the years, but its core features remain essential to understanding vocational preferences. Initially, the test was only for men, but a version for women was introduced in 1933 to promote inclusivity. The primary goal of the SVIB is to assess an individual's interests, defined as the desire to explore and learn about something or someone. The test is suitable for individuals aged 17 and above. It comprises 420 items, each requiring responses categorized as Like, Indifferent, or Dislike. The items are distributed across eight sections, which include Occupations, School Subjects, Activities, Leisure Activities, Types of People, Preferences between Paired Activities, Pairing between Four Items of Work, and Self-Descriptive Answers.

**Kuder Interest Inventories**- The Kuder Interest Inventory was designed by G. Frederic Kuder to help measure interest from different angles and purposes. It was designed for students of grade 9 and above in the form of three preference record i.e. Vocational, Occupational, Personal. The Kuder Vocational Preference Record contains 10 scales such as outdoor, mechanical, Computational, Scientific, persuasive, artistic, literary, musical, social service, and clerical. The forced-choice triad items method is used to find out the preference of the respondent. It presents respondents with sets of three occupations and requires them to choose the one that best reflects their preferences or opinions. The forced choice nature of these items makes respondents prioritize or rank their choices within each triad.

The second preference record i.e. Kuder Occupational Interest Inventory covers a wide variety of occupations to choose from such as farmer, newspaper editor, minister, mechanical engineer, architect, truck driver, lawyer etc.



The third one is a personality inventory that intends to evaluate five very broad characteristics of behavior. The characteristics are being active in a group, familiar and stable situation, working with ideas, avoiding conflicts, and directing others. The score in each characteristic suggests the respondent's preference. A high score suggests high preference and a low score suggests low preference.

**Thurstone Interest Schedule**-The Thurstone Interest Schedule was developed by L.L. Thurstone. It is a schedule that asks individuals to express their preferences for various occupations. The occupations are presented in pairs, and the participants are asked to indicate their preference by checking the preferred option. They should assume that there is no difference in income or prestige between the two options. If they like both options, they should circle both, and if they dislike both options, they should cross them out. There is no fixed time limit to complete the schedule but generally it takes 10 minutes to complete.

#### ***Self-check Exercise-5.5***

Explore the role of observation in assessing students' interest, highlighting its advantages and potential challenges in educational settings.

---

---

---

---

---

---

---

---

## **15.9. MEASURING SKILLS**

---

The traditional paradigm of education has shifted from a focus solely on academic achievement to a more comprehensive approach that acknowledges the importance of diverse skills. Skills are integral components of an individual's ability to perform tasks, solve problems, and navigate various aspects of life. In the field of education, the assessment of

skills goes beyond traditional academic measures, recognizing the importance of a holistic approach to student development.

A skill is a learned ability or capacity to carry out a task with expertise, efficiency, and effectiveness. Skills can be developed through training, practice, and experience, and they are often specific to activities or domains like cognitive, technical, interpersonal, and soft skills. In the educational context, skills extend beyond academic knowledge, encompassing a wide range of competencies that prepare students for success in both academic and real-world settings.

### **Essential Skills in Education**

In today's dynamic and interconnected world, students need to be equipped with a wide range of competencies to navigate complex challenges and opportunities. Let us discuss some skills that are essential in the field of education.

**Academic skills-** Academic skills form the core of traditional education, encompassing literacy, numeracy, critical thinking, and problem-solving. Proficiency in these areas provides a strong foundation for further learning and intellectual development. Standardized tests and assessments play a significant role in evaluating academic skills, providing a benchmark for student performance.

**Communication skills-** Effective communication is a cornerstone of success in both academic and real-world settings. Assessing verbal and written communication skills, along with active listening abilities, ensures that students can express themselves clearly, collaborate with others, and comprehend information from various sources. Beyond conventional exams, projects and presentations offer insights into students' communicative competence.

**Socio-emotional skills-** Education goes beyond the transmission of knowledge; it includes the cultivation of social and emotional intelligence. Skills such as teamwork, collaboration, empathy, and self-management contribute to a positive and supportive learning environment. Classroom observations and self-assessment tools are valuable in gauging the development of these interpersonal and intrapersonal skills.

**Creativity and Critical Thinking Skills-** Encouraging creativity and critical thinking fosters innovation and problem-solving abilities. Assessments that require students to think beyond memorization and apply knowledge in novel ways help identify their capacity for creative expression and analytical reasoning.

**Life Skills-** Life skills foster the ability to work effectively in teams, promoting cooperation and collaboration. These skills prepare students to adapt to changing circumstances, promoting flexibility and a positive response to new challenges. Moreover, life skill education contributes significantly to career readiness by instilling job interview skills, resume writing, and professional etiquette. It fosters a holistic approach to health and well-being, promoting healthy lifestyle choices and stress management.

### **Methods of measuring skills**

#### **Observation**

Observations conducted directly in the classroom setting offer a nuanced and real-time understanding of students' non-academic skills, encompassing social, emotional, and behavioral competencies. This method allows educators to gain valuable insights into the multifaceted aspects of a student's development beyond mere academic achievements. By keenly observing student interactions, educators can assess and analyze their abilities in areas crucial for holistic growth.

One of the key dimensions assessed through classroom observations is social competence. Educators can witness how students navigate social dynamics, engage with their peers, and form relationships. This includes the observation of collaborative efforts during group activities, discussions, and project work. The ability to work effectively within a team, share ideas, and contribute constructively to group tasks becomes evident through direct observation. This insight goes beyond what traditional assessments might capture, providing a more authentic portrayal of a student's social skills.

Moreover, classroom observations shed light on students' communication skills. Educators can witness firsthand how students express themselves, articulate their thoughts, and actively participate in discussions. Effective verbal communication, clarity of expression, and the ability to engage in meaningful dialogue are observable indicators that contribute to a comprehensive understanding of a student's communication competencies. Non-verbal

communication, such as body language and facial expressions, is also assessed, adding another layer to the observation process.

Behavioral competencies, including leadership and adaptability, are particularly well-suited for assessment through direct observation. In the classroom, educators can observe instances where students take on leadership roles, guide their peers, or demonstrate initiative. These observations provide insights into leadership potential and the ability to influence and inspire others positively. Additionally, the adaptability of students to different learning environments, their response to challenges, and their ability to adjust to unforeseen circumstances can be gauged in real-time, offering a more accurate reflection of their adaptability.

The real-time nature of classroom observations enhances the authenticity of the assessment. Unlike standardized tests or retrospective self-reporting, observations capture students' behaviors and interactions as they unfold naturally. This immediacy allows educators to identify strengths and areas for improvement promptly. It also facilitates the identification of patterns of behavior over time, enabling a more holistic and continuous evaluation of non-academic skills.

Furthermore, the direct observation method provides educators with opportunities to offer immediate feedback. Whether addressing a particular behavior, acknowledging a positive interaction, or suggesting improvements, timely feedback can positively impact students' awareness and growth in their non-academic skills. This feedback loop contributes to the iterative nature of skill development, reinforcing positive behaviors and guiding students toward continuous improvement.

In conclusion, direct observations in the classroom are a valuable and dynamic method for assessing students' non-academic skills. Through this approach, educators gain a nuanced understanding of social, emotional, and behavioral competencies, including teamwork, communication, leadership, and adaptability. The real-time nature of observations enhances the authenticity of the assessment, offering a holistic and continuous evaluation that goes beyond traditional measures of academic success. This method not only informs educators but also provides students with actionable feedback for their ongoing personal and interpersonal development.

## **Project-Based Learning**

Project-Based Learning (PBL) is an educational approach that immerses students in hands-on projects, fostering active engagement and facilitating deep learning experiences. Emphasizing the real-world application of knowledge, collaboration, critical thinking, and problem-solving skills, PBL transcends traditional teaching methods. Assessments within the context of Project-Based Learning are thoughtfully designed to extend beyond conventional measures, aiming for a comprehensive evaluation that encompasses various facets of student performance.

At the core of PBL are hands-on projects that initiate an exploration into real-world scenarios or issues, compelling students to delve into specific topics or problems. This immersive learning method allows students to actively explore, experiment, and apply theoretical knowledge to practical situations, creating an environment that mirrors authentic and complex scenarios.

