MEASUREMENT AND EVALUATION IN EDUCATION

MA [Education]

Second Semester

EDCN 804E

[ENGLISH EDITION]

Directorate of Distance Education

TRIPURA UNIVERSITY

Reviewer

Dr Aradhana Mani
Assistant Professor, Amity Institute of Education, Noida

Authors:

Dr Harish Kumar and Santosh Kumar Rout: Units (1.2, 2.3-2.4, 2.5, 3.3, 4.8) © Dr Harish Kumar and Santosh Kumar Rout, 2016

Dr Manisha Dalabh: Units (1.3.1, 3.4-3.4.1, 3.4.3, 5.4-5.5) © Dr Manisha Dalabh, 2016

Dr Jasim Ahmad and Dr Aerum Khan: Units (1.4, 1.6, 2.2, 3.2, 4.3, 4.6) © Dr Jasim Ahmad and Dr Aerum Khan, 2016

JS Chandan: Unit (4.4, 4.6.1) © JS Chandan, 2016

CR Kothari: Unit (4.5, 4.7) © CR Kothari, 2016

Lokesh Koul: Unit (5.2, 5.3) © Lokesh Koul, 2016

Books are developed, printed and published on behalf of **Directorate of Distance Education, Tripura University** by Vikas Publishing House Pvt. Ltd.

Vikas® is the registered trademark of Vikas® Publishing House Pvt. Ltd.

VIKAS® PUBLISHING HOUSE PVT. LTD.
E-28, Sector-8, Noida - 201301 (UP)
Phone: 0120-4078900 • Fax: 0120-4078999
Regd. Office: 7361, Ravindra Mansion, Ram Nagar, New Delhi – 110 055
• Website: www.vikaspublishing.com • Email: helpline@vikaspublishing.com

SYLLABI-BOOK MAPPING TABLE

Measurement and Evaluation in Education

Syllabi	Mapping in Book
Unit - I Concept of Measurement and Evaluation, Different Types of Measuring Scales Need for Measurement and Evaluation in Education, Placement, Diagnostic, Formative and Summative Evaluation-Role of Teachers in an Evaluation Programme, Taxonomy of Educational Objectives (Cognitive, Affective and Psychomotor) - Specification of Objectives Steps in the Process of Evaluation.	**Unit 1:** Measurement and Evaluation in Education: Overview **(Pages: 3-46)**
Unit - II Major Tool and Techniques in Educational Evaluation, Different Types of Tests-teacher made Vs. Standardized, Criterin-referenced vs. Norm-referenced Test, Essential Qualities of Good Measuring Instrument. Education Tests, Measurement of Achievement-Construction of Achievement Test and Standardization, Relative Merits and Demerits of Using Different Types of Test Items, Diagnostic Test-construction and Usefulness.	**Unit 2:** Major Tools and Techniques in Educational Evaluation **(Pages: 47-73)**
Unit - III Acquaintance with Psychological Tests in the Area of Intelligences, Attitude and Personality, Examination System-current Strategies-examination Reforms-open Book Examination-Semester System.	**Unit 3:** Psychological Testing **(Pages: 75-109)**
Unit - IV Statistical Treatment of Data: Frequency Distribution and Graphic Representation of Data, Measures of Central Tendency and Variability, Co-efficient of Correlation by Rank Differences and Product Moment Methods, Percentile and Percentile Rank, Skewness and Kurtosis, Normal Probability Curve, Derived Scores (Z-Score, Standard Score and T-Score).	**Unit 4:** Statistics in Measurement and Evaluation-I **(Pages: 111-209)**
Unit - V Reliability-Concept, Determining Factors-Methods of Determining Different Reliability Coefficient, Validity-Concept and Use-Types of Validity-Determination of Validity Co-efficient-relation between Validity and Reliability, Trends in Evaluation: Grading, Credit System, Cumulative Record Card. Computer in Evaluation.	**Unit 5:** Statistics in Measurement and Evaluation-II **(Pages: 211-248)**

CONTENTS

INTRODUCTION

NOTES

Quality of education has become very important in today's competitive environment. There is definitely a need to adapt to change in the educational processes in order to improve. Educational measurement requires the establishment of a strong feedback loop with evaluation being a continuous process and not just left until the end of the program of study.

Measurement refers to the process by which the attributes or dimensions of some physical object are determined. When used in the context of learning, it would refer to applying a standard scale or measuring device to an object, series of objects, events or conditions, according to practices accepted by those who are skilled in the use of the device or scale. On the other hand, evaluation is a complex and less understood term. Inherent in the idea of evaluation is 'value'. It involves engaging in some process that is designed to provide information that will help us make a judgment about a given situation. Generally, any evaluation process requires information about the situation in question.

This book *Measurement and Evaluation in Education* introduces the concept of testing and evaluation to its readers. This book is crucial in explaining how, even though different evaluation scales have been designed, the success is not entirely determined by the efficiency of the test. More often than not, the right test administered to the right individual makes all the difference. The way in which tests are constructed and comprehended is crucial to the process of evaluation. Different tests are designed to examine different attributes of intelligence and its evaluation is contingent on how well these tests have been understood. In other words, the lucidity of a test, inclusive of the kind of questions, is an important factor in determining its success and accuracy.

The book is divided into five units. It has been designed keeping in mind the self-instruction mode (SIM) format and follows a simple pattern, wherein each unit of the book begins with Introduction followed by Unit Objectives to the topic. The content is then presented in a simple and easy-to-understand manner, and is interspersed with Check Your Progress questions to test the understanding of the topic by the students. A list of Questions and Exercises is also provided at the end of each unit, and includes short-answer as well as long-answer questions. Key Terms and the Summary section is a useful tool for students and are meant for effective recapitulation of the text.

UNIT 1 MEASUREMENT AND EVALUATION IN EDUCATION: OVERVIEW

Structure

1.0 INTRODUCTION

The process of measurement is an essential component of the learning system. Measurement has more administrative utility in education than evaluation. An evaluation process is useful in modifying and improving learning system and instructional procedure. The process of measurement is more precise and objective in approach, when compared to evaluation.

Measurement is always done of a quality, attribute or variable of a thing or a person. Psychologists and educationists are mainly concerned with variables and attributes. The process of measurement converts the variables into variety which is used for drawing inferences. For example, intelligence is quantified in terms of IQ and achievement variable is measured in terms of scores.

This unit deals with the nature of measurement and evaluation in education. It also discusses the role of teachers in the evaluation programme as well as the taxonomy of educational objectives.

1.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Identify different measurement scales
- Discuss the nature of educational measurement and evaluation

- Describe the types and techniques of evaluation
- Explain the cognitive, affective and psychomotor domain and how their objectives are formulated and assessed

NOTES

1.2 CONCEPT OF MEASUREMENT AND EVALUATION

Let us begin by discussing the concept of measurement and measurement sales.

Concept of Measurement

Measurement refers to the process by which the attributes or dimensions of some physical object are determined. When used in the context of learning, it refers to: applying a standard scale or measuring device to an object, a series of objects, a events or conditions, according to practices accepted by those who are skilled in the use of the device or scale.

E.L. Thorndike stated that 'anything that exists at all, exists in some quantity; and anything that exists in some quantity is capable of being measured'. Measurement of any kind is a matter of determining how much or how little, how great or how small, how much more than or how much less than. The *Encyclopedia of Educational Research* explains measurement in more refined terms; to measure means 'to observe or determine the magnitude of a variant'.

Measurement answers the question of 'how much'. In our day-to-day life, we measure the height, weight, miles travelled, etc., the tailor measures the dimensions of an individuals body to prepare dress according to required size, the shopkeepers weigh different commodities like rice, wheat, sugar, fruits, vegetables etc. We have a wrong notion that measurement takes place only with tapes and scales. The ranking of contestants in a debate competition can be considered as measurement, rating of human behaviour comes under measurement. Hence, measurement may be defined as, 'the assignment of one of a set of numbers to each of a set of persons or objects according to certain established rules'.

James M. Bradfield defined measurement as 'the process of assigning symbols to the dimension of phenomenon in order to characterize the status of phenomenon as precisely as possible'. J.P. Guilford defined measurement as the 'assignment of numerals to objects or events according to certain rules'. According to Norman E. Gronlund, 'measurement results are some score or numerical value and quantitative descriptions of the pupils'.

Measurement involves the process of quantification. Quantification indicates to what extent a particular attribute is present in a particular object. It has been observed that measurement in any field always involves three essentials:

(i) Identification and definition of quantity, attribute or variable that is to be measured.

(ii) Determining the set of operations by which the attribute or variable may be made perceivable.

(iii) Establishing a set of procedure for translating observations into quantitative statement of degree, extent or amount.

Types of measurement

NOTES

Measurement is of two types: (i) physical measurement and (ii) mental measurement/ psychological measurement/educational measurement.

(i) Physical measurement: Physical measurement is the measurement of the object which has absolute existence. For example, we measure the height of individuals, the weight of rice, etc. Here, we directly measure the height or weight of an individual and all the measuring tools of physical measurement start from zero. Physical measurement is always accurate and quantitative, and there are some set of tools for physical measurement all over the world.

(ii) Mental measurement: Mental measurement is also known as 'educational measurement' or 'psychological measurement'. It is always relative and there is no absolute zero in case of mental measurement. For example, for measuring the intelligence of a person we have to take the help of intelligence tests which are subjective in nature. Through his response, we can know the level of intelligence of the person concerned. Mental measurement is both qualitative and quantitative in nature, and there are no fixed tools for such measurement i.e., the same set of tools may not be applied to different types of persons.

The application of the principles of measurement in the field of education is known as 'educational measurement'. In the educational system, measurement is the quantitative assessment of performance of the students in a given test. It can be used to compare performance between different students and to indicate the strengths and weaknesses of the students. It helps in classifying students into homogenous group, to assign educational and vocational guidance and to provide remedial measures to the low achievers. Measurement is a tool in the hands of educational psychologists to study human behaviour. Educational psychologists take the help of different valid and reliable psychological tests to know the level of different traits within an individual. The different kinds of such tests are: intelligence test, achievement test, attitude test, aptitude test, interest inventory, personality test, etc. The methods used for these tests are: observation, interview, checklist, rating scale, examinations, cumulative record card and anecdotal records etc.

In the teaching–learning situation, teachers should be competent enough to measure the student's achievement, intelligence, attitude, aptitude etc. To develop competency among the teachers in educational measurement, Ebel has suggested the following measures:

(i) Know how to administer a test properly, efficiently and fairly.

(ii) Know how to interpret test scores correctly and fully, but with recognition of their limitations.

(iii) Know how to select a standardized test that will be effective in a particular situation.

(iv) Know how to plan a test and write the test questions, to be included in it.

(v) Know the educational uses as well as the limitations of educational tests.

(vi) Know the criteria by which the quality of a test should be judged and how to secure evidence relating to these criteria.

NOTES

Characteristics of good measurement tool

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

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

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

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

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

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

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

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

(viii) Should be easy in scoring: The scoring procedure of the test should be clear and simple. The scoring directions and adequate scoring key should be provided to the scorer so that the test is easily scored.

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

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

Nature of Educational Measurement and Evaluation

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

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

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

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

NOTES

Characteristics of evaluation

The characteristics of evaluation are as follows:

- It is a systematic process.
- It measures the effectiveness of learning that experiences provide.
- It measures how far the instructional objectives have been achieved.
- It uses certain tools like tests, observation, interview etc.
- It is a continuous process.
- It is a subjective judgment.
- It is philosophical in nature.
- It includes quantitative description, qualitative description and value judgment.
- It gets data from measurement.
- It not only determines the magnitude, but also adds meaning to measurement.
- It involves values and purposes.

Evaluation from educational angle

Anything that needs to be evaluated has certain aims and objectives, and through evaluation we assess how far these objectives have been fulfilled. From an educational angle, we can evaluate many aspects which are the part and parcel of an educational system such as:

(i) Evaluation of a school site (with reference to its location, building, hygienic condition, strength of students and teachers etc.).

(ii) Evaluation of a school programme (school syllabus, co-curricular activities, guidance programmes etc.).

(iii) Evaluation of teaching methods (with reference to aims, purposes, suitability and efficacy).

(iv) Evaluation of total programme of instruction (with reference to cognitive, affective and psychomotor domain).

(v) Evaluation of school administration; discipline, control, management and organization.

(vi) Evaluation of textbooks and other course materials.

(vii) Evaluation of students' growth etc.

The steps involved in an evaluation process take place in a hierarchy. These steps are:

- Evaluating
- Planning of appropriate learning experiences
- Selecting appropriate teaching points
- Specification of desired student behaviour
- Identification and definition of specific objectives
- Identification and definition general objectives

CHECK YOUR PROGRESS

1. Define measurement.
2. What do you understand by objectivity of a test?
3. How does J.M. Bradfield define evaluation?

1.3 DIFFERENT TYPES OF MEASURING SCALES

'Scaling' in social sciences refers to the process of measuring. Scaling techniques are deployed in the service of estimating different individual levels of extraversion, or level of satisfaction. Comparative and non-comparative methods of scaling are the most common manifestation in the social sciences. With comparative scaling, items are directly compared with each other. For instance, 'Do you prefer beverage A or Beverage B?' In this case both beverages will be leading brands in the industry and the obtained feedback will allow market researches develop meaningful insight into the nature of market preferences. In non comparative scaling, each item will be subject to scaling independently of others. Questions formed under this will be more subjective in nature. For instance, it will ask an individual to state the exact reason for preferring one beverage over another.

Functions of Measurement

Lee J. Cronbach (1949) classified all applications of mental measurement under the following three main functions:

(i) Prognosis function: The first function of measurement is the prognosis function. Tests reveal differences among people's performance at this movement. All decisions involve prediction when psychological test is mentioned, for example, the IQ test that is given to students in school to predict their academic performance. The measurement provides the extent of a variable which has the specific purpose of predicting future behaviour. The prognosis function has an administrative function such as classification, selection, promotion and gradation of students. The college students ask counsellors to help them choose the best curriculum or job. Thus, guidance and counselling services are also based upon the prognosis function of measurement i.e., the effectiveness of method, instruction and treatment evaluated on the basis of student, achievement.

(ii) Diagnosis function: The second major function of measurement is diagnosis. The prognosis function reveals the level of student with regard to certain characteristics, whereas the diagnosis function identifies the weakness of the student-learning. The remedial instruction can be prepared on the basis of diagnosis. It also implies the prediction, but there is considerable justification in listing diagnosis as a separate function of measurement. The diagnostic function establishes the cause–effect relationship, but prediction implies the simple relationship. The instructional procedure can be improved by this function of psychological measurement.

NOTES

(iii) **Research function:** The third major function of measurement is verification of scientific hypotheses of research. The use of measurement for research purposes, however, is not much when compared to prediction and diagnosis. This is because a measurement is usually considered a completely valid measure of certain human characteristics. An investigator must treat test scores in this experiment as an accurate quantification of real and useful variable. Measurement provides a more objective and dependable basis for comparison than rough impressions. Thus, the valid generalizations are made on the basis of accurate measurement.

Methods of Measurement

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

The methods of measurement are as follows:

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

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

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

(iv) **Interview:** Interview is a face-to-face interaction between one interviewee and one interviewer or more than one interviewers. There are certain things which an individual does not want to express and they can be only assessed through interviews. The interview schedules may be used and the interviewer through a better personal support, and in congenial atmosphere, can succeed to bring out the inner feelings of the interviewee through carefully planned interviews.

(v) **Checklist:** A checklist consists of a series of items which needs response from the respondent. The presence or absence of an item may be indicated

by 'Yes' or 'No' (by a '√' or 'X' against the items). Checklists are popularly employed for appraisal of studies, school buildings, textbooks, outcomes, instructional procedures etc.

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

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

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

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

1.3.1 Need for Measurement and Evaluation in Education

The process of evaluation consists of the following three aspects:

(i) Quantitative measurements

(ii) Qualitative assessment

(iii) Teachers' opinion

(i) Quantitative measurement

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

(ii) Qualitative assessment

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

(iii) Teachers' opinion

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

Importance of Continuous and Comprehensive Evaluation

It does not suffice in the modern era that one is good only in academics. There are lots of opportunities for the students who somehow could not perform satisfactorily in academics. These activities need to be evaluated at a regular basis and are classified into scholastic and non-scholastic areas. In order to check the level of achievement and performance both in scholastic and non-scholastic areas, it is imperative that continuous and comprehensive evaluation should be practised in schools.

Evaluation on a regular basis should be used to monitor individual progress and to show the achievement level of each student in scholastic and non-scholastic areas throughout the year. With the help of such an evaluation, it can be assessed whether the objectives mentioned in the curriculum have been achieved or not. The objectives are set in order to improve not only cognitive domain in the students but also to the affective and psychomotor domains. This is so because these domains are complementary to each other and the absence of any one of these can create a sense of lack in the life of an individual.