Critical thinking and independent problem-solving are inherent components of Project-Based Learning. The nature of the projects often demands that students analyze information, make decisions, and devise solutions autonomously. Assessments in PBL focus on the depth of critical thinking demonstrated and the effectiveness of the solutions proposed during the project, thereby gauging the student's ability to apply learned concepts to real-world challenges.

Assessments within the realm of Project-Based Learning are deliberately multifaceted, capturing various dimensions of student performance. Unlike traditional assessments that predominantly focus on the final product, PBL assessments consider the entire project lifecycle. This holistic approach involves evaluating the planning process, research, ideation, and the iterative nature of project development. Recognizing the importance of the learning process itself, PBL assessments include process evaluation. Students are assessed on how well they manage their time, set goals, and navigate challenges throughout the project. This acknowledgment of the learning journey ensures that students are developing essential skills beyond the specific content of the project.

Project-based learning, as a method to measure essential skills, goes beyond traditional assessments by providing a holistic and authentic evaluation of students' capabilities. By integrating these skill assessments into real-world projects, it prepares students for success in

an ever-evolving and dynamic world, where a combination of cognitive and non-cognitive skills is crucial.

***Self-check exercise-5.6***

Design a project-based learning activity for students that integrates multiple skills, such as critical thinking, communication, teamwork, and problem-solving.

---

---

---

---

---

**Portfolio**

Portfolios serve as dynamic and comprehensive compilations that document a student's academic and personal journey, offering a multifaceted representation of their growth and proficiency across various skills over time. These collections, carefully curated and thoughtfully presented, encapsulate not only the tangible outcomes of a student's learning but also the evolution of their critical thinking, creativity, and application of knowledge.

The inclusion of written assignments within a portfolio offers insight into a student's ability to articulate thoughts, convey ideas effectively, and demonstrate mastery of academic content. These pieces not only showcase proficiency in the subject matter but also highlight communication skills, analytical thinking, and the application of theoretical knowledge to practical contexts. Project is another integral component of a portfolio, providing a tangible representation of a student's ability to apply classroom learning to real-world scenarios. Whether it be a science experiment, a research project, or a collaborative endeavor, projects illustrate not only academic competencies but also critical thinking, problem-solving, and teamwork. They offer a glimpse into a student's capacity to integrate knowledge across disciplines and to approach challenges with creativity.

Reflections, often included in portfolios, provide a metacognitive dimension to the collection. Students can articulate their learning experiences, personal growth, and insights gained through various academic and extracurricular activities. Reflections showcase self-awareness, the ability to learn from experiences, and a commitment to continuous improvement. This

element of the portfolio fosters a deeper understanding of the learning process and allows educators to gauge a student's capacity for self-assessment and reflective thinking. Extracurricular activities, documented in a portfolio, extend beyond the classroom, illustrating a student's engagement with the broader learning environment. The inclusion of evidence from clubs, sports, community service, or leadership roles provides a holistic view of a student's skills, including leadership, teamwork, communication, and adaptability. This demonstrates a student's commitment to personal development beyond academic pursuits.

The significance of portfolios lies in its ability to offer educators a panoramic view of a student's development. They serve as holistic assessment tools, allowing for a more comprehensive evaluation of essential skills that go beyond traditional measures of academic success. By examining the various components of a portfolio, educators can gain insights into a student's strengths, areas for improvement, and the interplay of skills across different contexts.

Portfolios are dynamic repositories that capture the essence of a student's educational journey, showcasing growth and proficiency in a myriad of skills. They transcend traditional assessments by providing a holistic view of a student's development, encompassing academic achievements, critical thinking, creativity, communication skills, and engagement in extracurricular activities. As a tool for assessment and reflection, portfolios empower both educators and students to appreciate the richness and diversity of the learning experience. It is very useful for teachers to monitor and assess the development of skills among students.

---

### **15.13. SUMMARY**

---

- Measurement, the process of associating numbers with physical quantities and phenomena.
- Achievement is very important for formal education as it determines the effectiveness of educational machinery. It refers to the level of knowledge, skills, and competencies that individuals have acquired through their educational experiences.
- Achievement Test-Achievement test is one of the most popular methods of measuring achievement. It is a commonly used instruments to measure an individual's knowledge, skills, or proficiency in a specific subject or set of subjects.

- Teacher-made achievement tests are created and used by educators for their classroom purpose. These assessments are tailored to align with the specific content covered during instruction, making it highly context-dependent
- Standardized achievement tests are developed externally by testing organizations or educational experts. These assessments are designed to be uniform, administered, and scored consistently across a broad and diverse population of students
- Aptitude refers to a person's inherent or natural ability to excel or perform well in a specific area or task. It is a quality or potential that individuals possess, often from birth or early in life, which enables them to acquire certain skills, knowledge, or talents more easily and effectively than others.
- Attitude refers to a psychological tendency or disposition to evaluate, react to, and behave towards people, objects, situations, or concepts in a particular way
- Likert scale is the most popular type of attitude scale. It was developed by Rensis Likert in the year 1932. It is also known as summative rating scale.
- Interest can be defined as an individual's psychological state characterized by a positive emotional response to a particular subject, activity, or domain. It is a multifaceted construct that encompasses both affective and cognitive aspects.
- Interest inventories are structured assessments that evaluate individuals' interests across various domains, such as careers, hobbies, or academic subjects.
- Strong Vocational Interest Blank (SVIB)-The "Strong Vocational Interest Blank" (SVIB) is a career assessment tool created by Edward Kellogg Strong, Jr. in 1927. Its purpose is to help individuals identify their interests and make informed career choices.
- The Kuder Interest Inventory was designed by G. Frederic Kuder to help measure interest from different angles and purposes. It was designed for students of grade 9 and above in the form of three preference record i.e. Vocational, Occupational, Personal.
- Project-Based Learning (PBL) is an educational approach that immerses students in hands-on projects, fostering active engagement and facilitating deep learning experiences. Emphasizing the real-world application of knowledge, collaboration, critical thinking, and problem-solving skills, PBL transcends traditional teaching methods.



- Portfolios serve as dynamic and comprehensive compilations that document a student's academic and personal journey, offering a multifaceted representation of their growth and proficiency across various skills over time

---

#### 15.14. UNIT END EXERCISE

---

- Define achievement? Explain about different methods of measuring achievement.
- Define attitude? Discuss about different methods of measuring aptitude.
- Define Intelligence? Explain about different methods of measuring intelligence.
- Define Interest? Discuss about different methods of measuring interest.
- What are different methods of measuring skills?

---

#### 15.15. FURTHER READING

---

Gronlund, N. E. (1965). *Measurement and evaluation in teaching*.  
<http://ci.nii.ac.jp/ncid/BA12623208>

Goswami, M. (2013). *Measurement and Evaluation in Psychology and Education*  
Lee, W. Y. (2010). *Assessment and evaluation in higher education*.  
<http://ci.nii.ac.jp/ncid/BB11596810>

Patel, R. N. (2014). *Educational evaluation theory and practice*.  
<http://ci.nii.ac.jp/ncid/BA29677030>

---

## **BLOCK 04:**

### **CHARACTERISTICS OF A GOOD TEST**

---

**Unit 16:** Validity-concept, types and methods of validation

**Unit 17:** Reliability- concept and methods of estimating reliability

**Unit 18:** Objectivity- concept and methods of estimating objectivity

**Unit 19:** Usability- concept

**Unit 20:** Usability- factors ensuring usability

## **UNIT -16**

### **NEW TRENDS OF EVALUATION**

## **STRUCTURE**

- **Learning Objectives**
- **Introduction**
- **New Trends of Evaluation**
- **Grading**
- **Summary**
- **Unit End Exercise**
- **Suggestion for Further Reading**

---

## **LEARNING OBJECTIVES**

---

After reading this unit, learners shall be able to;

- Define the new trends of evaluation system.
- Explain the Grading system and semester system
- Elaborate the continuous internal assessment and use of question bank
- Elaborate the use of computers in evaluation

---

## **INTRODUCTION**

---

Evaluation should serve the purpose of formative evaluation, providing instant feedback, outcome information, a diagnosis, and corrective action. One recent trend to assess a student's performance is through grading. Evaluation is a concept that has emerged as a prominent process of assessing, testing and measuring. Its main objective is Qualitative Improvement. Evaluation is a process of making value judgements over a level of performance or achievement.

Curriculum trends include integrating coding and computer science into various subjects, empowering students with the ability to understand and create technology. This integration allows students to develop problem-solving and critical-thinking skills while also fostering creativity and innovation.

---

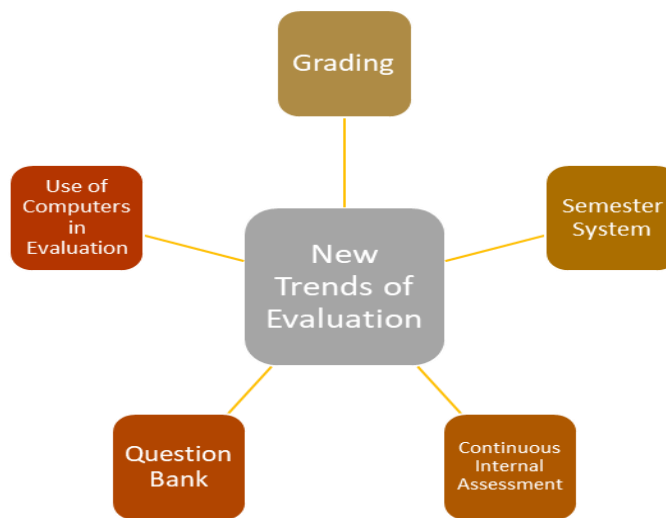
## **NEW TRENDS OF EVALUATION**

---

Education is a dynamic field that continually adapts to evolving societal needs, technological advancements, and pedagogical shifts. Central to this evolution is the way educators assess and evaluate students' progress and performance. Traditional evaluation methods, such as standardized testing and grades, have been supplemented and, in many cases, supplanted by

innovative approaches that harness the power of technology, adapt to diverse learning styles, and emphasize the development of real-world skills.

This section aims to discuss the new trends in evaluation in education, focusing on grading systems, the semester system, continuous internal assessment, the development of question banks, and the use of computers in evaluation. These trends are reshaping the way we assess and measure student performance, providing a more comprehensive and holistic view of their abilities and progress.



---

## GRADING

---

The word 'grade' originates from the Latin term '*gradus*' which means 'step'. Grading is a process used to categorize subjects based on predetermined standards. In an educational setting, grading is a way to communicate the level of achievement of students. It involves using a set of symbols that are clearly defined and understood by students, teachers, parents, and others involved. Without clear understanding of these symbols, the purpose of awarding grades is lost. It is crucial to define the meaning of each grading symbol while developing the grading system. Examiners must adhere to the specified system of grading. However, this does not limit the examiner's autonomy to determine the grade awarded to a student. A well-introduced grading system can not only allow for comparison of students' performance but also indicate the quality of performance in relation to the effort and knowledge acquired at the end of the course.

### Functions of Grades

Grades serve multiple purposes. Firstly, it shows how well a learner has achieved the instructional goals. This information is valuable not only for the learner and teacher but also for parents. Secondly, grades provide a permanent record of a learner's development, which can be useful for higher education institutions and prospective employers. Thirdly, grades help schools make decisions about placement and promotion. Fourthly, grades can assist in reviewing teaching strategies and curricular appropriateness. Additionally, grades factor into calculating a student's Grade Point Average (GPA) for awarding merit scholarships at many higher education institutions.

### **Methods of Assigning Grades**

Depending upon the reference point, grading can be done in a variety of ways. Based on the approach, grading can be of two types that is direct and indirect grading. Also based on the 'standard of judgment', the grading may be classified as *absolute grading and relative grading*.

**Direct grading-** The method of direct grading involves assessing the performance of examinees in qualitative terms and expressing the examiner's impression using letter grades. This approach is applicable to both cognitive and non-cognitive educational achievements. Direct grading is particularly useful for assessing non-cognitive learning outcomes. It is recommended to evaluate and report non-cognitive traits separately in terms of letter grades. The grading scale used should be determined based on the nature and quality of the attribute being evaluated, such as a three-point or five-point scale. Direct grading has the advantage of reducing inter-examiner variability and is easier to use than other methods. Nonetheless, it lacks clarity and diagnostic significance, and may not promote competition to the desired extent.

**Indirect grading-** In the indirect grading process, grades are not awarded directly. Rather, pupils are given marks, which are subsequently transformed into letter grades using various approaches while keeping the aim of assessment in mind. First marks are awarded for each question as usual. This procedure for awarding marks for individual questions is done using a prescribed marking scheme. Once marks are awarded to each question, the final score is calculated by adding up individual marks. Then the marks are converted into grades which can be done in two ways.

- *Absolute Grading*
- *Relative Grading*

**Absolute Grading:**

In this process of Grading, marks are given on a 100- or 101-point scale are converted into 5,7,9- or 11-point scale. For each grade, some fixed range of scores determined in advance. The score a student obtain in a paper are converted into their corresponding grade based on this predetermined range. It is also known as Criterion Referenced Grading or Reflect Absolute Performance. Let us consider some examples to understand the concept of Absolute Grading.

Score	Category Description	Grade
75% and above	Distinction	A
60% to <75%	First Division	B
45% to <60%	Second Division	C
33% to <45%	Third Division	D
Below 33%	Unsatisfactory	E

These categories and corresponding grades are associated with predefined standards.

*For example:* Those who score between 91 and 100 marks may be given ‘A’ grade. Those who score between 71 and 90 marks may be given ‘B’ grade and so on. If a student scores 95 marks, s/he gets an ‘A’ grade but if the score is 73, s/he gets a ‘B’ grade.

In Absolute Grading, the grades of a student are independent of the grades obtained by other students. Thus, it can be said that Absolute Grading is a type of Criterion-referenced grading which uses absolute standards. In this type of grading the teacher does not have advanced information on how many students will pass or fail the test.

**Relative Grading-**

In this process of grading, the grade range is not determined or fixed in advance. It is also known as Norm-Referenced Grading or Reflect Relative Performance. Instead, it varies according to the relative position of a student within a group. The performance of a student is compared with that of their peers. It is based on the premise that if the results of an evaluation are represented on a graph it will be in the form of a Normal Probability Curve (extends from  $-\infty$  to  $+\infty$ ). The number of students falling into each category is fixed based on Normal

Distribution which is a statistical principle. The same percentage of cases lie below and above the mean and the distance is measured in standard deviation units( $\sigma$ ). This gives the relative ranking of an individual. The normal distribution principle presupposes that the performance of individuals in a sizable group adheres to a specific pattern, which can be illustrated through a normal curve. This type of grading can be done on the curve.

*For example:* In a test the top 10 students may be awarded 'A' grade. The 'A' grade in one subject might be quite different from the 'A' grade in another subject. A teacher may award 'A' grade to students who scored 95 marks or above in Mathematics but for language subjects, the same teacher may place students who score 85 marks or above in 'A' grade.

***Self-check Exercise-5.7***

Differentiate between direct grading and indirect grading.

---

---

---

---

---

---

---

**Merits of Grading**

- It allows the assessment of examinees' achievements in different subjects separately.
- It minimizes errors of measurement, facilitating easy comparison of achievements among different students.
- Grading system enables inter-subject comparison of the same examinee and inter-examinee comparison within a specific subject.
- It shifts emphasis away from marks, promoting a focus on the grading system.
- The grading system is more significant and practical when contrasted with the marking system.
- The system is particularly beneficial for both low achievers and high achievers.
- It lessens anxiety and complexities among students by reducing unhealthy competitions.

## Demerits of Grading

- There is an absence of agreement among educators regarding the scale's points and corresponding marks.
- The grading system exhibits high sensitivity; transitioning from 70 to 75 is more fluid in marking than in grading.
- Subjectivity in evaluation, akin to the marking system, is possible in grading.
- Converting marks into grades is straightforward, but the reverse process is challenging.
- Lack of consistency in grading leads to confusion and difficulties in result interpretation.

Grading is one of the widely accepted ways of declaring results at all levels of education from school to universities. It has replaced the old marking system which created stress and anxiety among learners. Grading system has been accepted in Indian school boards such as CBSE, State Boards for the publication of results.

---

## SUMMARY

School measures students on different traits such as aptitude, intelligence, attitudes, skills along with achievement in different subjects. Achievement of students can be measured by essay test and objective test, teacher made test and standardized test. Teacher made-test are suitable for school as they are prepared by the teachers based on the school curriculum. Aptitude can be of different types which require a test battery like Differential Aptitude Testing (DAT). Aptitude test can indicate the students' suitability for a future course or career. Mental ability or intelligence can be measured by verbal, non-verbal and performance tests, each having its advantages and limitations. Intelligence tests can be administered individually or in groups. Other psychological traits like attitude, interest and skills can be measured by scales, checklists, and observations.