The assessment of the achievement and performance in all the three domains is 'comprehensive' evaluation. The students' education is a continuous process which is a result of their continuous learning both in formal and informal settings. They go through various experiences such as teaching, learning, practical observations, self study, etc. The expected behavioural changes in children are reinforced by their direct involvement and self-learning. Objective records of such changes are required to be maintained from time to time. They form a major part of continuous evaluation. It is obvious that evaluation of behavioural modifications observed in the students' personalities are to be undertaken continuously. These modifications may be related to the development of intellect, emotions and practical skills.

The evaluation in school should aim at assessing these modifications, stated in terms of behavioural objectives, continuously and comprehensively through formal or informal techniques. The techniques to be used may be in the form of a written

examination or observation tests, group work or individual assignments. Provision has to be made for their periodic evaluation. In the course of the teaching–learning process, it is of paramount importance to aim for a continuous and comprehensive evaluation that uses formal or informal assessment. These assessments may be based on observations that are made taking into account the pupil's overall progress by asking them oral or written questions individually or in a group.

The national curriculum framework of school education prescribes social and qualities that all children need to develop regardless of their academic aptitude. In other words, even if the student is not able to compete academically, he or she can explore other capabilities such as arts, humanities, sports, athletics, and music etc. Excelling in any field can increase the students' confidence manifold and may even inspire them to put in more efforts into the curricular studies. Students, with the help of this revolutionary evaluation system, can finally understand that their competition is with themselves and not with the others.

Not just children, even teachers benefit from this kind of evaluation since they are encouraged to take important decisions regarding the learning process, efficiency, ensuring quality, and accountability.

Continuous and comprehensive evaluation helps to empower teachers in the following ways:

- They get pointers on how to improve the evaluation procedure.
- They can formulate ways to help students master competencies in which the students are falling behind through planned remedial steps
- They can select and deploy media such as audio visual aids for teaching in classrooms, especially subjects which cannot be understood purely with theoretical learning.

To sum up, following are the reasons why continuous and comprehensive evaluation is important to improve the overall education system in India:

- For making sure that learning in schools is not just about rote methods
- For ensuring that the examination system is more flexible and tests capabilities of a wider range
- For making sure the curriculum enriches the individual as a whole besides promoting academic learning
- For ensuring that classroom knowledge is relevant to life outside as well

Check Your Progress

4. What is scaling?
5. What is a test?

NOTES

1.4 PLACEMENT, DIAGNOSTIC, FORMATIVE AND SUMMATIVE EVALUATION

Evaluation is a process of judgement of value or worth of a process or product, which may be the achievement, aptitude, interest, skill or other aspects of student's personality or the method of teaching and learning. There are many types of evaluation. These include formative, summative, placement and diagnostic evaluation.

(i) Formative Evaluation

The evaluation which is done during the teaching–learning process to assess the ongoing termed formation of knowledge and understanding of students is called as formative evaluation. It answers the following pertinent questions regarding the process and product of teaching and learning:

- Whether the objectives of teaching a particular topic or content or course or curriculum are being achieved?
- Whether student's cognitive, affective and psychomotor domains are forming properly?
- Whether learners' scholastic and non-scholastic areas of personality are developing properly?
- Learners' progress is satisfactory or not?
- Whether the predetermined objectives of any on-going educational program are being fulfilled?

The formative evaluation is a monitoring type of evaluation which is used to monitor the progress of students during the class, course or session. After formative evaluation, feedback is given to students, so that they can proceed accordingly. The formative evaluation is aimed at improving the quality of teaching-learning process.

(ii) Placement Evaluation

Through placement evaluation, the entry behaviour of the student is assessed. In this case, the students are given admission to new courses according to their intelligence, attitude, motivation, aptitude etc. This type of evaluation questions:

- Does the student possess the knowledge and skills needed to begin the planned instruction?
- To what extent has the student already developed the understanding and skills that are the goals of the planned instruction?
- To what extent do the student's interests, work habits and personality characteristics indicate that one mode of instruction might be better than other?

The goals of placement assessment are to determine for each student the position in the instructional sequence and the mode of instruction that is most beneficial. For example, the B.Ed. entrance test is conducted to give admission to the students' in B.Ed. course. This type of evaluation is called 'placement evaluation'.

(iii) Diagnostic Evaluation

It is concerned with the persistent learning difficulties that are left unresolved by the corrective prescriptions of formative assessment. It aims at identifying or diagnosing the weaknesses of students in a given course of instruction. Diagnostic evaluation involves the use of specially prepared diagnostic tests and various observational techniques. The aim of diagnostic assessment is to determine the causes of persistent learning problems of students and to formulate a plan for remedial action. When a teacher finds that in spite of the use of various alternative methods and techniques, the student still faces learning difficulties, he takes recourse to a detailed diagnosis. This type of evaluation includes vision tests, hearing tests and other tests used to determine how the student approaches a reading assignment, as well as it examines whether the student relies on pictures, sound, context clues, skip over unfamiliar words, etc.

(iv) Summative Evaluation

As the name indicates, summative evaluation is done at the end of a course semester, or a class or topic. It is meant to evaluate the quality of the final product and to find out the extent to which the instructional objectives have been achieved.

No remedial teaching is given after summative evaluation. The process of certification is done on the basis of the results of summative evaluation. Results of this evalauation reflect the effectiveness of the curriculum transaction process. Important examples of summative evaluations are annual exams, semester end exams and terminal exams. This is much more concerned with judging about the final product. The important tools of summative evaluation are achievement test, rating scales, project evaluation by experts, interviews, viva-voce examination, etc. The characteristic features of summative evaluation are as follows:

- This evaluation is conducted at the end of a topic, class, chapter, unit or course of instruction.
- Evaluation results give the final progress of the students in a class, in a topic, in a unit, in a course or in any educational programme.
- Summative evaluation results are used for preparing merit list, finalizing position, taking decisions of pass/fail/promotion and awarding degrees or diplomas.

Difference between formative and summative evaluation

Take 1.1 summarizes the difference between formative and summative evaluations.

Table 1.1 *Difference between Formative and Summative Evaluations*

Formative Evaluation	Summative Evaluation
1. Conducted during the process of teaching and learning, during the class, during the semester or a session. 2. Determines the level of achievement in a small task learned in a short duration. 3. Conducted regularly during the class, course or session. 4. Gives limited generalization. 5. Limited content area and ability are covered.	1. Conducted at the end of the process of teaching and learning, e.g., at the end of class, at the end of semester, at the end of session, etc. 2. Determines the level of achievement in a major task learned in a longer duration. 3. Conducted at the end of the course or session or programme. 4. Gives broad generalizations 5. Large content areas and abilities are covered.

External Evaluation

The evaluation procedure in which the evaluators or examiners are invited from outside is called external evaluation. The teachers who are teaching a particular group of students are not involved in the evaluation of their students. Hence, the teachers and the examiners are different; they are never common. The teacher who is teaching a particular group of students or class is not the examiner of the same group of students. Their performances or achievements are judged or evaluated by outside teachers.

External evaluation includes all board examinations, viva conducted by expert from outside, annual examinations in schools in which paper comes from outside and not set out by the teachers of the same school, etc.

The significance of external evaluation is that it eliminates the element of bias from the evaluation system.

Internal Evaluation

Internal evaluation is the evaluation in which the teacher and the examiner are the same person. The same teacher teaches a particular subject and himself or herself sets the paper and evaluates the achievement of the students. No external expert is invited in this type of evaluation.

Class test, unit test, weekly test, monthly test, quarterly test, etc. are the examples of internal evaluation

Significance

- It helps in finding out students' progress.
- Remedial teaching can be organized for weaker students.
- Teachers himself or herself comes to know about his/her own and their students' weak areas.
- It helps in building strong foundation of students.

CHECK YOUR PROGRESS

6. What are the goals of placement assessment?
7. What is the aim at diagnostic evaluation?

1.5 ROLE OF TEACHERS IN AN EVALUATION PROGRAMME

Teaching remains central to both learning and assessment. There is an interrelatedness between teaching objectives (ends), learning experiences (means) and assessment (evidence of what is taught and learnt). Assessment is the process of determining the following:

(i) The extent to which an objective is achieved.

(ii) The effectiveness of the learning experiences provided in the classroom.

(iii) How well the goals of teaching have been accomplished.

In assessment, one has to know where students were at the beginning if we are to determine what changes are occurring. It involves obtaining a record of the changes in pupil by using appropriate methods of appraisal; judging how good the changes are in the light of the evidence obtained and ensuring change is ushered in teaching technology and also in learning technology.

Thus, assessment comes in at the planning stage when teaching objectives are identified. At every point of learning, it is an attempt to discover the effectiveness of the learning situation in evoking the desired changes in students.

Assessment is integrated with the whole task of teaching and learning and its purpose is to improve learning and not merely to measure its achievement. In its highest sense, it brings out the factors that are inherent in student growth such as proper attitudes and habits, manipulative skills, appreciations and understanding in addition to the conventional acquisition of knowledge.

It has been rightly observed, 'The definition of assessment places it in the stream of activities that expedite the educational process; these activities can be reduced to four essential steps: identification of educational objectives, determination of the experiences students must have to attain these objectives, knowing the pupils well enough to design appropriate experiences and assessing the degree to which pupils attain these objectives.'

Objectives provide the starting point on which are based all the learning experiences are based, which in turn are the material of assessment. Our teaching objectives are the changes we wish to produce in the child. The changes that must take place through education are represented in:

1. The knowledge children acquire.
2. The skills and abilities children attain.

NOTES

3. The interest children develop.
4. The attitudes children manifest.

If education imparted is effective, then the child will behave differently from the way he did before he came to school. The pupil knows something of which he was ignorant before. He understands something which he did not understand before. He can solve problems he could not solve before. He can do something which he could not do before. He revises his attitudes desirably towards things.

These objectives must involve points of information, the skills and attitudes to be developed and interests that could be created through the particular topic or subject taken up for work in the classroom. A statement of classroom objectives:

1. Serves as a basis for the classroom procedures that should provide for suitable experiences to the children.
2. Serves as a guide in seeking evidence to determine the extent to which the classroom work has accomplished what it set out to do.

A learning experience is not synonymous with the content of instruction or what the teacher does. Learning results from the active reaction of the pupil to the stimulus situation which the teacher creates in the class. A pupil learns what he does. He is an active participant in what goes on in the class. Changes in a pupil's way of thinking and developing concepts, attitudes and interests have to be brought about gradually. No simple experience will result in the change. Many experiences, one reinforcing another, will have to be provided. They may have to be repeated in increasing complexity or levels in meaningful sequence extended over a period of time. A cumulative effect of such experiences will evoke the desired change of behaviour with reference to a specific objective.

The following considerations will be useful in the selection of such experiences:

1. Are they directly related to goals?
2. Are they meaningful and satisfying to the learners?
3. Are they appropriate to the maturity of the learners?

Why to Assess

1. **To assess the achievement of the students.** The abilities and the achievements of the students must be assessed. Examinations are conducted to discover whether or not the child had been able to acquire a certain amount of knowledge or skill.
2. **To measure personality.** They are used to test the power of clear thinking, quickness of mind, calmness and perseverance.
3. **To measure the efficiency of teachers and of the school.** They provide a suitable occasion for the authorities to measure the efficiency of the teachers. The efficiency of the institution is also measured through the examinations. They provide a proper occasion to the teachers to know whether or not their methods of teaching are appropriate.
4. **To help in diagnosis.** They help parents to know the progress of their children. They help to discover the specific weak points of an individual or

class and thus give an opportunity to the teachers as well as to the taught to remove these defects.

5. **To act as an incentive.** Stimulation to work hard is provided to the students through the institution of examinations. Some objectives are placed before the students and for the realization of those objectives the students develop in them the habits of constant hard work.
6. **To help in prognosis.** The examinations have a prognostic value as well. With this device, the aptitudes of the students are determined.
7. **To provide uniformity of standard.** The external examinations facilitate the problem of uniformity of standards attained by the students of different institutions.
8. **To help in guidance.** They facilitate the work of grouping individuals for the purpose of teaching by bringing those together who have more or less the same attainment.
9. **To measure fitness for admission to higher courses.** They are designed to determine the capacity and fitness of the candidate to pursue higher courses of general or professional study or training. Examinations which serve this purpose are called entrance or qualifying examinations.
10. **To help in selection by competition.** The examinations are also conducted to select the best candidates for appointment to public service or for award of prizes and scholarships.
11. **Study of every subject.** The students study all the subjects including what are known as 'dull ones' for fear of examinations.
12. **Parents' point of view.** Examinations provide an opportunity to the parents to know where their wards stand and whether money spent on them is being utilized properly.
13. **To certify competency.** A technological society needs qualified doctors, engineers and scientists, and so on, and the examination system certifies their presence even if inadequately.
14. **To link schools with the world.** Examinations fulfill an important role in linking schools with each other and with the world.

1.6 TAXONOMY OF EDUCATIONAL OBJECTIVES

The study of classification is known as taxonomy. You have various objectives of education related to personal needs, social needs, community needs, national needs, global needs, and various aspects to your personality. The first and the most successful effort in this direction was made by American educational psychologist Benjamin Bloom and his co-worker, who began their work in 1948. Bloom was teaching educational psychology at the University of Chicago. He developed and proposed the classification of educational objectives in 1956, which is popularly known as Bloom's Taxonomy of Educational Objectives. This is the classification of the different aims and objectives, which are required to be achieved through curriculum of any

NOTES

course and involves knowledge, skills, attitude values and abilities that a teacher, teacher educator or curriculum developer sets for students as teaching-learning objectives. This classification fits properly for all subjects and courses whether academic or technical. In all the areas of study, you try to develop content knowledge and understanding, skills, and values related to it. You will find these in the classification of educational objectives by Bloom.

Bloom's taxonomy of educational objectives also fits properly with Mahatma Gandhi's views on education. Mahatma Gandhi defines education as, 'By education, I mean the all-round drawing out the best in child and man-body, mind and soul'. Gandhi's definition remarkably reflects the idea that education should develop the whole personality of the child by developing the three most important dimensions of personality, i.e., body, mind and soul. Mind indicates intellectual aspects; soul indicates the emotional, spiritual and moral aspects; and body indicates the development of skills which requires the use of body parts like fingers, hands, legs, eyes, vocal cord etc., for performing various skills.

Bloom classified educational objectives into three 'domains', namely: (i) cognitive domain, (ii) affective domain (iii) psychomotor domain, as shown in Figure 1.1.

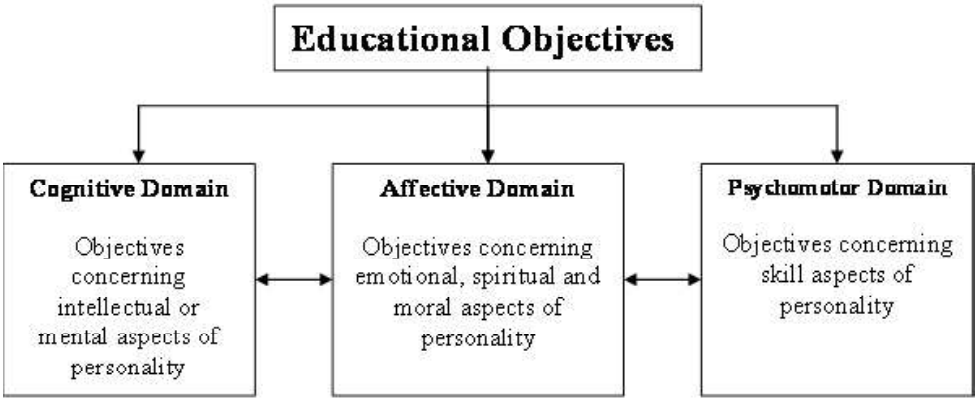


Fig. 1.1 *Domains of Educational Objectives*

All the three domains are not divided entirely by a water tight compartment. Each of the three domains is interrelated and achievements in one domain influence the achievement by other domain. However, all domains develop in a hierarchical order and constitute the whole personality of an individual. If an individual develops his or her three domains in any area of study, he or she is having a good personality, and if any one domain is underdeveloped, then he or she has not nurtured himself or herself as a groomed personality. Hierarchical means of learning at the higher level in any domain is entirely dependent on having acquired prerequisite knowledge and skills at lower level. The three domains may be represented in short as:

1. Cognitive domain (about knowing-intellectual aspects of personality of an individual)
2. Affective domain (about attitudes, feelings, interests, values and beliefs of an individual)

3. Psychomotor domain (about doing-skill aspects of the personality of an individual)

All the three domains have their own taxonomy or classification. In all domains, the levels of expertise or abilities are arranged in order of increasing complexity, i.e., in hierarchical order of difficulty levels. Learning outcomes that require higher levels of expertise require more effective teaching and more sophisticated classroom techniques and method of teaching.

Dimensions of learning basically form a framework of learning focused on preparing instructional planning, keeping in view cognition (the awareness part) and learning in practical classroom situations. This framework serves three major purposes. These are as follows:

1. It provides a process for planning and delivering curriculum and instruction that integrates much of the research on effective teaching and learning.
2. It offers a way of integrating the major instructional models by showing how they are connected and where the overlaps occur.
3. It provides a framework for organizing, describing and developing research-based teaching strategies that engage students in the types of thinking that can lead to meaningful learning.