The evaluation system has been changing with time and pedagogical & educational advancement. The recent trend in evaluation is grading, semester examination, internal assessment, question banks and use of computers in evaluation. The grading can be of absolute or relative in nature. Semester examination has simplified the assessment process by making it as internal part of education process. These innovations have made the evaluation system systematic, objective, unbiased, and faster.



---

**UNIT END EXERCISE**

---

1. Differentiate between Teacher-made Achievement Test and Standardized Achievement Test.
2. A child of 13 years scores 88 in an IQ test. Calculate his mental age.
3. Create a test paper with a mix of questions covering different aptitude areas.
4. Compare Thurstone scale and Likert scale, highlighting the differences in their methods and applications for measuring social attitudes.
5. Provide details about some other interest inventories including its name, purpose, target audience, and any unique features it offers for assessing interests?
7. Conduct a brief classroom observation and identify non-academic skills demonstrated by students. Provide examples of social, emotional, and behavioral competencies observed.
8. Brainstorm creative ways teachers can use question banks beyond traditional examinations, such as in classroom activities, formative assessments, or as a resource for personalized learning.
9. Develop a comparative analysis between direct grading and indirect grading. Highlight the advantages and disadvantages of each method. Provide scenarios where each grading method might be more suitable.
10. Reflect on the future trends and advancements in computer-based evaluation. Discuss emerging technologies and their potential influence on assessment methods.

---

**FURTHER READINGS**

---

- Gronlund, N. E. (1965). *Measurement and evaluation in teaching*.  
<http://ci.nii.ac.jp/ncid/BA12623208>
- Goswami, M. (2013). *Measurement and Evaluation in Psychology and Education*
- Lee, W. Y. (2010). *Assessment and evaluation in higher education*.  
<http://ci.nii.ac.jp/ncid/BB11596810>
- Patel, R. N. (2014). *Educational evaluation theory and practice*.  
<http://ci.nii.ac.jp/ncid/BA29677030>
- Kochhar S.K. (1984). *Guidance and Counseling in Colleges and Universities*.

## UNIT – 17

### SEMESTER SYSTEM

## STRUCTURE

- **Learning Objectives**
- **Introduction**
- **New Trends of Evaluation**
- **Grading**
- **Summary**
- **Unit End Exercise**
- **Suggestion for Further Reading**

---

## LEARNING OBJECTIVES

---

After reading this unit, learners shall be able to;

- Define the new trends of evaluation system.
- Explain the semester system
- Elaborate the key features, merits and demerits semester system

## INTRODUCTION

The **semester system** is a widely adopted academic structure used in educational institutions to organize and divide the academic year into distinct periods for instructional and assessment purposes. Typically, a semester lasts around 15–18 weeks, and it is designed to give students enough time to cover specific portions of the curriculum before they are evaluated and move on to the next phase of learning. The semester system contrasts with the **annual system**, which divides the academic year into a single long-term period, often without breaks between major evaluations.

In the semester system, the academic year is divided into two primary terms: the **Fall Semester** (often starting in August or September) and the **Spring Semester** (usually starting in January). Some institutions may also have a **Summer Semester** or **Winter Term**, which are optional or shorter periods offering more specialized or intensive courses.

---

## SEMESTER SYSTEM

---

The semester system was created to improve upon the annual examination system. Under this system, a programme of study is divided into equal parts based on months, and an examination is conducted after the completion of each part. For example, a two-year programme would have four semesters, while a three-year programme would be divided into

six semesters. If student fails in one subject in one semester, they are not declared to have failed. Instead, they are allowed to continue to the next semester and re-study the subject. They can then retake the exam in that subject in that semester. The semester system aims to reduce stress and strain among students by integrating examinations into the daily routine. It also broadens the outlook of students and instills in them a sense of responsibility and confidence. The semester system is very dynamic, engaging both faculty and students throughout the year in academic activity. It reduces the burden of examinations. Both systems have their merits and demerits. Let us explore the pros and cons of the semester system.

### **Key Features of the Semester System**

1. **Division of the Academic Year:** The semester system divides the academic year into two or three segments, each typically lasting between 15 to 18 weeks. These segments are referred to as **semesters** (fall and spring) or **terms** (quarter, trimester, or shorter periods).
2. **Course Load and Structure:** Each semester typically consists of a set number of credit hours, usually around 15 to 18 credits in total. Students generally take 4-5 courses per semester, depending on their academic program and institution.
3. **Regular Evaluation:** The semester system includes frequent assessments such as mid-term exams, quizzes, assignments, and projects, in addition to a final exam or project. These evaluations contribute to the overall grade for the course and offer students opportunities to improve their performance over the course of the semester.
4. **Time for In-depth Study:** The duration of a semester allows students to engage in more detailed and thorough study. It provides time for reflection, revision, and the application of knowledge in practical or project-based contexts.
5. **Breaks Between Semesters:** One of the most significant advantages of the semester system is the built-in breaks between the fall and spring semesters, typically a winter break (between December and January) and a summer break (usually between May and August). These breaks allow students time to recharge, pursue internships, or engage in personal or academic projects.

### **Merits of Semester System**

- It engages the pupils in studies throughout the semester as they must appear both internal and semester examinations.
- Students' progress is assessed constantly and continuously. So, their knowledge gets improved by getting remedial instructions from teachers and by self-effort.
- The interaction between teachers and students increases in the semester system due to continuous class and internal test.
- The workload of students decreases as programme get equally distributed in all semesters. It supports learner psychological needs to learn in chunks.
- The semester system discourages students from studying at the last minute and encourages consistent and regular academic engagement throughout the course duration. It fosters a proactive approach to learning, facilitates deeper understanding of the subject matter, and supports effective time management.
- Continuous internal assessment and periodic tests stand out as significant advantages of this system.
- Students have the liberty to converse about their performance, ensuring transparency in the assessment process.
- This approach diminishes stress and pressure, contributing to a purposeful, enjoyable, and pleasant learning experience.

#### **Demerits of Semester System**

- Continuous examinations in the semester system keep students consistently facing the challenge of examinations.
- This system is suitable primarily for higher education not for school education.
- Formulating an appropriate syllabus for each semester poses a challenging task for the curriculum developers.
- Gaps between semesters can occasionally result in losses of learning by students.
- Both teachers and students experience an increased workload and stress to complete the course and examinations.
- At times, there is a perception that examinations outweigh the actual study process.
- The curriculum engagement may impede other aspects of students' development.

#### ***Self-check Exercise-5.8***

Write about your experience with the semester system and how it influenced your learning.

Share your thoughts on the effectiveness of this system.

---

---

---

---

---

---

## SEMESTER SYSTEM AROUND THE WORLD

The **semester system** is commonly used in many countries, though the exact structure may vary:

- **United States and Canada:** The semester system is standard at most colleges and universities, with Fall and Spring semesters. Many institutions also offer a summer session for additional courses or research.
- **Europe:** Many European universities follow the semester system, though variations exist across countries. For example, the Bologna Process in Europe aims to standardize academic structures, and most European universities now follow a two-semester academic year.
- **India:** The semester system is widely used in universities and colleges in India, especially in professional courses like engineering, law, and business administration. Some Indian universities have adopted the semester system to offer more flexibility and better align with global standards.
- **Australia:** The semester system is also prevalent in Australian universities, with the academic year split into two main semesters (usually Feb-June and July-November), along with optional summer terms.

---

## SUMMARY

The **semester system** is one of the most widely adopted structures in modern education, providing a balanced and flexible framework for organizing courses, assessments, and

academic progress. Its division of the academic year into shorter, more manageable terms enables better time management, regular feedback, and focused learning. While the system has its challenges—such as pacing and the pressure of frequent evaluations—it remains popular for its adaptability, opportunities for specialization, and support for student success. As educational institutions continue to innovate and respond to evolving student needs, the semester system will likely remain an integral part of the educational landscape globally.

---

### UNIT END EXERCISE

---

1. How does the semester system impact student learning and retention compared to the annual system?
2. What are the advantages and disadvantages of the semester system for both students and teachers?
3. In what ways does the semester system promote or hinder the development of critical thinking and in-depth learning?
4. How does the semester system support or limit opportunities for students to engage in extracurricular activities, internships, or study abroad programs?
5. How can the semester system be adapted or improved to better accommodate diverse learning styles and needs of students?