The following five aspects of learning should be considered while finalizing curriculum, instruction and assessment:

- Attitudes and perceptions about learning
- Using knowledge meaningfully
- Identifying productive habits of the mind
- Acquiring and integrating knowledge
- Extending and refining knowledge

I. Cognitive Domain

Cognitive domain includes those objectives of education which attempt to develop mental faculties or intellectual abilities, i.e., the ability of knowing, understanding, thinking and problem solving. It develops our factual knowledge, conceptual understanding and all levels (lower, middle and higher) of thinking. It covers whole of mental abilities and mental operations.

Classification of cognitive domain

The classification of cognitive domain was done in 1956 by Bloom and is commonly known as Bloom's Taxonomy of Cognitive Domain (Bloom et.al, 1956). Cognitive taxonomy or classification of cognitive domain has knowledge-based goals. This domain has been classified into six major categories which are arranged in a hierarchical order based on the levels of complexities of cognitive or mental or intellectual tasks or operations. These are arranged from simple to complex and concrete to abstract starting from category one to category six, respectively.

1. **Knowledge:** This is the first and the lowest level of cognitive domain. It represents memory and constitutes recall and recognition of various facts, concepts, principles, theories and laws of physical science. No addition or deletion is done in this category; we simply recall and recognize things. In the revised Bloom's taxonomy, this category is given new name 'remembering'.

   Example: The symbol of iron is 'Fe'.

   The chemical formula of sulphuric acid is 'H_2SO_4'.

   Laws of motion were given by 'Isac Newton'.

A magnet has two poles– N (north) and S (south)

2. **Understanding:** This is the second level of cognitive domain and develops only after the development of the first category, i.e., knowledge or remembering in any particular area of study, also in physical science. Learners are expected to go beyond the level of recall and recognition. After having developed this level of understanding on any topic, learners become capable of doing the following major tasks, which in turn indicates that the learners have acquired the level of understanding in a given topic:
   (i) Translate, summarize or define the acquired knowledge in their own words.
   (ii) Describe, elaborate, extrapolate and explain natural phenomena or events or process or method, etc.
   (iii) Interpret the acquired information or knowledge in their own way and give their own examples. They can discriminate or differentiate between two or many objects or concepts. Classify and categorize various objects into groups on the basis of some criteria. Verify and generalize facts and concepts.

   Example: After having understood the structure of atom, learners not only recall protons, electrons and neutrons but also describe the structure of an atom. Now learners can also explain why an atom is neutral with the help of charges acquired by every fundamental particle and numbers of all three particles in an atom.

3. **Application:** After having acquired knowledge and understanding levels of any topic (may be a fact, concept, principle, theory or law), learners should be able to apply them in their day-to-day lives. Application of any concept, principle, theory or law in daily life and solving problems of varied nature is impossible without its knowledge and understanding. Unless the learner is able to apply whatever knowledge and understanding he or she has acquired, it has no meaning at all and indicates that the learners have not understood the content properly. by applying or implementing the gained knowledge and understanding of various contents, you can solve many problems of daily life, under concrete and abstract situations.

   Example: If learners know and understand the importance of natural resources, underground water crisis, electricity supply and demand relationship

and other such problems of daily lives, they will take care of these things in their day-to-day life, and by applying this understanding, they will try to minimize wastage of water and electricity in their homes, schools and society by proper and judicial use of these things.

4. **Analysis:** This is the fourth higher level category of cognitive abilities. At this stage, learners develop the potential to analyse and breakdown the whole into its various components or constituents and detect the relationship and organization of its various components. Learners develop the ability to break a law and theory into its various inherent facts, concepts and principles on the basis of which that theory or law has been created or proposed.

   Example: Learners are taught about the laws of motion. Suppose they know and understand the third law of motion which states, 'to every action there is an equal and opposite reaction'. They have also developed the ability to apply this knowledge and understanding in their daily lives. If the analytical ability has developed, they would be able to analyse this law in the event of some likely situations. They would also be able to describe its every concept like action and reaction. One can analyse anything if he or she has knowledge and understanding of that thing and also has the potential to apply it. In the process of analysis, three tasks are performed in general. These are as follows:

   (a) Analysis of elements or constituents making the whole

   (b) Analysis of relationship among various constituents

   (c) Analysis of the organizational patterns of the constituents

5. **Synthesis:** This is the process of putting together various constituents to make a whole. This is a higher level thinking ability and is complex in nature, which involves the creation of a new pattern or structure by manipulating various constituents. It has the elements of creativity attached with it. Development of creative personality requires this level of cognition to be achieved by the learners. All creative people have this ability in common. Synthesis involves the following three things:

   - Development of a unique communication
   - Development of a plan, procedure or proposed set of operation
   - Development of a set of abstract relations

6. **Evaluation:** This is the process of judgment about the worth or value of a process or a product. It includes all the content, i.e., facts, concepts, principles, theories and laws of physical sciences. It is the highest and the most complex level of cognitive ability and involves all the five categories discussed earlier. It is a quantitative as well as a qualitative process. It leads to the development of decision-making ability among the learners and involves judgment in terms of internal as well as external criteria.

Assessment of cognitive learning

Cognitive learning is defined as the acquisition of knowledge and skills by the mental process. Cognition includes representation of physical objects and events and other information processing.

NOTES

The cognitive learning is a result of listening, watching, touching or experiencing. Cognitive learning is mainly awareness towards the environment—the environment that provides some meaning to acquired knowledge.

After these activities, it is important to process and remember the information. No motor movement is needed in cognitive learning and the learner is active only in a cognitive way. Cognitive learning includes symbols, values, beliefs and norms.

When we say the word 'learning', we usually mean 'to think using the brain'. This basic concept of learning is the main viewpoint in the Cognitive Learning Theory (CLT). The theory has been used to explain mental processes as they are influenced by both intrinsic and extrinsic factors, which eventually bring about learning in an individual.

CLT implies that the different processes concerning learning can be explained by analysing the mental processes first. It posits that with effective cognitive processes, learning is easier and new information can be stored in the memory for a long time. On the other hand, ineffective cognitive processes result in learning difficulties that can be seen anytime during the lifetime of an individual.

Insight and motivation

In CLT, motivation and purpose are much the same. The learner's goal is the end result he anticipates and desires. The goal controls the behaviour of the learner. The teacher's most important responsibility is to help the learner find worthwhile goals which may be clear and realistic. They should recognize and use the prominence of social and achievement motives in school learning. The teacher must know about what is familiar to students and then he must introduce elements of novelty but not too rapidly. A good teacher must pace his presentation to maintain the interest and attention in learning.

The teachers' management of conflicting motives may be an important factor in student's success.

Guiding exploration and action

The cognitive theorists in teaching of reading begin with thoughts interesting and understandable to the learner. In every type of instruction, we start with meaningful wholes.

- An attempt is made to focus attention on elements and relationship that determine the correct response.
- Teachers' guidance must match the students' level of thought or ways of working. If a student has not advanced above the level of concrete thinking, information presented symbolically will not help him.
- The teacher can help students to find purpose or order in learning.

The formation of concepts may be regarded as the organization of experience. The teacher's role is to use appropriate means to clarify the critical features of both old and new experiences.

NOTES

Thinking skills required in cognitive learning

The following critical thinking skills are required in cognitive learning:

1. Convergent thinking

Convergent thinking is a term introduced by the American Psychologist Joy Paul Guilford. It generally means the ability to give the 'correct' answer to standard questions that do not require the utilization of much creative faculty. Divergent and convergent thinking skills are both important aspects of intelligence, problem solving and critical thinking. Bringing facts and data together from various sources and then applying logic and knowledge to solve the problem to achieve an objective is known as convergent thinking.

Assessing convergent thinking: Convergent thinking can be assessed by administering standard IQ (Intelligent Quotient) tests, by various recognition or knowledge tests, logical discussions and by giving some problems to the students.

2. Divergent thinking

Divergent thinking is thinking outwardly instead of inwardly. It is the ability to develop original and unique ideas and then come up with a solution to a problem or achieve an objective.

The goal of divergent thinking is to generate many different ideas about a topic in a short period of time. It involves breaking a topic down into its various component parts in order to gain insight into the various aspects of the topic. Divergent thinking typically occurs in a spontaneous, free-flowing manner, such that the ideas are generated in a random, unorganized fashion.

Self analysis

The following questions may help to find out the potential of a learner:

- What are my activities during a normal day?
- What do I know about?
- How do I spend my time?
- What are my areas of expertise?
- What would I like to change in my world or life?
- What are my strongest beliefs and values?
- What am I studying in school?
- What bothers me?
- What do I like? What are my hobbies? What are my interests?

Topic analysis

The following questions may help in refining a large topic into a specific, focused one:

- What are the most important aspects of some specific things?
- What are the effects of a thing?

NOTES

- How has an object changed? Why are those changes important?
- What are the different aspects of anything you can think of?
- What are the smaller parts that comprise an object?
- What do I know about a thing?
- What suggestions or recommendations can be made about that thing?
- Is something good or bad? Why?

Techniques to stimulate divergent thinking

The following are some of the techniques used to stimulate divergent thinking:

- **Brainstorming:** This is a technique which involves generating a list of ideas in a creative, but unstructured manner. The goal of brainstorming is to generate as many ideas as possible in a short period of time. All ideas are recorded during the brainstorming process.
- **Recording ideas:** By recording ideas, one can create a collection of thoughts on various subjects that later becomes a source book of ideas.
- **Free writing**: The idea is to write down whatever comes to mind about the topic, without stopping. This can help generate a variety of thoughts about a topic in a short period of time.
- **Mind or subject mapping**: This involves translating brainstormed ideas into the form of a visual map or picture.

3. Critical thinking

Critical thinking is described as reasonable reflective thinking focused on what to believe and what not to believe. It is that mode of thinking — about any subject, content, or problem — in which the thinker improves the quality of his or her thinking by skillfully taking charge of the structures inherent in thinking and imposing intellectual standards upon them. Michael Scriven, british polymath and academic, and Richard Paul, an internationally recognized authority on critical thinking, believe that critical thinking is the intellectually disciplined process of:

- Analysing the situation
- Synthesizing two or more pieces of information
- Applying knowledge
- Active and skilful conceptualization
- Evaluating gathered information based on certain universal intellectual values

***Characteristics of critical thinking*:** It comprises various modes of thinking, such as scientific thinking, mathematical thinking, historical thinking, anthropological thinking, economic thinking, moral thinking and philosophical thinking. The following are its four characteristics:

1. It is self-guided, self-disciplined thinking which attempts to reason at the highest level of quality.
2. Critical thinking of any kind is never universal in any individual; everyone is subject to episodes of irrational thought.

3. It is affected by the motivation.
4. Critical thinking can be seen as having two components: (a) a set of information and belief generating and processing skills, and (b) the habit, based on intellectual commitment to use those skills to guide behaviour.

***Importance of critical thinking skills*:** Critical thinking skills are essential to:

- Learn how to approach problems logically and confidently
- Balance using both the right and left sides of the brain
- Make wise decisions in life
- Put oneself on the path to knowledge

The list of core critical thinking skills includes observation, interpretation, analysis, inference, evaluation, explanation and meta-cognition. There is a reasonable level of consensus among experts that an individual or group engaged in strong critical thinking gives due consideration to the procedure involved.

Critical thinking calls for the ability to:

- Recognize problems
- Find workable means to solve those problems
- Comprehend and use the text with accuracy and clarity
- Interpret data to evaluate arguments
- Recognize unsaid assumptions
- Understand the importance of steps in problem solving
- Collect relevant information
- Recognize the existence (or non-existence) of logical relationships between propositions
- Draw conclusions
- Cross-check conclusions

4. Problem solving

Problem solving is a mental process that involves discovering, analysing and solving problems. The ultimate goal of problem-solving is to overcome obstacles and find a solution that best resolves the issue.

Formalized learning theory developed in late 1930, when proponents of various approaches attempted to build their own theory to explain the problems of learning. A theory of learning cannot be defined to satisfy all interested persons. We can quote the definition of a theory as 'a provisional explanatory proposition or set of propositions, concerning some natural phenomena and consisting of symbolic representations of: (a) the observed relationships among independent and dependent variables, (b) the mechanisms or structures presumed to underlie such relationships, or (c) inferred relationships and underlying mechanisms intended to account for observed data in the absence of any direct empirical manifestations of the relationships' (*Learning Theories* edited by Melvin H. Marx).

NOTES

Approaches to problem solving

Traditionally, two different approaches have been mentioned by psychologists, adhering to two families of learning theories: (a) cognitive field theory, and (b) stimulus-response theory.

Cognitive field theory emphasizes the importance of perception of total situation and relationship among its components, and restructuring the cognitive field. German Psychologist and Phenomenologist Wolfgang Köhler conducted his classical experiments on a chimpanzee named Sultan to study the process of problem solving in animals. He, from his study on problem solving, proposed that solution of a problem is arrived at, all of a sudden, after some initial efforts by the individual. Many studies have been conducted on children and adults which confirm that solution of a problem is reached, all of a sudden, through insight into the situation.

The second point of view has been advanced by stimulus-response theorists who emphasize the importance of trial and error. They hold that a problem is solved through a gradual process of elimination of errors and putting together correct responses. There has been considerable controversy as regards the superiority of one approach over the other as an interpretation of problem solving. Some psychologists are of the opinion that cognitive field theory, approach is most effective for solving problems which require higher mental processes, while stimulus-response approach is effective for solving simple problems.

To do away with the controversy of cognitive and stimulus response theorists' approach, American psychologist Harry Harlow (1959) proposed a third explanation. His approach is more realistic and rational in nature. He conducted a series of experiments on monkeys and human subjects of low mental abilities. He presented his human subjects with simple problems of discrimination. He observed that in the beginning, his subjects showed trial and error behaviour to solve a series of problems, but he noticed that when similar problems were presented to the subjects in future for the first time, they made correct discrimination. The later stage appears to be insightful learning, that is, suddenly getting the problem solved. According to Harlow, the underlying assumption is that in the previous trial and error learning, the subjects have learned 'how to learn'. They acquired what he called a learning set. They acquired a method of learning that transferred positively to other problem situations of similar type.

Harlow says, 'Generalizing broadly to human behaviour, we hold that original learning within an area is difficult and frustrating, but after mastery of the basic facts, learning within the same area becomes simple and effortless.'

The steps in problem solving

In order to correctly solve a problem, it is important to follow a series of steps. Many researchers refer to this as the problem-solving cycle, which includes developing strategies and organizing knowledge. The following are the steps required in the process of problem solving:

1. **Identifying the problem:** While it sounds like the simplest thing to do, it actually is not. It is very common for people to identify the reason incorrectly, which obviously renders all following efforts useless.

2. **Defining the problem:** Once the problem has been correctly discovered, the next step is to define it. Defining the problem carefully offers solutions and insights into the problem.
3. **Forming a strategy:**Next, a strategy must be devised which would be as per the individual's mindset, attitude, experience and available resources.
4. **Organizing information:**The strategy should ideally lead to a solution, but first, the information needs to be organized in a coherent manner, such that it leads naturally to the best possible solution.
5. **Allocating resources:** It is good to allocate more resources to solve a problem, so that the available resources can be used to find out a solution.

Problem solving techniques

These techniques are usually called problem-solving strategies. These are as follows:

- **Brainstorming:** Suggesting a large number of solutions or ideas and combining and developing them until an effective solution is found
- **Abstraction:** Solving the problem in a model of the system before applying it to the real system
- Assessing the output and interactions of an entire system
- **Lateral thinking:** Approaching solutions indirectly and creatively
- **Dividing and conquer:** Breaking down a large, complex problem into smaller, solvable problems
- Employing existing ideas or adapting existing solutions to similar problems
- **Analogy:** Using a solution that solved an analogous problem
- **Hypothesis testing:** Assuming a possible explanation to the problem and trying to prove one's perspective
- Synthesizing seemingly non-matching characteristics of different objects into something new
- Transforming the problem into another problem for which solutions exist
- Eliminating the cause of the problem
- Testing possible solutions until the right one is found

5. Decision making

Decision making can be regarded as the mental processes resulting in the selection of a course of action among several alternative scenarios. The end result of each decision making process is a final selection. The output can be an action or a suggestion.

Steps in decision making

The following are the steps that are to be followed in the decision-making process:

- Objectives are to be established first.
- Objectives must be classified and placed in the order of importance.
- Alternative actions must be developed.
- The alternative must be evaluated.

NOTES

- A tentative decision can be made.
- The tentative decision is evaluated and analysed.
- Few steps are followed that result in a decision model. This can be used to determine an optimal plan.
- In a situation of conflict, role-playing can be used.

Each step in the decision-making process may include social, cognitive and cultural problems. It has been suggested that becoming more aware of these obstacles allow one to better anticipate and overcome them. The following few steps can help in establishing good decision making:

- Creating and nurturing the relationships, norms and procedures that will influence how problems are understood and communicated.
- Recognizing that a problem exists.
- Identifying an explanation for the problem and evaluating it.
- Finding more suitable justice among many responses.
- Following through with action that supports the more justified decision. Integrity is supported by the ability to overcome distractions and obstacles, developing implementing skills and ego strength.