---

### FURTHER READINGS

---

- Gronlund, N. E. (1965). *Measurement and evaluation in teaching*.  
<http://ci.nii.ac.jp/ncid/BA12623208>
- Goswami, M. (2013). *Measurement and Evaluation in Psychology and Education*
- Lee, W. Y. (2010). *Assessment and evaluation in higher education*.  
<http://ci.nii.ac.jp/ncid/BB11596810>
- Patel, R. N. (2014). *Educational evaluation theory and practice*.  
<http://ci.nii.ac.jp/ncid/BA29677030>
- Kochhar S.K. (1984). *Guidance and Counseling in Colleges and Universities*.

## UNIT -18

# INTERNAL ASSESSMENT

## STRUCTURE

- **Learning Objectives**
- **Introduction**
- **Continuous Internal Assessment**
- **Summary**
- **Unit End Exercise**
- **Suggestion for Further Reading**

---

## LEARNING OBJECTIVES

---

After reading this unit, learners shall be able to;

- Define the Continuous Internal Assessment (CIA).
- Explain the key features of Continuous Internal Assessment (CIA)
- Elaborate the merits and demerits of continuous internal assessment
- Elaborate the implementation of Continuous Internal Assessment (CIA)

## **INTRODUCTION**

Continuous Internal Assessment (CIA) is an approach to evaluating students' performance over an extended period, rather than relying solely on end-of-term exams or final assessments. It focuses on the ongoing evaluation of students' progress through a variety of assessments that are conducted regularly throughout the academic term. The CIA system aims to assess a broader range of student skills, including academic knowledge, problem-solving, critical thinking, creativity, participation, and personal development.

In the CIA system, teachers continuously monitor and evaluate students' performance through various methods such as quizzes, assignments, projects, presentations, class participation, group work, and even regular feedback. The idea is to provide a more holistic view of a student's abilities, offering opportunities for improvement throughout the term rather than placing undue pressure on a single exam or assignment.

---

## CONTINUOUS INTERNAL ASSESSMENT

---

It is often said that the classroom is the second home for students. With most of their time spent in school, it is important to ensure that their progress is being monitored and evaluated on a regular basis. Teachers, being the ones who interact with the students every day, are the most suitable individuals to assess their performance in various areas. To ensure that students are receiving the best education possible, it is highly recommended that internal assessments

be conducted. These evaluations should be done continuously and form an integral part of the teaching and learning process. It is through these assessments that teachers can identify areas where students may be struggling or excelling, and offer the necessary support and guidance. Continuous internal evaluations help in the identification of gaps in the curriculum, which can be addressed immediately. This ensures that students are kept up-to-date with the latest information and are well prepared for any assessments that may be coming up. Additionally, these evaluations provide valuable feedback to students, helping them to understand their strengths and weaknesses and work on improving their skills. In conclusion, internal assessments play a vital role in the education system. It is imperative that teachers conduct these evaluations regularly to ensure that students are receiving the best education possible. By identifying areas where students may be struggling, teachers can offer the necessary support and guidance, helping students to achieve their full potential. The objectives of implementation of continuous internal evaluation are as follows:

- To conduct a comprehensive evaluation of a student's cognitive and non-cognitive areas of learning.
- To inspire both students and teachers to enhance the effectiveness of the teaching-learning process.
- To offer feedback to teachers, students, and parents about learning progress and suggesting suitable remedial measures.
- To reduce the emphasis on memorization and rote learning as contents are taught in small sections.
- To prioritize the assessment of non-academic facets of a child's personality that can facilitate in holistic development.
- To enhance the meaningfulness, reliability, validity, and objectivity of the evaluation system.

### **Key Features of the CIA System**



1. **Ongoing Evaluation:** The CIA system ensures that assessments occur throughout the course, providing a more accurate reflection of a student's abilities over time. Instead of a one-time examination, students' learning is evaluated continuously.
2. **Diverse Assessment Methods:** A wide variety of assessment methods are used under the CIA system. These can include:
  - Written tests/quizzes (both announced and surprise)
  - Assignments and homework
  - Projects and presentations
  - Group discussions or seminars
  - Class participation
  - Practical or lab work
  - Peer assessments and self-assessments

These methods ensure that different learning styles and abilities are taken into account.

3. **Feedback and Improvement:** One of the key strengths of CIA is that it allows for regular feedback. This continuous feedback loop helps students understand their strengths and areas for improvement early in the term. They can act on the feedback to improve their learning, giving them the chance to adapt and adjust their study methods.
4. **Formative and Summative Assessments:** The CIA system blends both formative (ongoing) and summative (final) assessments. While formative assessments help students improve throughout the course, summative assessments evaluate their final level of achievement in the subject.
5. **Holistic Evaluation:** In addition to academic performance, CIA often assesses non-cognitive aspects of learning such as critical thinking, collaboration, communication, and creativity. This holistic approach helps in the overall development of the student, not just their ability to memorize facts.

#### **Merits of Internal Evaluation**

- It involves continuous observation and occasional testing of students by teachers.
- Continuous internal assessment serves dual purposes, being both formative for instructional improvement and summative to complement final exam results.
- It strengthens the relationship between students and teachers.
- It encompasses evaluation in cognitive as well as other dimensions of students' personalities.

- This system encourages students to maintain attentiveness and regularity throughout the duration of their studies.

### **Demerits of Internal Evaluation**

- Internal evaluation system requires a reasonable teacher-taught ratio in the class for its effective use.
- There is a chance of biasness and favoritism by teachers in the internal assessment.
- A positive atmosphere within the institution/school is essential for the system's effectiveness.
- In this system, students who seek recommendations, impress their teachers, or establish close relationships may receive higher marks/grades than warranted.
- Due to undue pressure, teachers may experience feelings of insecurity and stress.
- The distribution between internal and external assessment in the system is a matter of debate.

### ***Self-check Exercise-5.9***

Reflect how continuous internal assessment enhances rapport between students and teachers.

---

---

---

---

### **Implementation of CIA in Education**

To implement a successful Continuous Internal Assessment system, institutions need to:

1. **Develop Clear Guidelines:** Schools and universities must establish clear criteria and guidelines for assessments to ensure fairness and consistency. Rubrics and grading scales for assignments, projects, and presentations should be made available to students at the start of the course.
2. **Balanced Assessment:** A mix of different assessment types should be used (e.g., quizzes, assignments, projects, class participation, etc.) to ensure a well-rounded

evaluation. It is important to ensure that no single form of assessment disproportionately affects the overall grade.

3. **Use of Technology:** Educational institutions can leverage technology, such as Learning Management Systems (LMS), to facilitate the submission of assignments, track progress, and provide feedback. This can also help reduce administrative burden on teachers.
4. **Regular Feedback Mechanisms:** Students should receive timely and constructive feedback throughout the course to ensure they understand their progress and how to improve. Regular interaction between teachers and students is crucial for a successful CIA system.
5. **Teacher Training:** Teachers need to be trained in the CIA methodology and how to fairly assess students across various formats. This includes developing rubrics, conducting peer assessments, and offering constructive feedback.

---

## SUMMARY

The **Continuous Internal Assessment (CIA)** system represents a shift towards a more holistic and student-centered approach to evaluation in education. By offering regular feedback, a variety of assessment methods, and a focus on a wide range of skills, the CIA system aims to enhance learning, reduce stress, and foster the overall development of students. However, successful implementation requires careful planning, clear guidelines, and an ongoing commitment to fairness and equity. When done well, CIA can significantly improve the educational experience by providing a more comprehensive and balanced evaluation of student progress.

---

## UNIT END EXERCISE

1. How does the implementation of the Continuous Internal Assessment (CIA) system impact student motivation and engagement in their learning process?
2. What are the challenges teachers face in managing Continuous Internal Assessments, and how can these challenges be addressed?
3. How does the CIA system contribute to the development of critical thinking, creativity, and problem-solving skills in students?

4. What are the potential biases or limitations in the Continuous Internal Assessment system, and how can they be mitigated?
5. How can the integration of technology enhance the effectiveness of the Continuous Internal Assessment system in modern education?

---

#### FURTHER READINGS

---

- **Buchberger, F., & Klieme, E. (2004).** *Assessment and Evaluation: Continuous Assessment in Schools*. *European Educational Research Journal*, 3(4), 357–370.
- **Guskey, T. R. (2007).** *Closing the Achievement Gap: A Vision for Changing Beliefs and Practices*. Corwin Press.
- **Reddy, P., & Andrade, H. (2010).** *A Review of Rubrics in Higher Education: Accuracy, Consistency, and Utility*. *Assessment & Evaluation in Higher Education*, 35(4), 423–438.
- **Wiggins, G. (1998).** *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. Jossey-Bass.
- **Black, P., & Wiliam, D. (1998).** *Assessment and Classroom Learning*. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- **Assessment Reform Group (2002).** *Testing, Motivation, and Learning*. *Research Papers in Education*, 17(1), 87–113.
- **Lizzio, A., & Wilson, K. (2006).** *Course Completion: The Influence of Perceived Quality of Teaching and Assessment*. *Studies in Higher Education*, 31(1), 45–64.

# UNIT 19

## QUESTION BANK

### STRUCTURE

- Learning Objectives
- Introduction
- New Trends of Evaluation
- Grading
- Summary
- Unit End Exercise
- Suggestion for Further Reading

---

### LEARNING OBJECTIVES

---

After reading this unit, learners shall be able to;

- Define the new trends of evaluation system.
- Explain the Grading system and semester system
- Elaborate the continuous internal assessment and use of question bank
- Elaborate the use of computers in evaluation

### INTRODUCTION

A **question bank** is a collection of pre-written questions used in the process of assessing and evaluating students' knowledge and understanding in an academic setting. It serves as a valuable resource for educators to generate assessments like quizzes, exams, and practice tests. These questions are typically organized according to different subject areas, difficulty levels, types of questions (e.g., multiple-choice, short-answer, essay), and learning objectives. The use of a question bank allows for greater efficiency in the creation of assessments while maintaining consistency and fairness.

---

### QUESTION BANKS

---

Question banks refer to collections of questions that can be shared across various courses and programs of study. It is essentially a list of questions from a specific subject, which is compiled through collaborative efforts for the benefit of students, teachers, and assessors.

Question banks may also be known as "item banks," "item pools," "item collections," "item reservoirs," or "test item libraries."

A question bank is a collection of large test items developed by a group of trained and experienced professionals and printed on index cards or stored in the memory of a computer along with certain supporting data and capable of being reproduced whenever needed (Agrawal, 2005). Question banks can be searched to find questions that meet specific criteria and create assessments. Users with appropriate access can create questions in the banks. Question banks serve several purposes. Teachers can use it during pre-testing, to create question papers, measure student achievement, etc. Practicing teachers should prepare the questions for the question banks. Question enrichment should be an ongoing process that includes updating, rejecting, replacing, modifying, and adding new questions. In question banks, all types of questions - objective, short answer, as well as long answer - are included for a particular topic. Question banks are crucial for teaching, general exams, competitive exams, and entrance tests. They can also increase the importance of large-scale public exams by covering a wider range of content. The primary reasons for creating question banks include:

- Addressing the requirements of an expanding student population across various programmes.
- Enhancing knowledge and striving for competitiveness in all facets of life.
- Cultivating testing techniques that are more reliable, valid, comprehensive, fair, and objective.
- Structuring testing programs that are more efficient, cost-effective, and comprehensive.
- Employing scientific and technological tools for data processing.
- Elevating the efficiency of testing programs.

Question banks can enhance testing techniques by making them more efficient and objective. Let us discuss the advantages and disadvantages of using question banks.

### **Key Features of a Question Bank**

1. **Variety of Question Types:** A well-designed question bank contains a diverse range of question formats, including:

- **Multiple-choice questions (MCQs)**
  - **True/false questions**
  - **Short answer questions**
  - **Long answer or essay-type questions**
  - **Matching and fill-in-the-blank questions**
  - **Problem-solving or case study questions** (particularly for practical subjects)
2. **Categorization:** Questions are often categorized by various criteria:
- **Subject Area:** Grouped by topics, units, or themes (e.g., mathematics, history, science).
  - **Difficulty Level:** Questions may be classified into easy, moderate, and difficult categories.
  - **Learning Objectives:** Questions may be aligned with specific learning goals or competencies, such as understanding concepts, applying knowledge, analyzing data, or evaluating outcomes.
3. **Reusable:** One of the primary benefits of a question bank is that it can be reused for multiple tests or assessments over time, saving teachers significant time in question preparation.
4. **Randomized Question Selection:** Many electronic or computer-based question banks support the randomization of questions. This means that when an exam is generated from a question bank, the questions can be selected randomly, ensuring that each student receives a unique set of questions. This helps reduce the chances of cheating or collaboration during exams.
5. **Alignment with Curriculum and Learning Outcomes:** Effective question banks are closely aligned with the course syllabus, curriculum guidelines, and learning outcomes. This ensures that the questions assess the content and skills that students are expected to master.

### Advantages of Using a Question Bank

1. **Efficiency in Exam Preparation:** Creating exams or quizzes from a pre-existing question bank allows teachers to save time. Instead of writing new questions for every exam, educators can select or modify questions from the bank based on the topics they want to assess.
2. **Consistency and Fairness:** Since the questions in the bank are pre-written and categorized, it becomes easier to create exams that are consistent in terms of difficulty level, coverage of content, and alignment with the learning objectives. This helps ensure fairness in testing, as all students are assessed using similar question types and standards.
3. **Variety of Assessment:** A question bank allows for a diverse range of assessment tools, from multiple-choice tests to long-form essays, enabling teachers to assess students' knowledge from different angles and in various formats. This can cater to different learning styles and reduce the reliance on one single mode of assessment.
4. **Personalized Assessment:** With larger question banks, teachers can create different versions of an exam or quiz, allowing them to personalize assessments for individual students or groups of students. This can be especially useful in large classes, helping to mitigate the risks of cheating.
5. **Tracking Student Progress:** When teachers use a question bank over multiple terms or courses, they can track which types of questions are most challenging for students. This allows for the identification of areas where students consistently struggle, enabling instructors to adjust their teaching methods accordingly.
6. **Helps in Formative Assessment:** Question banks can be used not just for summative assessments (final exams), but also for **formative assessments**—regular quizzes, practice tests, or in-class exercises that gauge student learning throughout the course. This approach provides continual feedback to both students and teachers.

### Merits of Question Bank

- There is minimal risk of question paper leaks, as even experts are unaware of whether their questions are included in the test.



- Teachers are aware of the types of questions expected in the examination, allowing them to tailor their instruction accordingly.
- Question papers can be promptly prepared, even in emergency situations by using question bank.
- The question bank functions as a tool for facilitating comprehensive learning from diverse perspectives.
- Paper setters can use the question bank as a reference guide.
- The involvement of various instructors and specialists in item construction ensures objectivity in evaluation.
- Question banks are frequently employed for admission and examination purposes.
- The standardized scoring procedure upholds the reliability and objectivity of test results.
- The issue of non-comparability of marks across different times or years is addressed by assuming that sample tests chosen from the bank are parallel or equivalent in all aspects.

### **Demerits of Question Bank**

- Educators hold diverse views regarding the confidentiality of the question bank.
- The creation of questions for question papers often lacks originality if it is drawn from the question bank.
- In the present system, where the objective of an examination is merely to ascertain a pass or fail outcome, there seems to be no justification for the existence of such banks for students or teachers.
- Item writers need subject expertise and specialized training to modify the items of question bank before its use for formal testing.
- The item writer might be unfamiliar with the psychological traits of the students intended to take or use the test.

#### ***Self-check Exercise-5.6***

Develop a small question bank for a subject of your choice. Include various types of questions (objective, short answer, long answer).

-----

-----
-----
-----
-----
-----

### How to Build an Effective Question Bank

To build an effective and useful question bank, educators should follow these steps:

1. **Define Learning Objectives:** Clearly define what students are expected to learn and be able to do by the end of the course. Questions should be mapped to these objectives.
2. **Variety and Balance:** Include a variety of question types (MCQs, essays, practical problems) to assess different skills. Balance between easy, moderate, and difficult questions to ensure the assessment reflects a range of student abilities.
3. **Regular Updates:** Regularly review and update the question bank to ensure relevance and fairness. This includes ensuring that questions are free from bias and reflect any changes in the curriculum.
4. **Pilot Testing:** Before fully implementing a new set of questions, pilot them with a small group of students or colleagues to test their clarity and difficulty level.
5. **Review and Refine:** Continuously assess the effectiveness of the question bank by collecting feedback from students and teachers, and refine it based on performance data.