Decision-making stages

There are four stages that are involved in all group decision making. These stages, or sometimes called phases, are important for the decision-making process to begin. These were developed by communication researcher B. Aubrey Fisher. The four stages are as follows:

1. **Orientation stage**: This is introductory stage, when members meet for the first time and get to know each other.
2. **Conflict stage**: Once group members become familiar with each other, disputes, little fights and arguments take place. Nevertheless, group members eventually work it out.
3. **Emergence stage**: The group begins to clear up vague opinions by talking about them.
4. **Reinforcement stage**: Members finally make a decision, while justifying to themselves that it was the right decision.

II. Affective Domain and Formulation of Specific Objectives

The affective domain of Bloom's taxonomy of educational objectives is related to the development of emotions, values, attitudes, and the development of those aspects of personality which are more influenced by heart than the mind. It also includes the development of interests, appreciation, feelings, likes and dislikes towards something.

Classification of Affective Domain

The classification of this domain was done by American educational psychologists D. R. Krathwohl, B. S. Bloom and B. B. Masia in 1964. The categories in this domain are also arranged hierarchically from the lowest to the highest level of complexity.

1. **Receiving:** This is the ability, inclination and readiness of learners to receive information. It requires attention, awareness, listening, seeing and willingness on the part of the learners. These are preconditions of learning, personality development, and imbibing culture and values. It needs sensitization of learners to stimuli, phenomena or environment. On the whole, learners should be made receptive in their habit and attitude. Whatever you want learners to learn, you should make them receptive toward those things.

   Examples:

   - Reading newspapers, magazines, journals, books, reports, etc., of interest to the learner
   - Watching news, shows, reports, programmes as per interest
   - Listening patiently and attentively to teachers, parents, seniors, friends and more experienced persons
   - Having curiosity to learn from various sources

2. **Responding:** This is the second level objective under the affective domain. Learners are required to be responsive along with being receptive; otherwise it will not serve the purpose. Responding behaviour reflects that the learners are receiving or trying to receive. Continuity in attention and motivation behaviour (receiving) leads to the development of responding behaviour.

   This category of ability is represented by interest, which is the tendency to respond to a particular object or event or situation. This creates the way for two-way communication and facilitates the process of teaching and learning. Students 'listen' to the teachers attentively and 'respond' to them to give their reflection and share their experiences.

   Example:

   - Response of students in class
   - Interaction of students with teachers, friends and seniors on various issues or problems
   - Visit to clubs, libraries, museums and other knowledge resource centres
   - Participation in various activities, competition, seminars, conferences, cross word and such other programmes

3. **Valuing:** During the cyclic process of receiving and responding, learners are automatically inclined towards taking value judgment regarding the things they are concerned with. These things may be an object, an event, an idea, a rule, any ritual, a set norm or any traditional or modern aspects of our culture. Through the process of valuing, individuals set guidelines for regulating their own behaviour. Character formation or value inculcation in the growing generation is done through the following three sequential steps:

   (a) Value acceptance

   (b) Value preference

   (c) Value commitment

   Example: A class is taught by several teachers. All teachers practice various values in which some are common and some are unique for individual teachers. Students attend their classes and interact with them. They observe and analyse

NOTES

various values being practiced by their teachers. Through regular observation and analysis, students develop their own value based on their preference, acceptance and commitment.

4. **Organization:** Through the process discussed above, students absorb various values from their teachers, parents and society. They analyse various values absorbed from different sources and finally construct relatively enduring value system through the process of synthesis and organization of values for a balanced conduct and behaviour pattern. This leads to the development of a set value structure or philosophy of life for every individual. It assists individuals in decision-making process about conduct in real life situations and in forming opinions on major issues of social and personal concern.
5. **Characterization of values or value complex:** This is the highest level category of objectives under affective domain. At this level, individuals develop a set of values, attitudes and beliefs for themselves that build their character and give shape to their philosophy and personality. This process goes on continuously throughout life resulting into the shift of preference of various values, depending upon situation, age and experience.

Assessment of Affective Learning

The affective domain given by Krathwohl, Bloom and Masia includes the manner in which we deal with things emotionally, such as feelings, values, appreciation, enthusiasm, motivations and attitudes. The five major categories are listed from the simplest behaviour to the most complex in Table 1.2.

Table 1.2 *Five Major Categories Describing the Affective Domain*

Category	Key Words (Verbs)
1. **Receiving phenomena** (Awareness, willingness to hear)	Asks, choose, describe, follow, give, identify, locate, name
2. **Responding to phenomena** (Active participation on the part of the learners. Attends and reacts to a particular phenomenon)	Answer, assist, aid, complie, conform, discuss, greet, help, perform, practice, present, read, recite, report, select, tell, write
3. **Valuing** (Valuing is based on the internalization of a set of specified values, while clues to these values are expressed in the learner's overt behaviour and are often identifiable)	Complete, demonstrate, differentiate, explain, follow, form, initiate, invite, propose, read, report, select, share, study, work
4. **Organization** (Organizes values into priorities by contrasting different values, resolving conflicts between them, and creating a unique value system)	Adhere, alter, arrange, combine, compare, complete, defend, explain, formulate, generalize, identify, integrate, modify, order
5. **Internalizing values** (characterization): Has a value system that controls their behaviour	Act, discriminate, display, influence, listen, modify, perform, practice, solve

NOTES

Attitude and Values

'An attitude can be defined as a positive or negative evaluation of people, objects, events, activities, ideas, or just about anything in your environment.'

—Philip George Zimbardo, 1999

In the opinion of American sociologist Read Bain, attitude is 'the relatively stable overt behaviour of a person which affects his status'. An attitude is a state of mind or a feeling or disposition. It is important to have a positive attitude about work.

Values

Affective learning also involves internalizing a set of values expressed in behaviour. Teachers affect the values and behaviour in a student by setting examples by their own behaviour.

Attitude formation

Attitudes are expected to change as a function of experience, environment and education. American psychologist Abraham Tesser has argued that hereditary variables may also affect attitudes.

Measurement of attitude

A number of techniques for measuring attitudes are in use. However, they all suffer from different kinds of limitations. Largely, the different types of techniques focus on the components of attitudes, namely the cognitive, the affective and the behavioural components. The two basic categories that attitude measurement methods can be divided into are as follows:

1. Direct measurement, such as Likert scale
2. Indirect measurement, such as projective techniques

Direct observation

This is a simple and logical method which records the behaviour patterns of people under study. This method is widely used for various purposes. However, even if the individuals to be studied are easily accessible, observing the behaviour of a large sample of individuals is not practically feasible.

Direct questioning

This method involves asking pre-set questions on certain topics on which the individual's behaviours are to be evaluated. While it seems like the most straightforward approach to simply ask questions to test attitude, the results may not be accurate because an individual may try to hide his or her real opinions and attitudes.

Some other approaches

In projective techniques, attitude gauging objects are hidden and results are interpreted on the basis of pre-set criteria. While this technique overcomes some limitations of the direct observation technique, the projective technique falls short when it comes to objective and reliable interpretation of data.

Thurstone scale

The first official method for measuring attitude was formulated by Louis Leon Thurston, a US pioneer in the fields of psychometrics and psychophysics, in 1928. His objective was to measure people's attitudes regarding religion. The Thurstone scale contains statements regarding the issue in question and every statement is assigned a numerical value based on the value the evaluator considers it to have. When people have selected the statements for each question, the values are added up and the average is calculated, which corresponds to a particular attitude.

Likert scale

Likert introduced the Likert scale of attitudes in *A Technique for the Measurement of Attitudes*. The scale covered a range of attitudes, from extremely positive to neutral to extremely negative. This scale also consists of statements and the subjects are asked to express their opinion regarding the statement through a 0–5 point scale. Once each statement has been assigned a numerical value, again they are added up and the mean is calculated. This kind of scale is often used in career assessment programmes to gauge a learner's interests and tendencies, so that they can help themselves select the right career path.

The Likert scale is applied in the form of questionnaires. A Likert scale questionnaire would contain a statement, which would need to be evaluated by theindividual on the basis of the following kind of responses:

(a) Strongly disagree

(b) Disagree

(c) Neither disagree nor agree

(d) Agree

(e) Strongly agree

The individual will tick or circle one response for each statement or question. Each statement/question and its possible responses are together known as a Likert item. A Likert scale is, in turn, a sum of responses to multiple Likert items. A Likert scale is considered a 'balanced' form of attitude testing because it contains an equal number of positive and negative options.

In some cases of Likert scale questionnaires, the neutral response, such as 'neither disagree nor agree', is removed. Such an evaluation method is known as the forced choice method, because the individual cannot remain neutral on any statement. The forced choice method is used when it is thought that the individual might select the neutral response to avoid controversy or hide ignorance.

Generally, three kinds of common biases may render Likert scale questionnaires unreliable. These are as follows:

- **Central tendency bias:** In a professional setting especially, the individual may be reluctant to admit strong feelings about an issue.
- **Acquiescence bias:** The individual may feel obliged to agree with the presented statement.

- **Social desirability bias:** The individual may select an option to preserve the self-image or to be popular.

Once the respondent gives in the complete questionnaire, the responses may be evaluated individually or in a grouped form, depending on the pattern being studied. When studied in a grouped form, the items provide a score, which can be categorized and evaluated. The Likert scale is also sometimes known as the Summative Scale.

Interest and its measurement

The reason for finding out the interest of the students is to help them in getting their dream careers. This inventory helps the teachers in knowing the interest areas of their students so that they can encourage and get various opportunities to grow in a particular field.

Thus, an interest inventory is often used in career assessment. The goal of this assessment is to give insight into the students' interests, so that they may face less difficulty in deciding on an appropriate career choice for themselves. It is also frequently used for educational guidance as one of the most popular career assessment tools. The test was developed in 1927 by psychologist E. K. Strong to help people who were quitting the military jobs to find suitable jobs.

Supporting students

Prior to selecting a career, students need to identify the right path for themselves. This can be done through an assessment, which would help them get an insight into their own interests, preferences and personal styles. Analysing these aspects will direct them into identifying the right courses, jobs, internships and activities that are suitable for them.

Self-concept and its assessment

Self-concept defines how we assess ourselves as individuals or what we think of ourselves. Self-concept is a commonly used term. There are two aspects to the development of self-concept. These are as follows:

1. **The existential self:** This aspect of the self can be defined as,'the sense of being separate and distinct from others and the awareness of the constancy of the self'. (Bee 1992)
2. **The categorical self:** Once a child realizes that he hasa distinct identity, hegradually become aware of the world around him and his own place in the world. He starts relating to the world and starts thinking of himself as more than hisphysical characteristics, such as hair colour or height, etc. Finally, he becomes aware of the fact that others perceive himin a certain way, which may or may not be similar to how he perceives himself.

According to American psychologist Carl Rogers, self-concept has three different components. These are as follows:

1. The view you have of yourself (self-image)
2. How much value you place on yourself (self-esteem or self-worth)
3. What you wish you were really like (ideal self)

NOTES

Factors affecting self-concept

The following factors affect self-concept:

1. **How others react to us:** People's approval and recognition and affection helps to develop a positive self-image.
2. **Comparison with others:** Comparing oneself with people who seem to be doing better financially or more popular socially than oneself can lead to a negative self-image. However, if the comparison is made with people who are less successful, it would lead to a positive self-image.
3. **Social roles:** There are some roles that are associated with prestige and positive self-image, such as that of a doctor.
4. **Identification**: Self-concept is also influenced by the role we play in the group that we belong to.

III. Psychomotor Domain

Psychomotor domain is concerned with those objectives which are intended to develop various skills. For example: typing, painting, drawing, dissecting, preparing food, beautification, carpentry, book binding, sculpture, photography, operation of computer or any other machine, working with any tools to produce something. It includes all manipulative skills. For any motor activity or skill work, psychological readiness is an essential condition. If a person is psychologically ready, he or she will also be mentally ready and will act towards the desired, skilled work.

Classification

Psychomotor domain is classified by many psychologists. Some of them are Ragole (1950), E. J. Simpson (1966), J. P. Guilford (1958), R. H. Dave (1969), A. J. Harrow (1972), Allyn & Bacon (1994). In the following paragraphs, we are going to discuss the classification given by R. H. Dave (1969) in an adapted form. There are five categories under this domain, arranged from 1 to 5 in the order of increasing complexity, difficulty level and fineness in the skill being developed.

1. **Initiation-observation or observation-initiation:** For learning any skill (simple or complex), learners need psychological readiness. It is common to most of the learners that they hesitate of any skilled work at the time of beginning. They generally hesitate to take initiative. Contrary to this, there are some learners who are highly motivated to start as they have observed somebody doing it and they are highly motivated to perform that skill.

   If learners are not interested or motivated, they need to become motivated by means of promise of some reward, discussion, fulfillment of some desires and aspirations of the learners, etc. In the above case, 'initiation' is the first step towards the development of skill. On the other hand, if a learner has observed some person performing some skill and is highly motivated and encouraged to do that skill, then he takes initiative automatically. In this case, observation is followed by initiation.
2. **Manipulation:** When the learner is ready to take initiative, he observes others performing that skill and sees how performers are manipulating tools required

in that skill and he also starts manipulation of those tools to produce or copy that skill. Manipulation and observation work together continuously for quite some time. As a result, improvement in performance is achieved and it leads towards perfection. During this process, learners perform the following three tasks:

(i) Perform selected steps

(ii) Follow directions

(iii) Fix their performance through necessary practice

3. **Precision:** Repeated observation of the expert performers and continuous practice leads learners to the performance of the skill with a desired level of precision, i.e., accuracy and exactness. They reach at a higher level of refinement. They achieve this level through the following points of consideration:
   - Controlling faults
   - Eliminating errors
   - Reproducing the desired skill with precision
4. **Articulation:** This is the level at which learners bring some novel attributes or features to their skill performance in addition to the general attributes.
5. **Naturalization:** This is the highest level of performance in skill development. The act of the performer becomes automatic or natural. Achiever of this level of proficiency, which is rare, performs with the highest degree of refinement and convenience as natural as possible. For performer as well as audience or observer, it looks like an effortless performance.

Assessment of Performance or Psychomotor Learning

Performance assessments are designed to judge a student's abilities to use specific knowledge and research skills. Most performance assessments require the student to manipulate equipment to solve a problem or make an analysis. Rich performance assessments reveal a variety of problem-solving approaches, thus providing insight into a student's level of conceptual knowledge.

Performance assessment is a form of testing that requires students to perform a task rather than select an answer from a ready-made list. Experienced teachers or other trained staff then judge the quality of the student's work based on a pre-decided set of criteria. This new form of assessment is most widely used to directly assess writing ability based on the text produced by students under test instructions.

How does it work?

The student's work is assessed by teachers or other staff on the basis of pre-set criteria. The teachers assess the student's learning abilities when given specific instructions. It is important to understand how the assessment works. The following methods have been used successfully to do so:

- **Open-ended answers:** The student has to come up with the answer independently after some degree of analysis, either orally or in writing. No

NOTES

options are given and the students are usually asked to discuss their observations about some material given to them.

- **Extended tasks:** In this kind of assessment, the student is required to research, draft and revise and come up with a well-though out analysis of the given problem. It is like an assignment or a project. It may take the student many hours to complete.
- **Portfolios:** They are pictorial descriptions of the student's past work and the teacher's evaluation of it. Usually, portfolios include a person's best work and some work in progress which showcases the student's strengths.

What is common among all these methods is that they are all specific tasks completed by the student according to the instructions from the teachers and the students are made well-aware of the standards expected from them. They are also aware that the teachers will be grading them based on their work. This aspect renders performance assessment distinct from other testing methods.

Purpose of assessment

Performance assessments evaluate a student's work and assess how well the student is able to apply the learned concepts. Assessments may be of different types—some may require the student to replicate a procedure learnt in class, some may require simple recall or even intricate analysis.

A limitation of performance assessments is that such assessments may not turn out to be suitable for simply testing the knowledge of facts.

Teaching goals of performance assessment

Performance assessment helps the student to:

- Apply systematic procedures
- Make use of resource texts, laboratory equipment and computers
- Use scientific methodology
- Evaluate and apply several approaches
- Work out difficult problems

The three main purposes of performance assessments are as follows:

1. **Diagnostic purposes:** The assessment is expected to diagnose the students' knowledge about solving specific categories of problems.
2. **Instructional purposes:** A well-crafted assessment can also be used for instructional purposes, except for the fact that it is standardized and graded.
3. **Monitoring purposes:** Since assessments are undertaken at regular intervals, it is very useful for tracking a student's levels of learning, especially in subjects like science and mathematics, which signifies a student's skills of problem solving.

This methodology is revolutionary, in the sense that it is not just about testing, but also about the learning process. In this methodology, students are encouraged to think and create the finished product, during the course of which they are automatically testing their skills and knowledge.

Unlike the old model of teaching where students were tested after each lesson or unit was covered in the classroom, here the approach is to use the book knowledge merely as a resource or means to an end, and not the end itself.

Let us study a relevant example to understand performance assessment better. If a coach assembles 11 players and tells them all the rules of the game of cricket, the spirit in which the game must be played and the dangers of playing without protective padding, will not make them world champions. The players need to go and play an actual game. The performance assessment method involves students 'playing the game'. This way, what goes on in the class is actual preparation for life after school.

Typically, a performance assessment activity can be broken down into the use of the following skills:

1. **Obtaining information:** Students would need to research, compile and organize information as per the specified goal.
2. **Employing the information:** After they have organized the information, they will need to work with it or infer, sequence, analyse and synthesize it to create something new.
3. **Use the information:** When they have the information in the required form, for example, a flowchart, they can use it to convince a target group, or simply present ideas and concepts.
4. **Communication:** Making presentations, oral and visual, doing projects and demonstrating skills is all a part of communicating with others. When a student or a group of students presents their ideas in any form to others, they are automatically honing their communication skills.

Here are the key benefits of employing the performance assessment methodology:

1. **The variety of tasks call for use of various learning styles and skills:** Each student has a core competency, which simply put, means a skill they have more refined than others. Someone may be good at writing descriptive book reports and someone else may be proficient at making Powerpoint presentations. Some learners are exceptional at organizing information and still others may be brilliant speakers, and so on. Not only does performance assessment allow students to display their particular skills, it also encourages them to acquire new ones. For example, a student who likes writing detailed, informational pamphlets should be encouraged to make a persuasive speech to a group of adults describing the pamphlet he or she has made.
2. **Performance assessment tasks encourage teamwork:** Working in teams is a skill that will help a student throughout life, especially if he or she chooses to work in a corporate set up. Therefore, it is a useful skill to acquire. Making group presentations, working on projects in a team, and so on, helps students to learn the value of collaboration at an early age. Division of tasks, cooperation, empathy are all valuable lessons that each student must learn.
3. **There is focus on personal habits and attitudes:** In performance assessment tasks, success depends on the learner's attitude almost as much

NOTES

as book learning. Inculcating habits such as flexibility, diligence, cooperation, tolerance, persistence, problem solving, planning and many more are all inherent to the performance assessment methodology.

It is important to have predetermined criteria to evaluate the students' performance. Students should not be scored or graded against their classmates and should be provided with the criteria before the assessment.

Student's feedback reflects levels of competency, rather than comparative scores. It is always useful to try to find in students' performance patterns of appropriate and inappropriate responses.

1.6.1 Specification of Objective Steps in Evaluation Process

Bloom's taxonomy of educational objectives was revised in 2001 under the title of *A Taxonomy for Learning, Teaching and Assessing*. The revision work was done by authors Lorin. W. Anderson, David R. Krathwohl, Kathleen A. Cruikshank and others. Due to the changes brought in the cognitive domain, the instructional objectives for this domain also need changes accordingly.

The following changes were made in the classification of educational objectives:

The nomenclatures of the categories under cognitive domain were changed.

- Knowledge was replaced by remembering.
- Comprehension was replaced by understanding.
- Application was replaced by applying.
- Analysis was replaced by analysing.
- Synthesis was replaced by creating.
- Evaluation was replaced by evaluating.

Rationale behind the changing concepts

It is evident that the six categories in the original classification of Bloom are in the form of 'noun', which has been changed to 'verb'. The rationale behind this change is simple, logical and easy to digest. Taxonomy of educational objectives, i.e., various categories reflect different forms of thinking. Thinking, being an active process, represents the forms of verbs more accurately than the noun. Therefore, some of the sub categories were reorganized and renamed. Knowledge, the first category, is a product of thinking and was considered inappropriate to describe a category of thinking and, therefore, it was replaced with the word 'remembering', the lowest level of cognitive domain activities.

Knowledge is a product of thinking (lower as well as higher order products) and was not clarifying its level of category; hence change was needed. Remember is the right word because it indicates the product of lower order thinking. We can cram, recall, recognize and remember something without understanding it. Comprehension is replaced by understanding, synthesis is replaced by creating, and so on. All these changes reflect the nature of the thinking process and the outcomes described by each category in a better way.

NOTES

Functions of Educational Objectives/Usefulness of Taxonomical Classification

Educational objectives perform the following functions:

- Bloom's taxonomy motivates educators to focus on all three domains of educational objectives.
- Creates a more holistic form of education to develop the whole personality of the learners.
- Whole personality development makes an individual more useful for his ownself and for the society.
- Allows teachers to select appropriate classroom assessment techniques for student evaluation and for the performance or skills to be evaluated.
- Understanding Bloom's taxonomy develops the competencies in the teachers through which a teacher can design effective learning experiences.
- Helps in devising and organizing fruitful and interesting co-curricular activities.
- Takes care of the all-round development of the personality of the child.
- Helps teachers in framing instructional objectives in behavioural terms.

Guidance Functions of Evaluation

Guidance and counselling of students have become one of the most important areas of school education. Due to diverse kinds of problems like stress and suicide committed by students, this area has become very important for students, parents and educational administration. As a result of these reasons, counsellors are being appointed in schools and several guidance and counselling clinics and centres are coming up at different places to support students and parents. Some of the important guidance functions of evaluation are listed as follows:

- Evaluation results are very much useful for the guidance and counselling of needy students.
- Proper educational and vocational guidance can be given to students on the basis of their evaluation on various abilities, interests, skills, attitudes, aptitude, values, and after knowing the strength and weakness of the needy students.
- On the basis of evaluation results of the test of anxiety and stress among students, proper guidance may be given to students to reduce their level of stress.
- Various types of personality disorders may be diagnosed by using variety of tests and the subject may be given proper guidance to cure or control them properly.
- Students may be helped in choosing the right kind of vocation or profession or career in which they may succeed and grow.
- Students may be helped in choosing suitable courses of study which suit them and in which they may excel.

NOTES

- Based on the findings of evaluation, students may be guided and counselled properly to nurture and develop their life skills.

CHECK YOUR PROGRESS

8. What does assessment determine?
9. What is taxonomy?
10. How did Bloom classify educational objectives?
11. What is problem solving?
12. What is the affective domain of Bloom's taxonomy related to?
13. What is the psychomotor domain concerned with?
14. What does self-concept defines?
15. List the three main purposes of performance assessment.

1.7 SUMMARY

- Measurement refers to the process by which the attributes or dimensions of some physical object are determined.
- Measurement is of two types: (i) physical measurement and (ii) mental measurement/ psychological measurement/educational measurement.
- 'Scaling' in social sciences refers to the process of measuring. Scaling techniques are deployed in the service of estimating different individual levels of extraversion, or level of satisfaction.
- The process of evaluation consists of the following three aspects:
  - Quantitative measurements
  - Qualitative assessment
  - Teachers' opinion
- There are many types of evaluation, but the two important types which are commonly used by teachers are formative evaluation and summative evaluation.
- The formative evaluation is a monitoring type of evaluation which is used to monitor the progress of students during the class, course, or session. It is aimed at improving the quality of teaching-learning process.
- Through placement evaluation, the entry behaviour of the student is assessed.
- Diagnostic evaluation is concerned with the persistent learning difficulties that are left unresolved by the corrective prescriptions of formative assessment.
- The summative evaluation is done at the end of a course, semester or a class or topic.
- The formative evaluation is conducted during the process of teaching and learning, during the class, during the semester or a session. The summative

evaluation, on the other hand, is conducted at the end of the process of teaching and learning, e.g., at the end of class, at the end of semester, at the end of session, etc.

- In external evaluation, the teachers who are teaching a particular group of students are not involved in the evaluation of their students. Hence, the teachers and the examiners are different; they are never common.
- In internal evaluation, the teacher and the examiner both are the same person. The same teacher teaches a particular subject and himself or herself sets the paper and evaluates the achievement of the students.
- The term 'norm' has two meanings. One is the established or approved set of behaviour or conduct to be followed or displayed by all the members of a family, or society or any organization.
- In a norm-referenced evaluation, there is no pass or fail, as there is set marks for passing the test. This type of evaluation is termed as the norm- referenced evaluation and the test is considered as the norm referenced test.
- The study of classification is known as taxonomy.
- Bloom developed and proposed the classification of educational objectives in 1956 which is popularly known as Bloom's Taxonomy of Educational Objectives.
- Bloom classified educational objectives into three 'domains' which are as follows:
  - Cognitive domain
  - Affective domain
  - Psychomotor domain
- Cognitive domain includes those objectives of education which attempt to develop mental faculties or intellectual abilities, that is, the ability of knowing, understanding, thinking and problem solving.
- Bloom has further classified cognitive domain into six major categories which are knowledge, understanding, application, analysis, synthesis and evaluation.
- Affective domain of Bloom's taxonomy of educational objectives is related to the development of emotions, values, attitudes and the development of those aspects of personality which are more influenced by heart than the mind.
- Psychomotor domain deals with the development of motor abilities which in turn are responsible for skill development for performing any task.
- The assessments used in affective assessment are attitudes and values, interest, and self-concept.
- An attitude is a state of mind or a feeling or disposition. It is important to have a positive attitude about work.
- Various methods of measuring attitudes have been developed. Attitude measurement can be divided into two basic categories: direct measurement (A likert scale) and indirect measurement (projective techniques).

- Performance assessments are designed to judge a student's abilities to use specific knowledge and research skills. Performance assessment is a form of testing that requires students to perform a task rather than select an answer from a ready-made list.

1.8 KEY TERMS

- **Rating scale:** It refers to a scale used to evaluate the personal and social conduct of a learner.
- **Diagnostic evaluation:** It aims at identifying or diagnosing the weaknesses of students in a given course of instruction.
- **Anecdotal records:** Those records which maintain the description of significant event and work or performance of the students
- **Grading:** It is a means for reporting the result of measurement.
- **Formative evaluation:** It is the type of evaluation which is done during the teaching-learning process to assess the ongoing construction or formation of knowledge and understanding of students.
- **Summative evaluation:** It is a type of evaluation that evaluates the quality of the final product and finds out the extent to which the instructional objectives have been achieved.
- **Cognitive domain:** It includes those objectives of education which attempt to develop mental faculties or intellectual abilities, that is, the ability of knowing, understanding, thinking and problem solving.
- **Affective domain:** It is related to the development of emotions, values, attitudes and the development of those aspects of personality which are more influenced by heart than the mind.
- **Psychomotor domain:** It deals with the development of motor abilities which in turn are responsible for skill development for performing any task.
- **Likert scale:** It is a tool used to determine opinions or attitudes; it contains a list of declarative statements, each followed by a scale on which the subject is to indicate degrees of intensity of a given feeling.
- **Self-concept:** It is the composite of ideas, feelings and attitudes that a person has about his or her own identity, worth, capabilities and limitations.

1.9 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Measurement refers to the process by which the attributes or dimensions of some physical object are determined.
2. Objectivity of a test refers to two aspects: (a) item objectivity (i.e., objectivity of the items), and (b) scoring objectivity (i.e., objectivity of scoring).

3. J.M. Bradfield defines evaluation as 'the assignment of symbols to phenomenon in order to characterize the worth or value of the phenomenon usually with reference to some social, cultural and scientific standards'.
4. Scaling in social sciences refers to the process of measuring.
5. A test consists of a set of questions to be answered or tasks to be performed.
6. The goals of placement assessment are to determine for each student the position in the instructional sequence and the mode of instruction that is most beneficial.
7. Diagnostic evaluation aims at identifying or diagnosing the weaknesses of students in a given course of instruction.
8. Assessment is the process of determining the following:
   (i) he extent to which an objective is achieved.
   (ii) The effectiveness of the learning experiences provided in the classroom.
   (iii) How well the goals of teaching have been accomplished.
9. The study of classification is known as taxonomy.
10. Bloom classified educational objectives into three 'domains', namely: (i) cognitive domain, (ii) affective domain, and (iii) affective domain.
11. Problem solving is a mental process that involves discovering, analysing and solving problems.
12. The affective domain of Bloom's taxonomy of educational objectives is related to the development of emotions, values, attitudes, and the development of those aspects of personality which are more influenced by the heart than the mind.
13. The psychomotor domain is concerned with those objectives which are intended to develop various skills.
14. Self-concept defines how we assess ourselves as individuals or what we think of ourselves.
15. The three main purposes of performance assessment are:
    - Diagnostic purposes
    - Instructional purposes
    - Monitoring purposes

1.10 QUESTIONS AND EXERCISES

Short-Answer Questions

1. Discuss the two types of measurement.
2. Differentiate between formative and summative evaluation.
3. What are aptitude tests?
4. How is cognitive learning defined?

NOTES

5. What are the thinking skills required in cognitive learning?
6. What are the steps in the problem solving process?
7. What are the features of external evaluation?
8. What was the rationale behind changing the six categories of Bloom's educational objectives?
9. State few ways which can ensure that one makes good decisions.
10. What are the factors that affect self-concept?

Long-Answer Questions

1. Discuss the characteristics of a good measurement tool.
2. Discuss the importance of measurement scales. You may give your argument with reference to various measurement scales.
3. Describe qualitative evaluation techniques, as well as the five categories that come under it.
4. How are the objectives of the cognitive domain classified? Discuss in detail.
5. Describe various types of evaluation with proper examples.
6. Describe the techniques that help to stimulate divergent thinking.
7. Explain the classification of psychomotor domain under educational objectives.
8. What are the characteristics of critical thinking?

1.11 FURTHER READING

Bloom, B. S. *Taxonomy of Objectives of Cognitive Domains*. New York: Longmans Green &Co.

Ebel, R.L. *Essentials of Educational Measurement.* New Jersey: Englewood Cliff.

Grunlund, N. E. *Measurement and Evaluation in Teaching* (3rd Ed). New York: Mcmillan Publishing Co.

Julian, C., Stanley and Hopkins. *Educational and Psychological Measurement and Evaluation*. New Delhi: Prentice Hall of India Pvt. Ltd.

Lindermon, Richard, H. *Educational Measurement*. Mumbai: B. S. Taraporevala Sons & Co.

UNIT 2 MAJOR TOOLS AND TECHNIQUES IN EDUCATIONAL EVALUATION

Structure

2.0 INTRODUCTION

In the previous unit, you learnt about the concept of measurement and evaluation. In it, difference types of measuring scales and the need for measurement and evaluation were discussed. In this unit, the discussion will turn towards the major tools and techniques in educational evaluation. Evaluation is an attempt to appraise the quality/ suitability of a resource. There are different types of evaluation techniques. Some are standardized tests, while others are made by the teachers. Moreover, while some tests are norm-referenced, other tests are criterion referenced. This unit will look at all of these in detail.

2.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Discuss how tests can be classified
- Describe the merits and demerits of different types of tests.
- Explain how a criterion-referenced test can be constructed
- Discuss the difference types of questionnaires that are used by researchers

2.2 DIFFERENT TYPES OF TESTS: TEACHER-MADE VS. STANDARDIZED

NOTES

Tools or question papers prepared by the teachers to evaluate their own students whom they have been teaching are called as teacher made tests. These tests are not standardized and are prepared for a small number of students, generally for a section or a class or a school.

Standardized tests or tool

A standardized test is a test, which is developed in such a way that it reaches up to a specific level of quality or standardization. The test is standardized with respect to the form and construction, administration procedure and test norms. It is a kind of test, which is standardized in terms of its development, administration, scoring and interpretation. The test is standardized to bring objectivity, reliability, validity and to have all the other characteristics of a good test. Standardized tests have a manual with it, which instructs and guides its users regarding its administration, scoring and interpretation. The following are some of the important definitions of standardized tests to clarify the concept of standardized test or tool of measurement and evaluation.

According to C. V. Good, 'a standardized test is that for which content has been selected and checked empirically, for which norms have been established, for which uniform methods of administration and scoring have been developed and which may be scored with a relatively high degree of objectivity.'

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

The most important benefits of a standardized test are that it minimizes or reduces four types of errors. These are personal error, variable error, constant error and interpretive error.

Characteristics of standardized tests

Following are the important characteristics of a standardized test:

- It has norms, which contains everything about the test, starting from its preparation to its scoring and interpretation. Norms of the test describe every aspect of the test in detail so that any user can use it properly by using the norms of the test.
- It has norms developed for transferring raw score to a standard score.
- Instruction for administration of the test is pre-determined and fixed.
- Duration of the test is fixed.
- The test is standardized on a suitable sample size selected from a well-defined population.
- It has high reliability and validity.
- The test has high objectivity.

- Answer key and scoring procedure of the test is fixed and is predetermined.
- Test manual is properly developed.
- The errors in the standardized tests are minimized or reduced.

NOTES

2.2.1 Criterion-referenced vs. Norm-referenced Tests

Norm-referenced Tests

To understand non-referenced type of evaluation, we have to first learn about the term 'norm'. The term 'norm' has two meanings. One is the established or approved set of behavriour or conduct to be followed or displayed by all members of a family or society or any organization**.** It is the established customs of the society which most of the people follow without question. The other meaning of the term, which is meaningful for us here, is the average performance of the group.