---

### SUMMARY

A **question bank** is an essential tool in modern educational assessment. By providing a repository of pre-made, organized questions, it saves time in exam preparation, ensures consistency, and allows for varied types of evaluation. When designed and used thoughtfully, it can improve both the quality and fairness of assessments. However, it is important to ensure that the question bank is continually updated, diverse, and aligned with the curriculum to prevent issues like over-reliance on factual recall and ensure that higher-order thinking is also adequately tested.

---

### UNIT END EXERCISE

---

1. Differentiate between Teacher-made Achievement Test and Standardized Achievement Test.
2. A child of 13 years scores 88 in an IQ test. Calculate his mental age.
3. Create a test paper with a mix of questions covering different aptitude areas.
4. Compare Thurstone scale and Likert scale, highlighting the differences in their methods and applications for measuring social attitudes.
5. Provide details about some other interest inventories including its name, purpose, target audience, and any unique features it offers for assessing interests?
7. Conduct a brief classroom observation and identify non-academic skills demonstrated by students. Provide examples of social, emotional, and behavioral competencies observed.
8. Brainstorm creative ways teachers can use question banks beyond traditional examinations, such as in classroom activities, formative assessments, or as a resource for personalized learning.
9. Develop a comparative analysis between direct grading and indirect grading. Highlight the advantages and disadvantages of each method. Provide scenarios where each grading method might be more suitable.
10. Reflect on the future trends and advancements in computer-based evaluation. Discuss emerging technologies and their potential influence on assessment methods.

---

### FURTHER READINGS

---

- Gronlund, N. E. (1965). *Measurement and evaluation in teaching*.  
<http://ci.nii.ac.jp/ncid/BA12623208>
- Goswami, M. (2013). *Measurement and Evaluation in Psychology and Education*
- Lee, W. Y. (2010). *Assessment and evaluation in higher education*.  
<http://ci.nii.ac.jp/ncid/BB11596810>
- Patel, R. N. (2014). *Educational evaluation theory and practice*.  
<http://ci.nii.ac.jp/ncid/BA29677030>
- Kochhar S.K. (1984). *Guidance and Counseling in Colleges and Universities*.

## UNIT -20

### USE OF COMPUTER IN EVALUATION

#### STRUCTURE

- **Learning Objectives**
- **Introduction**
- **New Trends of Evaluation**
- **Grading**
- **Summary**
- **Unit End Exercise**
- **Suggestion for Further Reading**

---

#### LEARNING OBJECTIVES

---

After reading this unit, learners shall be able to;

- Define the new trends of evaluation system.
- Explain the Grading system and semester system
- Elaborate the continuous internal assessment and use of question bank
- Elaborate the use of computers in evaluation

#### INTRODUCTION

The integration of **computers** and **information technology** in educational assessment has transformed traditional evaluation practices. Computers have significantly enhanced the efficiency, accuracy, and accessibility of evaluation systems, providing both educators and students with powerful tools for assessment, feedback, and learning. The use of computers in evaluation spans multiple dimensions—from test administration and grading to data analysis and personalized feedback.

---

#### Use of Computer in evaluation

---

In recent years, the incorporation of computers into educational practices has become increasingly prevalent. One of the areas witnessing significant change in the evaluation process is use of computers. Traditional methods of assessment are giving way to more dynamic, flexible, and efficient computer-based evaluation strategies. The computers can be used in diverse ways from planning to publication of results in the education system.

Traditionally, teachers write the draft questions many times before finalization manually. But now computers can be utilized to draft questions, saved, retrieved, and edited as and when required. That means the computer has made the task of preparing questions easy. Computers can be useful for administering the test items on students through online testing at same time across countries and school boards. Now the publication of results and communication of examination results becomes very easy due to computer software and mailing services.

**1. Online Assessments-**The advent of online assessments has revolutionized the way educators evaluate students. With the capacity to create and administer quizzes and exams digitally, the geographical and temporal constraints of traditional evaluations are dismantled. Educators can leverage various question formats, including multiple-choice, true/false, and fill-in-the-blank, while automated grading systems streamline the evaluation process, providing prompt feedback to both educators and students.

**2. Interactive Learning Platforms-**Learning Management Systems (LMS) have become the cornerstone of modern education. Platforms such as Moodle, Canvas, and Blackboard offer a plethora of tools for designing, delivering, and evaluating assignments. The interactive nature of these platforms facilitates collaborative learning, allowing students to engage in discussions, submit assignments, and receive feedback in a digital environment.

**3. Simulations and Virtual Labs-** Practical assessments are integral to evaluating hands-on skills and real-world application of knowledge. Simulations and virtual labs provide a risk-free yet immersive environment for students to apply theoretical concepts. Educators can assess problem-solving abilities, critical thinking skills, and the practical application of knowledge in fields ranging from science and engineering to healthcare.

**4. E-Portfolios-**The traditional portfolio has undergone a digital transformation with the advent of e-portfolios. These digital showcases allow students to compile a comprehensive collection of their work, projects, and achievements. E-portfolios provide educators with a holistic view of a student's progress, emphasizing not only the final outcomes but also the learning process, growth, and reflective insights.

**5. Automated Writing Evaluation-** The emergence of artificial intelligence (AI) in education has extended to the evaluation of written assignments. Automated writing evaluation tools, powered by Natural Language Processing (NLP), analyze written content

for grammar, style, and content. This not only expedites the grading process but also provides students with instant, personalized feedback to support their learning journey.

**6. Adaptive Learning Platforms-** Addressing the diverse learning needs of students, adaptive learning platforms personalize assessments based on individual performance. These platforms dynamically adjust the difficulty level of questions to match each student's proficiency, ensuring a customized learning experience. The data generated from adaptive systems further contribute to informed decision-making for educators.

**7. Remote Proctoring-** Maintaining the integrity of online assessments is a paramount concern. Remote proctoring tools address this challenge by monitoring students during exams. These tools utilize webcam and microphone technology to ensure a secure testing environment, mitigating the risk of academic dishonesty and upholding the credibility of evaluation outcomes.

**8. Data Analytics and Learning Analytics:** The use of data analytics plays a pivotal role in shaping educational strategies. Data analytics enable educators to analyze trends in student performance, identifying areas of strength and weakness. Learning analytics, on the other hand, dig into how students interact with educational content, offering insights that can inform instructional design and interventions.

**9. Applications for assessment:** Many computer and mobile applications have been developed by the educationist and computer scientist during the Covid-19 for conducting assessment. The applications such as Google form, Google classroom, Multimeter, Kahoot etc. can be used for testing students in offline and online classes. The beauty of these applications is that students get immediate feedback about their results and teachers can do very quick result analysis and identification of students who need improvement.

**10. Rubrics for assessment:** The computer not only helps in writing and administering the test items but also helps in systematic evaluation of answer scripts and projects/ assignments. Rubrics is a set of criteria or guidelines to be used for evaluating the student's product such as essay, project, and assignments. The online and free rubric like Rubi star is very helpful for teachers for developing rubrics which can be useful for evaluation of students' answers.

## Key Areas of Computer Use in Evaluation

### 1. Automated Testing and Exam Administration

- **Online Quizzes and Exams:** Computers enable the creation, distribution, and administration of online quizzes and exams. Students can take assessments remotely, which makes testing more flexible and accessible. These online tests can include multiple-choice questions (MCQs), short-answer questions, true/false questions, and even essay-style responses.
- **Randomization of Questions:** Computerized assessment systems can randomize the order of questions and answers, ensuring that each student receives a unique version of the test. This helps reduce cheating and ensures fairness.
- **Timed Assessments:** Computers can track time during assessments, automatically ending the exam when the allotted time expires. This reduces human error and ensures that all students are given equal testing time.

### 2. Automated Grading and Feedback

- **Instant Grading:** One of the most significant advantages of using computers in evaluation is **automated grading**. Systems can quickly grade objective-based questions (e.g., MCQs, true/false, matching) without any human intervention. This provides immediate results to students and teachers.
- **Error-Free Grading:** Computerized grading eliminates the risk of human error in scoring, ensuring accuracy and consistency. This is especially important in large classes or for standardized testing.
- **Detailed Feedback:** Some computer systems are designed to provide **instant feedback** to students after they complete an assessment. This feedback can include explanations for correct and incorrect answers, offering students an opportunity to learn from their mistakes immediately.