Example: A group of students are tested for the awareness towards environmental pollution through a written test. The test consists of 50 objective type questions of one mark each. There was no negative marking in the test. The full mark of the test is obviously 50. After the conduction of the test, it is marked by the examiner. There are 150 students in the group. Marks of all students are added and the additive marks are divided by 150 to find the average performance of the group. Suppose it is found to be 30, so 30 marks is the average obtained by the whole group in which some achieve 49 out of 50 and some other achieve very less, say 12 out of 50. The 30 mark, i.e., the average of the group is said to be the norm of this group.

Now, the evaluation of all 150 students is done considering this 30 (the norm) as a point of reference. All students who have got marks above 30 are considered as above average, all those who have achieved below 30 are considered as below average and all those who have achieved just 30 are supposed as average. There is no pass or fail in this type of evaluation as there is set marks for passing the test. This type of evaluation is called as norm-referenced evaluation and the test is considered as norm-referenced test.

Criterion-referenced evaluation

The type of evaluation in which the performance of the testees are evaluated with reference to some predetermined criteria is called as criterion-referenced evaluation. No weightage is given to norm or average performance of the group in this evaluation. All decisions, such as pass or fail, distinction, excellent, etc., are taken with reference to criteria set out in advance.

In the preceding example, if some criteria is set before the test with reference to which the performance of each students will be evaluated, it will become criterion-referenced devaluation. Suppose the following criteria are finalized for this test:

Pass marks: 40 %

Distinction: 80%

In the test discussed above, all those students who get 20 or more than 20 (40%) are declared as pass. All those students who score less than 20 are declared

NOTES

as fail. All those who get 40 or more (80%) are declared as distinction. If any prize is given to those who score at least 90%, then only the students who will get 45 or more will get the prize. As all decisions are taken on the basis of some criteria, this evaluation is called as criterion-referenced evaluation. Let us now look at how a creterion-referenced evaluation can be constructed.

Construction of criterion referenced test

During the preparation of a criterion-referenced test, the test constructor is required to take the following steps:

Step 1: Indentifying the purpose of the test

First of all, the objectives of the test are finalized. The test developer must know the purpose of the test. He should be well informed and well aware of the purpose of the test for which it is going to be prepared. In addition to the purpose, he should also know the following aspects of the test:

- Content areas of the test from where the items will be developed
- Level of students or examinees for whom test is being prepared
- Difficulty level of the test items
- Types of the test objective type or subjective type or mixed type of test
- Criteria for qualifying the test

After having understood these points, the test developer starts the work of constructing the criterion-referenced test. He moves on to the second step.

Step 2: Planning the test

This step requires the following work to be done by the test constructor:

(i) Content analysis: The test developer analyses the content of the test. It involves the selection of contents, i.e., the testing areas and its peripherals. He also decides the key areas of the content from where more questions are to be developed.

(ii) Types of items: The decisions regarding the type of items are taken at this stage. In case of subjective type, it may be essay type, short-answer type and very short-answer type. In case of objective type, it may be multiple- choice question, fill in the blanks, true or false type, sentence completion type, one word answer type, etc., If the test is of mixed type, then questions are developed accordingly. But what is planned at this stage is the proportion of objective and subjective type items in terms of marks.

(iii) No of items: Total number of questions of each type which is to be included in the test is decided.

(iv) Weightage: It is very important to decide the weightage of each type of items and each content area. It depends upon the level of the student being tested. As we move from lower to higher level, the percentage of knowledge domain items decreases and higher order thinking abilities, such as

understanding, application and skill, increases. The test developer also decides the weightage of each of the content areas being included in the test considering its relevance.

(v) **Duration of the test:** With consideration to the total number of questions in the test, the level of examinees, difficulty levels of the test items and the duration of the test are decided.

(vi) **Mechanical aspects:** It includes the quality of paper, ink, diagrams, type setting, font size and printing of the test papers.

(vii) **Development of key for objective scoring:** To bring objectivity in the process of evaluation, it becomes essential to achieve interpersonal agreement among the examiners with regard to the meaning of the test items and their scoring. For this purpose, a 'key' is prepared for each paper and given to all examiners while scoring the test. They are supposed to score the test following the key.

(viii) **Instructions for the test:** The test developer also prepares instructions for its administration, and scoring and evaluation procedure 'test manual'. It shows the whole procedure of testing. It acts as a guide to the individuals involved in testing procedure at all stages. This manual is strictly applied to bring objectivity in the test.

Step 3: Preparing blueprint of the test

Blueprint is a specification chart which shows the details of the test items to be prepared. It shows all the content areas and the number and type of questions from those areas. It also reflects the objectives to be tested. The blueprint describes the weightage given to various content areas, objectives, types of items and all other details of the test. It serves as a guideline or frame of reference for the person constructing the test.

Step 4. Construction of test items

According to the blueprint all questions are constructed, covering all the content areas, all the objectives or abilities and all types of items. Questions may be objective or subjective type as mentioned in the blueprint.

Examples of objective-based and objective-type questions:

1. Multiple choice questions

(i) Which of the following metal is present in blood?
(a) Cobalt (b) Iron (c) Calcium (d) Sulpher

(ii) Which of the following is not associated with photosynthesis?
(a) Dark reaction (b) Light reaction (c) Calcium (d) Chlorophyll

(iii) Which of the following is a port city?
(A) Patna (B) Ahmadabad (C) Calcutta (D) Durgapur

NOTES

(iv) The capital of India was transferred from Calcutta to Delhi in the year:

(A) 1911 (B)1905

(C) 1912 (D) 1906

2. True false type questions

State true or false for the following:

(i) Pacemaker is related to lungs.

(ii) Evaluation is a quantitative process.

(iii) Man is mortal.

(iv) Privatization and globalization are the same concept.

3. Matching type items

Match the items in column A to those in column B in the following table.

Column A	Column B
Ganga	Bhakra
Gomti	Patna
Yamuna	Lucknow
Sutlej	Delhi

4. Fill in the blanks

Fill in the blanks selecting the appropriate words from the brackets against each of the following.

(i) The longest cell of human body is …….(Neuron, Tibia fibula)

(ii) Largest continent is ……..(Asia, Africa)

(iii) The deepest ocean is …….(Atlantic, Pacific)

(iv) India became republic in the year ……..(1947, 1950)

(v) The 20th common wealth game will be held in …… city (Glasgow, Mexico)

5. One word-answer Questions

Give one word answer for the following questions:

(i) India's biggest bird sanctuary is situated at which place?

(ii) How many chambers are there in human heart?

(iii) Write the full form of GATT.

(iv) How many countries participated in the 19th Commonwealth Games held in Delhi?

(v) Name the discoverer of America.

NOTES

The questions prepared in the box are objective in nature as all of them have only one correct answer. The students will get either full marks (for correct response) or no marks (for wrong response). The experience, feelings and other academic, personal or social factors of the examiners or scorers will not influence the marking of these types of questions. We can say that, no subjectivity is possible in scoring these types of tests. Hence, these are called objective type and objective-based tests.

Step 5: Selecting the items for the test

Items have already been constructed as per the guideline of the blue print. It may be that some items are not suitable or up to the mark. To avoid this fact, generally some more questions are prepared so that if any question is rejected at any stage, it might be able to be replaced immediately. The following steps are followed to select the right item for the test. This process of selection is done through a process known as 'try out'. The process of try out involves the following steps:

(i) Sampling of subjects: As per the size of population for which the test is being prepared, a workable sample, say around 150 subjects, is selected on a random basis. This is the sample on which the prepared items are tested for its functionality, workability and effectiveness.

(ii) Pre-try out: It is also known as preliminary try-out. The prepared items are administered on a sample of around ten subjects taken form the sample. The answersheets are checked, evaluated and discussed with the candidates for any kind of problems they would have faced during the test. It is probable that they might have faced the language difficulty, words ambiguity and some other problems of this kind. These problems are sorted out. The items having these problems are rewritten or rephrased to improve and modify the language difficulties and ambiguity of the items. At the end of the pre try- out, the initial draft of the test is prepared.

(iii) Proper try-out: At this stage, around fifty candidates are selected from the sample and the initial draft of the test is administered on them. Answersheets are scored and item analysis is done. Difficulty value and discrimination power of each item are calculated. The items which come within the acceptable range of difficulty value and discrimination power are selected for the test and others are rejected.

(iv) Final try-out: The final try-out is done on a comparatively large sample. The sample size may be more than 100 or even more, depending upon the size of population. After administration and scoring of the test, reliability and validity of the test are measured. If it is proved to be reliable and valid, it gets green signal.

Step 6. Evaluating the test and preparing final draft of the paper

For establishing quality, an index test manual is prepared which informs about the test's norms, scoring key, reliability and validity. The final draft of the test paper is prepared. Instructions for the examinees as well as for the test administration are

determined. Item analysis is performed to find out the item workability for the test. The required changes are done and final draft of the paper is ready for printing.

NOTES

2.2.2 Essential Qualities of Good Measuring Instruments

The following are the characteristics of a good test:

- **Objectivity:** An evaluation tool should be objective in nature. The test should fetch almost the same score to an individual, irrespective of the examiner who scores it. It should be free from all kinds of biases at all levels of testing.
- **Reliability:** It is the consistency of scores obtained by an individual on different times on the same test or parallel test, scored by same or different examiner at same or different times. The difference in scores (if any) should be insignificant.
- **Validity:** An evaluation tool should measure what it is supposed to measure. If a tool is developed or prepared to measure a particular aspect of personality, it should measure that aspect only and nothing else. If the test does so, it is supposed to be valid.
- **Practicability**: An evaluation tool should have practicability or usefulness in terms of time, energy and resources. It should also be easy to administer.

These shall be discussed in detail in the next section.

CHECK YOUR PROGRESS

1. What are teacher-made tests?
2. Define criterion-referenced evaluation.

2.3 EDUCATION TESTS

An educational test is not just a test 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 test is essentially an objective and standardized 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 samples of verbal or non-verbal responses, or by means of other behaviours'.

Test is a stimulus selected and organized to elicit responses which can reveal certain psychological traits in the person who deals with them. The diagnostic 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 a psychological test is the quantitative and qualitative measurement of the various aspects of behaviour of the individual for making generalized statements about his total performances.

NOTES

The aspects which affect the characteristics of a good test are as follows:

- Validity of the test
- Reliability of the test
- Objectivity of the test
- Usability of the test
- Comprehensive and preciseness of the test
- Administration of the test
- Test from economic viewpoint
- Availability of the test
- Appearance of the test
- Standardization of the test
- Norms of the test

Some of the important characteristics of a test are analysed below.

Validity: Validity of a test refers to its truthfulness; it refers to the extent to which a test measures what it intends to measure. Standardization of a test requires the important characteristic viz., validity. If the objectives of a test are fulfilled, we can say that the test is a valid one. The validity of a test is determined by measuring the extent to which it matches with a given criterion. Let us take an example, suppose we want to know whether an 'achievement test in mathematics' is valid. If it really measures the achievement of students in mathematics, the test is said to be valid, or else not. So 'validity' refers to the very important purpose of a test and hence it is the most important characteristic of a good test. A test may have other merits, but if it lacks validity, it is valueless.

Freeman states, 'an index of validity shows the degree to which a test measures what it is supposed to measure when compared with the accepted criteria'. Lee J. Cronback held the view that validity 'is the extent to which a test measures what it purports to measure'.

Reliability: Reliability refers to consistency of scores obtained by some individuals when re-tested with the test on different sets of equivalent items or under other variable examining conditions. It refers to the consistency of scores obtained by the same individuals when they are re-examined with the same test on different occasions or with different sets of equivalent items or under different examining conditions. Reliability paves the way for consistency that makes validity possible and identifies the degree to which various kinds of generalizations are justifiable. It refers to the consistency of measurement i.e., how stable test scores or other assessment results are from one measurement to another.

Reliability refers to the extent to which a measuring device yields consistent results upon testing and retesting. If a measuring device measures consistently, it is reliable. The reliability of a test refers to the degree to which the test result obtained is free from error of measurement or chance errors. For instance, we administer an achievement test in mathematics for students of class IX. In this test, Paresh scores

NOTES

52. After a few days, we administer the same test. If Paresh scores 52 marks again, we consider the test to be reliable, because we feel that this test accurately measures Paresh's ability in mathematics. H.E. Garrett stated, 'the reliability of test or any measuring instrument depends upon the consistency with which it gauges the ability to whom it is applied'. The reliability of a test can also be defined as 'the correlation between two or more sets of scores on equivalent tests from the same group of individuals'.

Objectivity: Objectivity is an important characteristic of a good test. Without objectivity, the reliability and validity of a test is a matter of question. It is a pre-requisite for both validity and reliability. Objectivity of a test indicates two things—item objectivity and scoring objectivity.

'Item objectivity' refers to the item that must call for a definite single answer. In an objective-type question, a definite answer is expected from the test-takers. While framing the questions, some points to be kept in mind are: ambiguous questions, lack of proper direction, double barrelled questions, questions with double negatives, etc. These concepts affect the objectivity of a test. Let us take an example of an objective item. Suppose we ask students to write about Gandhi. This question does not have objectivity. Because, here the answers will have different perceptions for different individuals and also the evaluation. If we ask the students 'what was Gandhi's father's name', this obviously will have only one answer and even the biasness of the evaluator will not affect the scoring. So all the items of a test should be objective.

Objectivity of scoring refers to by whosoever checked the test paper would fetch the same score. It refers to that the subjectivity or personal judgment or biasness of the scorer should not affect the scores. The essay-type questions are subjective and the scores are affected by a number of factors like mood of the examiner, his language, his biasness, etc. Essay-type questions can have objectivity if the scoring key and proper directions for scoring are provided.

Usability: Usability of a test refers to the practicability of a test. It refers to the degree to which the test can be successfully used by the teachers/evaluators. Usability of a test depends on certain aspects which are expressed in the following manner:

(a) *Comprehensibility:* The test items should be free from ambiguity and the direction to the test items and other directions to the test must be clear and understandable. The directions for scoring and the interpretation of scores must be within the comprehension of the user.

(b) *Ease of administration:* If the directions for administration are complicated or if they need more time and labour, the users may lag behind to use such tests. The directions for administration must be clear and concise. The test paper should be constructed according to the availability of time. Lengthy tests involving more time may not be preferred for use.

(c) *Availability:* If a test is not available at the time of necessity, it lacks its usability. Most of the standardized tests are of high validity and reliability,

NOTES

but their availability is very less. So it is desirable that in order to be reliable, the tests must be readily and easily available.

(d) *Cost of the test:* The cost of the test must be cheap, so that the schools and teachers can afford to purchase and use them. If it will be costly, then every school cannot avail it. So a good test should be of reasonable price.

(e) *Ease of interpretation:* A test is considered to be good if the test scores obtained can be easily interpreted. For this, the test manual should provide age norms, grade norms, percentile norms and standard score norms like standard scores, T-scores, Z-scores etc. So 'interpretability' of test refers to how readily the raw scores of test can be derived and understood.

(f) *Ease of scoring:* A test in order to be usable must ensure ease of scoring. The scoring procedure must be a simple one.

All the directions for scoring and the scoring key should be available, to make the scoring an objective one. The examiner's biasness, the handwriting of the examinee should not affect the scoring of a test.

Classification of Tests

Tests are divided into different types taking into consideration their content, objective, administration system, scoring style etc. According to mode of administration, tests are of two types:

(i) *Individual test:* When a psychological test is administered upon an individual at a particular time, it is known as 'individual test'.

(ii) *Group test:* When a test is administered upon a group of individuals at a particular time, it is known as 'group test'. It is mostly applicable on adult literates.

According to the ability of the student, tests are of two types:

(i) *Speed test:* This type of test is applicable upon the individuals to know the mental speediness. Here, the time is limited and the number of questions is more and all the questions are equal in difficulty level. Railway examinations, banking examinations are the examples of speed test.

(ii) *Power test:* This type of test is applicable upon the individuals to know the mental power or the ability. Here, time limit is not there and the individuals are expected to answer the question within as much time they like. All the questions of this test are arranged according to difficulty level and discriminating power. The essay competitions by the media are the bright examples of power test.

According to the type of items involved in the test, it can be of three types:

(i) *Essay-type test:* Essay-type tests are otherwise known as open-ended tests. The essay question is especially useful for measuring those aspects of complex achievement that cannot be measured well by more objective means. These include: (a) the ability to supply rather than merely identify interpretations and

NOTES

application of data, and (b) the ability to organize, integrate and express ideas in a general attack on a problem. Outcomes of the first type are measured by restricted-response questions and outcomes of the second type by extended-response questions. For example, 'discuss the educational philosophy of M.K. Gandhi'.

(ii) *Short-answer type test:* This type of test requires to be written in a short-cut manner regarding a concept. It is suitable for measuring a wide variety of relatively simple learning outcomes, and it is used almost exclusively to measure the recall of memorized information. *For example:* 'What is measurement? Write within 50 words.'

(iii) *Objective-type test:* In objective-type questions, the individual is expected to answer the question with the help of a word, a phrase, a number or a symbol. The test with multiple-choice items, true-false items, matching type items, fill-in-the blanks items, one-word substitution are the examples of objective type test.

According to the method of scoring, test are of two types:

(i) *Machine-scored test:* The tests which are scored or assessed by the machines like computer are known as 'machine-scored test'. The Bank P.O. examination is an example of machine-scored test.

(ii) *Hand-scored test:* The tests which are assessed by the human beings are known as 'hand-scored tests'. The classroom achievement tests is an example of hand-scored tests.

According to the principle of test construction, tests are of two types:

(i) *Teacher-made test:* Generally 'teacher-made tests' are prepared by classroom teachers to assess pupils' growth. It is related to action research. Teacher-made tests serve different purposes viz., to measure pupil's achievement, to know how far the specific objectives have been fulfilled, to diagnose the learning difficulties, and to arrange specific remedial measures to award grades etc. This type of test only follows two steps—planning and preparation.

(ii) *Standardized test:* Standardized tests measure the common objectives of a wide variety of schools. They have standard procedures for administration and scoring, and provide norms for interpreting the scores. A test manual and other necessary material are typically provided to aid in the administration of the test and the interpretation and use of the results. The test items are generally of high quality because they have been prepared by specialists, subject experts, pre-tested and selected on the basis of their effectiveness and their relevance to a rigid set of specification. They are specially useful for measuring general educational development, determining student's progress from one year to the next, grouping students, analysing learning difficulties, and comparing achievement with learning ability.

Standardized test and teacher-made test have been discussed later in this unit in detail.

NOTES

According to the nature of the test, they are classsified as:

(i) *Oral test:* It is a kind of verbal test. In oral test, the individual is expected to answer orally. This type of tests is mostly applicable to illiterates or small children. In the public survey, the people are asked to speak something regarding the issue. In the interview board, the interviewers ask questions to the interviewee and the interviewee answers orally. This type of test is known as 'oral test'.

(ii) *Written test:* Here, the individual has to respond the questions in writing form. So the respondent should have the writing ability. It is only applicable upon the literates. All the written examinations are the examples of written test. It is a kind of verbal test.

(iii) *Performance test:* This type of test is also known as 'non-verbal test'. The respondent is not expected to respond verbally. He has to perform the task. The running competition, jumping competition held by physical examination are the examples of performance test.

Check Your Progress

3. What is validity of a test?
4. What does the objectivity of a test indicate?

2.4 MEASUREMENT OF ACHIEVEMENT

An achievement test is an instrument designed to measure the relative achievement of students. It is an indispensable instrument in the teaching–learning process. The following definitions given by experts enable us to have a comprehensive view of an achievement test. According to Gronland, achievement test is 'a systematic procedure for determining how much a student has learned through instruction'. According to Popham, 'the achievement test focuses upon an examinee's attainments at a given point in time'.

This most common technique for measuring cognitive development is supposed to yield the intended evidences about students' learning and indirectly the instructional effectiveness. Since most of the judgments about the students' achievement and teaching effectiveness are based on the evidences gathered, it is necessary that the question papers are designed well and prepared scientifically. For this, it is necessary that teachers in general and paper-setters in particular are thoroughly made conversant with the following concepts and steps for developing a good question paper (achievement test):

(i) Knowledge of criteria of a good achievement test

(ii) Preparation of design of question paper

(iii) Development of blueprint based on the design provided or developed

(iv) Framing of different types of questions in accordance with the blueprint

NOTES

(v) Development of key, model answers and marking schemes

(vi) Consolidation of questions and formatting of question papers

(vii) Preparation of question-wise analysis

(viii) Final editing and review of the question paper

(ix) Moderating of the question paper

Practical uses of achievement tests

Achievement tests can be used for various purposes in the teaching–learning process. The value of such tests is directly proportional to the extent to which the results from its use are translated into improved instructional, guidance and administrative practices in the school. The instructional uses of achievement tests are:

- For class analysis and diagnosis
- For individual pupil diagnosis
- Guidance use of achievement tests
- Alternative uses
- Administrative uses
- Aid to the parents

We know that the three basic criteria of a good question paper are: (i) validity, (ii) reliability, and (iii) usability or practicability that makes a question paper good. Evidences collected are to be judged against these criteria, which the paper-setter must understand.

2.4.1 Construction of Achievement Tests and Standardization

A questionnaire is a tool for research, comprising a list of questions whose answers provide information about the target group, individual or event. Although they are often designed for statistical analysis of the responses, this is not always the case. This method was the invention of Sir Francis Galton. Questionnaire is used when factual information is desired. When opinion rather than facts are desired, an opinionative or attitude scale is used. Of course, these two purposes can be combined into one form that is usually referred to as 'questionnaire'.

Questionnaire may be regarded as a form of interview on paper. The procedure for the construction of a questionnaire follows a pattern similar to that of the interview schedule. However, because the questionnaire is impersonal, it is all the more important to take care of its construction.

A questionnaire is a list of questions arranged in a specific way or randomly, generally in print or typed and having spaces for recording answers to the questions. It is a form which is prepared and distributed for the purpose of securing responses. Thus a questionnaire relies heavily on the validity of the verbal reports.

According to Goode and Hatt, 'in general, the word questionnaire refers to a device for securing answers to questions by using a form which the respondent fills himself'.

NOTES

Barr, Davis and Johnson define questionnaire as, 'a questionnaire is a systematic compilation of questions that are submitted to a sampling of population from which information is desired' and Lundberg says, 'fundamentally, questionnaire is a set of stimuli to which literate people are exposed in order to observe their verbal behaviour under these stimuli'.

Types of Questionnaire

Figure 2.1 depicts the types of questionnaires that are used by researchers.

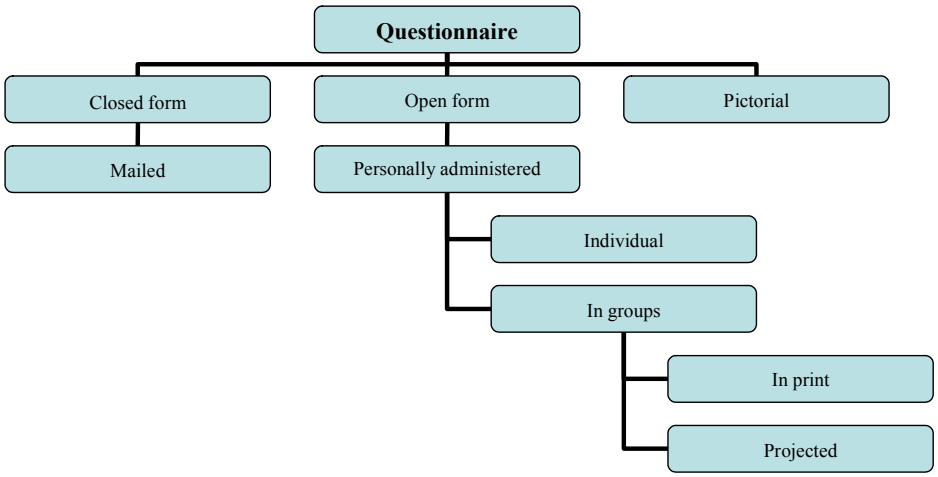


Fig. 2.1 *Types of Questionnaires*

Commonly used questionnaires are:

(i) Closed form: Questionnaire that calls for short, check-mark responses are known as closed-form type or restricted type. They have highly structured answers like mark a yes or no, write a short response or check an item from a list of suggested responses. For certain type of information, the closed form questionnaire is entirely satisfactory. It is easy to fill out, takes little time, keeps the respondent on the subject, is relatively objective and is fairly easy to tabulate and analyse.

For example, How did you obtain your Bachelors' degree? (Put a tick mark against your answer)

(a) As a regular student

(b) As a private student

(c) By distance mode

These types of questionnaires are very suitable for research purposes. It is easy to fill out, less time consuming for the respondents, relatively objective and fairly more convenient for tabulation and analysis. However, construction of such type of questionnaire requires a lot of labour and thought. It is generally lengthy as all possible alternative answers are given under each question.

(ii) Open form: The open form, or unrestricted questionnaire, requires the respondent to answer the question in their own words. The responses

NOTES

have greater depth as the respondents have to give reasons for their choices. The drawback of this type of questionnaire is that not many people take the time to fill these out as they are more time consuming and require more effort, and it is also more difficult to analyse the information obtained.

Example: Why did you choose to obtain your graduation degree through correspondence?

No alternative or plausible answers are provided. The open form questionnaire is good for depth studies and gives freedom to the respondents to answer the questions without any restriction.

Limitations of open questionnaire are as follows:

- They are difficult to fill out.
- The respondents may never be aware of all the possible answers.
- They take longer to fill.
- Their returns are often few.
- The information is too unwieldy and unstructured and hence difficult to analyse, tabulate and interpret.

Some investigators combine the approaches and the questionnaires carry both the closed and open form items. In the close ended questions, the last alternative is kept open for the respondents to provide their optimum response. For example, 'Why did you prefer to join B.Ed. programme? (i) Interest in teaching (ii) Parents' wish (iii) For securing a government job (iv) Other friends opted for this (v) Any other.'

(iii) Pictorial form: Pictorial questionnaires contain drawings, photographs or other such material rather than written statements and the respondents are to choose answers in terms of the pictorial material. Instructions or directions can be given orally. This form is useful for working with illiterate persons, young children and persons who do not know a specific language. It keeps up the interest of the respondent and decreases subjects' resistance to answer.

Questionnaire administration modes

Main modes of questionnaire administration are:

- **Through Mail:** Mailed questionnaires are the most widely used and also perhaps the most criticized tool of research. They have been referred to as a 'lazy person's way of gaining information'. The mailed questionnaire has a written and signed request as a covering letter and is accompanied by a self-addressed, written and stamped envelope for the return by post. The method of mailing out the questionnaire is less expensive in terms of time, funds required; it provides freedom to the respondent to work at his own convenience and enables coverage of a large population.
- **Personal contact/face-to-face:** Personally administered questionnaires both in individual and group situations are also helpful in some cases and have the following advantages over the mailed questionnaire (i) the investigator can establish a rapport with the respondents; (ii) the purpose of the questionnaire

can be explained; (iii) the meaning of the difficult terms and items can be explained to the respondents; (iv) group administration when the respondents are available at one place is more economical in time and expense; (v) the proportion of non-response is cut down to almost zero; and (vi) the proportion of usable responses becomes larger. However, it is more difficult to obtain respondents in groups and may involve administrative permission which may not be forthcoming.

- **Computerized questionnaire:** It is the one where the questions need to be answered on the computer.
- **Adaptive computerized questionnaire:** It is the one presented on the computer where the next questions are adjusted automatically according to the responses given as the computer is able to gauge the respondent's ability or traits.

2.4.2 Relative Merits and Demerits of Different Test Items

Different types of test include multiple-choice questions, short-answer questions, essay –type questions, true-false questions, and so on. Let us look at their merits and demerits.

(a) Multiple-choice questions

Merits

The merits of multiple-choice questions are as follows:

- Quick and easy to score, by hand or electronically
- Can be written so that they test a wide range of higher-order thinking skills
- Can cover lots of content areas on a single exam and still be answered in a class period

Demerits

The demerits of multiple-choice questions are as follows:

- Often test literacy skills: "if the student reads the question carefully, the answer is easy to recognize even if the student knows little about the subject" (p. 194)
- Provide unprepared students the opportunity to guess, and with guesses that are right, they get credit for things they don't know
- Expose students to misinformation that can influence subsequent thinking about the content
- Take time and skill to construct (especially good questions)

(b) True-false questions

Merits

The merits of truth-false questions are as follows:

- Quick and easy to score

NOTES

Demerits

The demerits of truth-false questions are as follows:

- Considered to be one of the most undependable forms of evaluation
- Often written so that most of the statement is true save one small, often trivial bit of information that then makes the whole statement untrue
- Encourages guess work, and rewards correct guesses

(c) Short-answer questions

Merits

The merits of short-answer questions are as follows:

- Quick and easy to grade
- Quick and easy to write

Demerits

The demerits of short-answer questions are as follows:

- Encourage students to memorize terms and details, so that their understanding of the content remains superficial

(d) Essay questions

Merits

The merits of essay questions are as follows:

- Offer students an opportunity to demonstrate knowledge, skills, and abilities in a variety of ways
- Can be used to develop student writing skills, particularly the ability to formulate arguments supported with reasoning and evidence

Demerits

The demerits of essay questions are as follows:

- Require extensive time to grade
- Encourage use of subjective criteria when assessing answers
- If used in class, necessitate quick composition without time for planning or revision, which can result in poor-quality writing.

CHECK YOUR PROGRESS

5. How does Popham define achievement test?
6. What are the instructional uses of achievement tests?

2.5 DIAGNOSTIC TEST CONSTRUCTION AND USEFULNESS

NOTES

Diagnostic tests are a kind of educational tests. These tests have two purposes: (i) prognostic purpose, and (ii) diagnostic purpose. The prognosis of the students done in the specific subjects which has been taught to them. The diagnosis means to identify the causes of their weakness of their poor attainment of the students. The prognosis and diagnostic functions are complementary to each other, and both are essential in educational measurement and evaluation.

A diagnostic test is designed to real specific weakness or failures to learn in some subject of study such as reading or arithmetic. In a diagnostic test, the main interest is the performance on individual items or on small groups of highly similar items. In the diagnostic test, score or mark is not assigned for the correct answer but wrong answer provide the basis for the causes of his failure.

Diagnostic tests are those which help us to know the particular strength and weakness of the student. These tests are also known as 'analytical tests'. The correct answer provides a student's strength and wrong answer indicates his weakness. The achievement tests provide the overall scores on the basis of correct answers of the subject items. The wrong answers are assigned zero marks. These attainment tests do not provide reason for the poor scores of an individual.

The term 'diagnostic' as applied to tests is fraught with danger and ambiguity. Educationists consider the certain test as diagnostic, while others are achievement tests which have no diagnostic characteristics. A diagnostic test undertakes to provide a picture of strengths and weaknesses. Hence, any test that yields more than a single overall score is, in a sense, diagnostic. Even if there are only two part scores, for example, one for arithmetic computation and one for arithmetic reasoning, diagnostic tests makes it possible to say that the student performed better in computation than in reasoning answers to problems. It means that the diagnostic tests are qualitative, not quantitative. A diagnostic test does not yield the total scores of an individual in a subject which he has studied and taken the test.

Analysis as the basis of diagnosis

The successful development of school learning depends upon the care with which the underlying and basic skills of the subjects themselves are recognized and utilized in teaching. For instance, teaching a child how to do additions comprise not only in developing the habit of responding automatically and correctly to the basic combinations, but also involves higher levels of skill, such as control of the attention span, and carrying from one column to the next. The teacher's task is made obvious and objective if he understands this. Similarly, it can be shown that silent reading comprehension is not a single isolated ability. It is a composite of many elements,

NOTES

such as knowledge of word meanings, ability to get meaning from sentences, ability to arrange through unit and sentence units into logically organized wholes, and ability to find desired material quickly. The teacher has a real basis for instructional procedures with this knowledge.

Language is another basic subject in which many delicately balanced skills are interwoven in an extremely complex manner. Here again the elements of achievement in the total process must be identified. Blind trust in general practice on the total skill must necessarily give way to the exact identification and discovery of the particular points of pupil weakness as a basis for special emphasis.

Good diagnosis must parallel the process of good teaching. Effective diagnostic materials in any school subject can be prepared only after the skills contributing to success in that subject have been isolated and identified. Psychologically, the reason for this is that on the whole the child learns to do what he practises or does. Remedial work, accordingly, can function only when the point at which pupil mastery breaks down has been located. Thus, the analysis must be penetrating and the diagnosis must be precise.

Specific nature of diagnosis

Diagnosis must be more exact than merely a broad statement of general functions. It is not enough to discover that a child is unable to read silently. The exact nature of his handicap must be revealed before it is possible to undertake a remedial programme. The more specific the diagnostic information revealed, the more exactly the remedial material can be made to fit the need. To return to a frequently used illustration, if it is found by diagnosis that a child is unable to add, unless the exact point at which his mastery of addition breaks down can be determined by the diagnosis, teaching, or remedial efforts are largely wasted. One of the outstanding reasons why more effective teaching and remedial work has not been done in certain fields is that no adequate analysis of basic skills can be made or has been made.

Importance of diagnostic use of test results

Tests as such are incapable of improving instruction because of any inherent power. Existing conditions are merely revealed by them. Remedial or corrective teaching is the result of deliberate constructive effort by the teacher after the particular points of weakness in the instruction of the pupils has been revealed by the tests. The ease, the clarity and the directness with which these needs are revealed by the tests are a measure of their real educational value.