### 3. Data Collection and Analysis

- **Performance Tracking:** Computers can collect and store data on students' performance over time. Teachers can use this data to identify trends in student

learning, track individual progress, and monitor class-wide performance. This information helps in making informed decisions about future teaching strategies.

- **Learning Analytics:** Advanced software can analyze patterns in student responses, engagement, and behavior to predict learning outcomes. Learning analytics tools can generate reports that highlight areas where students are excelling or struggling, allowing for targeted interventions.
- **Statistical Analysis:** Computers facilitate **statistical analysis** of student performance, enabling the creation of sophisticated evaluation metrics such as grade distributions, correlation between question difficulty and student performance, and analysis of bias in assessments.

#### 4. Personalized Learning and Adaptive Assessments

- **Adaptive Testing:** Adaptive assessments adjust the difficulty of questions based on the student's performance. If a student answers a question correctly, the system presents a more challenging question next. If a student answers incorrectly, the system may provide an easier question. This approach helps evaluate the student's actual level of competence and provides a more personalized assessment experience.
- **Tailored Feedback:** Computers can generate personalized feedback based on a student's performance. For example, if a student struggles with a particular topic, the system can suggest resources, exercises, or remedial lessons to help them improve.

#### 5. Computer-Based Simulation and Practical Assessments

- **Virtual Labs and Simulations:** In fields like science, engineering, medicine, and economics, computers enable the creation of **virtual simulations** where students can demonstrate their practical skills without the need for physical labs. These computer-based assessments allow students to engage in hands-on learning in a controlled, virtual environment.
- **Simulated Scenarios:** For subjects like medicine, law, or business, simulations of real-world scenarios (e.g., diagnosing patients, legal case simulations, or business management exercises) can be used to assess decision-making, problem-solving, and practical application of knowledge.



## 6. Online Portfolios and E-assessment Tools

- **E-portfolios:** Computers facilitate the creation of **electronic portfolios** where students can submit and store evidence of their learning, such as essays, projects, presentations, and videos. These portfolios can be assessed regularly, providing a more holistic view of a student's progress over time.
- **E-assessment Tools:** Many educational platforms now provide comprehensive tools for conducting assessments. These systems offer features such as question banks, automated grading, progress tracking, and real-time data analytics, all of which streamline the evaluation process.

## 7. Support for Multiple Learning Styles

- **Multimedia Assessments:** Computers allow educators to incorporate a variety of media (such as videos, images, sound clips, and interactive elements) into assessments, catering to different learning styles. For example, visual learners might benefit from assessments that include diagrams or infographics, while auditory learners could benefit from audio-based questions.
- **Interactive Assessment Formats:** Interactive quizzes and gamified assessments can engage students more effectively. For example, interactive drag-and-drop questions or simulation-based assessments can be used to test knowledge in a more engaging way.

## Merits of Computer for Assessment

- It reduces the time and effort required for evaluation, allowing educators to focus on providing meaningful feedback and adapting instructional strategies.
- Computer-based evaluations provide flexibility in terms of timing and location. Students can access assessments remotely, accommodating diverse learning styles and schedules.
- It offers instant feedback to students, enabling them to quickly identify and rectify errors. This immediate feedback promotes a continuous learning process and timely intervention when needed.

- Adaptive learning platforms tailor assessments to individual student proficiency levels, providing a personalized learning experience. This approach caters to diverse learning needs and supports students in reaching their full potential.
- Computers enable the use of multimedia elements in assessments, such as video submissions and interactive simulations. This allows for a more comprehensive evaluation of students' skills and abilities beyond traditional written tests.
- Over time, the use of computer-based evaluations can be cost-effective, especially when considering the reduction in paper, printing, and manual grading expenses. It also eliminates the need for physical storage space for documents.

### **Demerits of Computers for Assessment**

- Unequal access to technology can create disparities in students' ability to participate in computer-based assessments.
- The initial costs associated with implementing computer-based evaluation systems, including software, training, and infrastructure, can be substantial. Educational institutions may face budget constraints in adopting such technologies.
- Educators may require time and training to adapt to new evaluation technologies. The learning curve for mastering assessment tools and platforms can initially impact efficiency.
- Ensuring standardized evaluation across diverse settings and institutions can be challenging.
- Computer-based evaluations may lack the personal touch and nuanced understanding that educators can provide through face-to-face interactions.
- While automated grading systems are efficient, they may not capture the complexity of certain subjective assessments, such as essays or projects.
- The use of data analytics raises ethical concerns regarding student privacy. Protecting confidential information and guaranteeing compliance with data protection laws is an essential aspect to consider

### **Challenges of Using Computers in Evaluation**

#### **1. Technical Issues and Accessibility:**

- Not all students have access to computers, the internet, or the necessary digital literacy skills. This digital divide can create inequities, especially in lower-income or rural areas.
- **System Downtime:** Technical failures, such as server crashes or software glitches, can disrupt the exam-taking process, potentially leading to unfair or incomplete evaluations.

## 2. **Security and Integrity:**

- Online assessments are vulnerable to cheating, especially in cases where students have access to the internet during assessments. Educators must implement security measures like **proctoring software**, **randomized question banks**, and **browser lockdowns** to ensure the integrity of the exam.
- **Plagiarism:** Students may copy content from the internet or other sources, particularly in assignments and essays, making it harder to evaluate their original thinking.

## 3. **Over-Reliance on Automated Grading:**

- While computers excel at grading objective questions, they may struggle with more complex assessments, such as essays or long-answer questions. Some systems attempt to grade these using artificial intelligence, but this technology is still not perfect and may fail to appreciate nuances in student responses.

## 4. **Data Privacy and Security:**

- Storing large amounts of student data online raises concerns about data privacy and security. Institutions need to ensure that the data is protected from unauthorized access and that students' personal information is handled securely.

## 5. **Lack of Human Judgment:**

- While automated systems are effective for grading objective questions, they cannot replicate the nuanced judgment of human evaluators. Some assessments, such as group work, presentations, or essays, require a more

holistic evaluation that combines subject knowledge with critical thinking, creativity, and communication skills.

---

## SUMMARY

---

The use of **computers in evaluation** offers numerous advantages, including efficiency, scalability, and accessibility. It enables the automation of many aspects of assessment, such as test creation, grading, and feedback, which improves the overall learning experience for students and reduces administrative burdens for educators. However, there are challenges related to technical issues, security, and data privacy that need to be addressed. When implemented thoughtfully, computer-based evaluation can enhance the quality, fairness, and reach of educational assessments, making it an essential tool in modern education.

---

## UNIT END EXERCISE

---

1. Differentiate between Teacher-made Achievement Test and Standardized Achievement Test.
2. A child of 13 years scores 88 in an IQ test. Calculate his mental age.
3. Create a test paper with a mix of questions covering different aptitude areas.
4. Compare Thurstone scale and Likert scale, highlighting the differences in their methods and applications for measuring social attitudes.
5. Provide details about some other interest inventories including its name, purpose, target audience, and any unique features it offers for assessing interests?
7. Conduct a brief classroom observation and identify non-academic skills demonstrated by students. Provide examples of social, emotional, and behavioral competencies observed.
8. Brainstorm creative ways teachers can use question banks beyond traditional examinations, such as in classroom activities, formative assessments, or as a resource for personalized learning.
9. Develop a comparative analysis between direct grading and indirect grading. Highlight the advantages and disadvantages of each method. Provide scenarios where each grading method might be more suitable.
10. Reflect on the future trends and advancements in computer-based evaluation. Discuss emerging technologies and their potential influence on assessment methods.

---

**FURTHER READINGS**

---

- Gronlund, N. E. (1965). *Measurement and evaluation in teaching*.  
<http://ci.nii.ac.jp/ncid/BA12623208>
- Goswami, M. (2013). *Measurement and Evaluation in Psychology and Education*
- Lee, W. Y. (2010). *Assessment and evaluation in higher education*.  
<http://ci.nii.ac.jp/ncid/BB11596810>
- Patel, R. N. (2014). *Educational evaluation theory and practice*.  
<http://ci.nii.ac.jp/ncid/BA29677030>
- Kochhar S.K. (1984). *Guidance and Counseling in Colleges and Universities*.