Very few existing tests are so constructed as to permit the interpretation of their results directly in terms of an effective remedial procedure. However, this seems to be no good reason for the failure of teachers to apply more directly the results of this work in testing to the improvement of their teaching practice. Just as the data revealed by the navigator's instruments require calculation and interpretation, so is it necessary to analyse test data carefully in order to make them the basis of genuine remedial programme.

NOTES

Diagnosis as the basis for remedial work

Accurate diagnosis of class and individual pupil difficulties, coupled with application of remedy is not only important but necessary for teachers. The success of the remedial or corrective teaching depends upon the accuracy and detail with which the specific skills involved in successful achievement in the subject are identified and isolated in the test. Tests of the general survey type, or tests that report single unanalysed scores, cannot supply this information in sufficient detail.

Diagnosis as the basis for preventive work

Diagnosis as applied in education has taken on a meaning indicative of a breakdown in method, a failure of instructional techniques. Unquestionably, one of the basic purposes of diagnosis is the location of weaknesses and the determination of their causes, but there is nothing in the method that precludes its use in the prevention of weaknesses through anticipation of their causes. Out of the knowledge gained through the use of diagnostic procedures should become the basis for preventive work of all types. The existence of a weakness implies a failure at some point in the programme. The real importance in the discovery should lie rather in the prevention of its reappearance elsewhere under similar conditions.

An illustration from the field of medicine may make this point somewhat more concrete. In every medical examination for diagnostic purposes, a complete analysis is made and an exact case record of all observations is kept. Due to the analysis of these records, a better understanding of the causes and characteristics of certain types of human ailments is possible. Out of this same type of analysis has also come the basis for much of the preventive work that characterizes modern medical science.

In a similar way, accurate and detailed educational diagnosis may ultimately offer the basis for the development of a programme of preventive work in education. For example, if after diagnosing the addition of fractions in the fifth grade, it is found that the failure of pupils to reduce the fractions to their lowest terms in the answers is a common weakness, the obvious thing to do is to correct the defects at once, and then proceed to reconstruct the first instruction so that in the following year the causes for this particular weakness may not operate so powerfully. Similarly, any weakness identified now should afford the basis for decisions calculated to reduce the probability of their recurrence in the future.

2.5.1 Types of Diagnostic Tests

The diagnostic tests are broadly classified into two categories: (i) educational diagnostic tests, and (ii) physical or clinical diagnostic tests.

(i) Educational diagnostic tests: These tests are related to different subjects for specific level or class or grade.

(ii) Clinical diagnostic tests: These tests are also of several types relating to hearing, vision and other aspects.

NOTES

The regular classroom teacher should take help of a clinical specialist. Due to their knowledge about the students, classroom teachers can be more confident in their orientation towards group approaches. The teacher would expect normal actions and behaviours on the part of all students. They would be prepared, however, to deal with reasonable departure from normal behaviour, and for the unusual cases, the teacher would feel free to invite the person with special skill in diagnosing and treating such cases to bring them at par to the normal students of the class.

Functions of diagnosis

Although diagnosis is an individual activity, it has four functions:

(i) Classification
  (a) Intellectual level
  (b) Vocational level
  (c) Aptitude or Musical level

(ii) Assessment of specific ability
  (a) Level of adjustment
  (b) Level of abnormality
  (c) Level of depression and anxiety

(iii) Aetiology function refers to study the diagnosis or causes.

(iv) Remediation may be of following types:
  (a) Clinical treatment for physical ailments
  (b) Counselling for mental ailments
  (c) Remedial teaching for learning weakness
  (d) Special education for handicapped

These functions are highly individualized.

Methods of diagnosis

In view of the above functions, the following methods are used for diagnosis:

(i) Observation method: It is most popular method used for both prognosis and diagnosis. It is used for children. It is a subjective method, as well as being the most commonly used method.

(ii) Testing procedure used for diagnosis: This may include: (a) Clinical testing method, (b) Psychological testing method, and (c) Educational testing method.

It has been mentioned that the teacher has to invite other skilled persons to know the deficiencies of the students to provide proper remediation.

Steps for construction of diagonostic test

The following steps are used for preparing the diagnostic tests:

- Formulation of objectives and outline of the content or topic.
- Content analysis into sub-topics and its elements: (a) Sequence of sub-topics and elements within the sub-topic; (b) Sequence of learning points.
- Identifying in difficulty order of sub-topics.

- Deciding types of the items.
- Preparing items and try-out.
- Item analysis of test items and modification of items.
- Analysis of logical sequence of content.
- Preparing the final draft of the test.
- Preparing manual of the test.
- Remedial devices or measures.

Diagnostic test in school subjects

Let us look at the different diagnostic tests in school subjects.

(i) Diagnostic reading test: Reading is an important skill for school subjects. Low silent reading test includes the following sub-tests:

(a) Role of reading and comprehension of prose.

(b) Poetry comprehension and appreciation.

(c) Vocabulary in different areas.

(d) Meaning of sentences.

(e) Paragraph comprehension.

With these tests, the following types of errors are committed by the students: wrong pronunciation of words, spelling errors, omission error, repetition error, placement of words, combining word, adverse reading etc. The teacher has to locate the real causes for these errors, only after that remediation would be provided.

(ii) Diagnostic test of mathematical skills: The compass diagnostic test in arithmetic is used on the topic: addition, subtraction, multiplication and division. The following types of errors are committed by the students:

(a) Carry over in addition.

(b) Borrowing in addition.

(c) Borrowing and reducing one from next.

(d) Placement of decimal error.

(e) Tables are not remembered in multiplication.

(f) Placement of decimal in multiplication.

(g) Tables are not remembered in division.

(h) Noting wrong numbers.

Example of early over error in addition

(1) $\begin{array}{r} 43 \\ +56 \\ \hline 99 \end{array}$ (2) $\begin{array}{r} 54 \\ +74 \\ \hline 128 \end{array}$ (3) $\begin{array}{r} 67 \\ +35 \\ \hline 912x \end{array}$ or $\begin{array}{r} 67 \\ +35 \\ \hline 102\surd \end{array}$

NOTES

Example of borrowing error in subtraction

(1) $\begin{array}{r} 78 \\ -46 \\ \hline 32 \end{array}$ (2) $\begin{array}{r} 96 \\ -75 \\ \hline 21 \end{array}$ (3) $\begin{array}{r} 52 \\ -38 \\ \hline 26x \end{array}$ or $\begin{array}{r} 52 \\ -38 \\ \hline 14\surd \end{array}$

The teacher has to identify the real cause of wrong answers, then he will be able to provide the remedial teaching for these errors.

Achievements *vs.* Diagnostic Tests

Some important characteristics of the two type of educational tests are enumerated as follows:

(i) An achievement test provides the extent of the leering outcomes of the students. The performance is evaluated in terms of a single one, which is intended to locate specific weakness in the learning of content, but does not provide a single score.

(ii) An achievement test is quantitative, while a diagnostic test is qualitative. An achievement test has prognostic purposes.

(iii) The items of an achievement are arranged in the order of difficulty value. The easiest item in the beginning of the test and most difficult item is placed at the end of the test. But in a diagnostic test, items are arranged in the learning sequence in order of positive transfer of learning.

(iv) In achievement tests, items are analysed on the basis of correct responses of the students on an item, 'device method' of item analysis is used for selecting and rejecting the items. A diagnostic test items are analysed on the basis of wrong responses of the students on items. Stanley method of item analysis is used for diagnostic test.

(v) The contest of a subject is the same in both the tests. The objective-type items are used in both types of tests. In an achievement test, one score is assigned for correct answer and zero for wrong answer. In diagnostic test, scores are not assigned for correct answer by wrong responses, considered in view of the sequence of the content to identify the cause for the wrong answer.

(vi) The scoring and interpretation is easy for achievement tests scores. The subject specialist is not required, but scoring and interpretation of diagnostic test scores is difficult. The subject specialist is a must for identifying the causes of the wrong answers.

(vii) Essay-type and objective-type tests are used in the attainment tests, but only objective type can be effectively employed in diagnostic tests. Essay- type cannot be used in diagnosis purpose.

(viii) Reliability and validity are essential characteristics of an achievement test. But diagnostic sub-tests or one learning unit are adequately reliable. A diagnostic test is an individual test from interpretation point of view, but otherwise it is a group test.

NOTES

(ix) In constructing an achievement test, the sampling of content is done for consideration; no learning point is omitted or ignored.

(x) The diagnostic tests are effective tools for teachers for planning and organizing remedial teaching. Tutorial groups are formed on the basis of diagnostic test. Teacher can help the poor students and can raise their level of performance by removing their difficulties. The purpose of diagnostic tests is the remedial teaching and instruction. The achievement tests are used for selection, promotion gradation, classification and also in research work.

CHECK YOUR PROGRESS

7. What are the two purposes of diagnostic tests?
8. What are educational diagnostic tests?

2.6 SUMMARY

- Tools or question papers prepared by the teachers to evaluate their own students whom they have been teaching are called as teacher made tests.
- A standardized test is a test, which is developed in such a way that it reaches up to a specific level of quality or standardization. The test is standardized with respect to the form and construction, administration procedure and test norms.
- The type of evaluation in which the performance of the testees are evaluated with reference to some predetermined criteria is called as criterion-referenced evaluation. No weightage is given to norm or average performance of the group in this evaluation.
- An educational test is not just a test that measures achievement in subjects of study, but is also a psychological test that leads to an assessment of the overall development of a student.
- The aspects which affect the characteristics of a good test are as follows:
  - Validity of the test
  - Reliability of the test
  - Objectivity of the test
  - Usability of the test
  - Comprehensive and preciseness of the test
  - Administration of the test
  - Test from economic viewpoint
  - Availability of the test
  - Appearance of the test
  - Standardization of the test
  - Norms of the test

- Tests are divided into different types taking into consideration their content, objective, administration system, scoring style etc.
- An achievement test is an instrument designed to measure the relative achievement of students. It is an indispensable instrument in the teaching–learning process.
- A questionnaire is a tool for research, comprising a list of questions whose answers provide information about the target group, individual or event. Although they are often designed for statistical analysis of the responses, this is not always the case. This method was the invention of Sir Francis Galton.
- Diagnostic tests are a kind of educational tests. These tests have two purposes: (i) prognostic purpose, and (ii) diagnostic purpose.
- The prognosis of the students done in the specific subjects which has been taught to them. The diagnosis means to identify the causes of their weakness of their poor attainment of the students.
- The diagnostic tests are broadly classified into two categories: (i) educational diagnostic tests, and (ii) physical or clinical diagnostic tests.

2.7 KEY TERMS

- **Standardized test:** A standardized test is any form of test that requires all test takers to answer the same questions, or a selection of questions from common bank of questions, in the same way, and that is scored in a 'standard' or consistent manner, which makes it possible to compare the relative performance of individual.
- **Criterion-referenced test:** Tests that are designed to measure student performance against a fixed set of predetermined criteria or learning standards.
- **Objectivity:** Objectivity is a noun that means a lack of bias, judgment, or prejudice.
- **Pictorial:** Something that is expressed in pictures or is illustrated.
- **Remedial:** A remedial action is intended to correct something that is wrong or to improve a bad situation.

2.8 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Tools or question papers prepared by the teachers to evaluate their own students whom they have been teaching are called as teacher-made tests.
2. The type of evaluation in which the performance of the testees are evaluated with reference to some predetermined criteria is called as criterion-referenced evaluation.
3. Validity of a test refers to its truthfulness; it refers to the extent to which a test measures what it intends to measure.
4. Objectivity of a test indicates two things: item objectivity and scoring objectivity.

5. According to Popham, 'the achievement test focuses upon an examinee's attainments at a given point in time.
6. The instructional uses of achievement tests are:
   - For class analysis and diagnosis
   - For individual pupil diagnosis
   - Guidance use of achievement tests
   - Alternative uses
   - Administrative uses
   - Aid to the parents
7. Diagnostic tests have two purposes. These are:
   (i) Prognostic purpose
   (ii) Diagnostic purpose
8. Educational diagnostic tests are tests related to different subjects for specific level or class or grade.

NOTES

2.9 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is a standardized test? What are its characteristics?
2. Differentiate between criterion-referenced tests and norm referenced tests.
3. What are the essential characteristics of a good test?
4. Discuss the merits and demerits of different types of tests.
5. What are the different types of diagnostic tests?

Long-Answer Questions

1. Describe how a criterion-referenced test can be constructed.
2. Discuss how tests can be classified.
3. Discuss the difference types of questionnaires that are used by researchers.
4. Describe how a diagnostic test can be constructed.

2.10 FURTHER READING

Bloom, B. S. *Taxonomy of Objectives of Cognitive Domains*. New York: Longmans Green &Co.

Ebel, R.L. *Essentials of Educational Measurement.* New Jersey: Englewood Cliff.

Grunlund, N. E. *Measurement and Evaluation in Teaching* (3rd Ed). New York: Mcmillan Publishing Co.

Julian, C., Stanley and Hopkins. *Educational and Psychological Measurement and Evaluation*. New Delhi: Prentice Hall of India Pvt. Ltd.

Lindermon, Richard, H. *Educational Measurement*. Mumbai: B. S. Taraporevala Sons & Co.

UNIT 3 PSYCHOLOGICAL TESTING

NOTES

Structure

3.0 INTRODUCTION

In the previous unit, you learnt about the major tools and techniques used in evaluation. In it, standardized tests, criterion-referenced tests and diagnostic tests were discussed. In this unit, the discussion will turn towards psychological testing. Psychological testing refers to the administration of psychological tests. A psychological test is an objective and standardized measure of a sample of behaviour. There are many types of psychological tests - intelligence tests, personality tests, occupational tests, and so on. This unit will discuss psychological tests in detail. The unit will also discuss different types of examination systems such as semester system and the open book system.

3.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Discuss the characteristics and types of attitude
- Describe the determinants of personality
- Discuss some of the projective techniques of examining personality.
- Examine some of the major trends in examinations and evaluations seen in recent times

3.2 PSYCHOLOGICAL TESTING IN THE AREA OF INTELLIGENCES

NOTES

Intelligence of an individual or group of individuals is tested through various types of tests that are called intelligence tests. Intelligence test can be studied by classifying it in the following ways:

1. Classification of intelligence test on the basis of administration:
   (i) Individual intelligence test
   (ii) Group intelligence test
2. Classification of intelligence test on the basis of the nature of the test:
   (i) Verbal intelligence test
   (ii) Non-verbal intelligence test
   (iii) Performance intelligence test

- **Individual intelligence test:** This test can be given to each individual separately. The first practical general intelligence test, called as Binet–Simon scale, was administered individually.
- **Group intelligence test:** It is the intelligence test, which is administered in groups. The first practical group intelligence test was developed for the Armed Forces during the First World War.

  Table 3.1 shows the comparison between individual and group intelligence test.

Table 3.1 *Comparison between Individual and Group Intelligence Test*

Individual intelligence test	Group intelligence test
• Can be administered on one individual at a time. • It is developed and administered on young children. • It is better predictors for young children. • It is more expensive and time consuming. • It is highly useful in clinical settings. • It is difficult to construct.	• Can be administered on a group of individuals at a time. • It is more frequently used for testing normal adults (adolescents or older). • It is better predictors for adults' intelligence. • It is much less expensive and time consuming. • It is less useful in clinical setting. • It is easy to construct.

- **Verbal intelligence test:** A verbal intelligence test needs the following verbal capabilities in the tests regarding the language in which the intelligence test is developed and administered:
  - o Understanding spoken language
  - o Understanding written language

NOTES

- Ability to speak language
- Ability to write language
- Verbal comprehension factor

A test may require the first four items from the given list or may require all. It is also possible to compose a test in which none of the first four aspects is present and the fifth aspect, verbal comprehension, is the only aspect acting as a cardinal requirement. Each test is examined in terms of its combination of verbal requirement rather than simply classified as 'verbal' or 'non-verbal'.

- **Non-verbal intelligence test**: The intelligence test that is based on symbols or figures and in which language is either very less used or not used is referred to as non-verbal intelligence test. The subject indicates the correct answer either through the use of language or by marking one of a number of choices. The symbolic response might be made with respect to objects rather than printed materials.
- **Performance intelligence test**: In this type of intelligence test the subject is required to handle or manipulate objects in such a way as to complete a specified task. Intelligence regarding skills of performing some task is tested through this test. Performance test usually measures more coordination, speed, perceptual and spatial factors. Any kinds of instrument that is in a printed form is referred to as verbal test as this emphasizes verbal comprehension and requires symbolic responses. Table 3.2 shows the comparison between verbal, non-verbal and performance test.

Table 3.2 *Comparison of Verbal, Non-Verbal and Performance Test*

Verbal intelligence test	Non-verbal intelligence test	Performance test
1. It requires understanding of written language.	It requires symbolic responses on the part of testees.	It requires manipulative responses by the testees.
2. Considers verbal comprehension factors.	Considers symbolic relation reasoning factor.	Considers perceptual and spatial factor.
3. Conducted by writing language.	By marking correct choices.	By handling the objects correctly.
4. Consists of verbal comparative, numerical computation and numerical reasoning.	Consists of reasoning for abstract and seeing the relationship perception.	Consists of perceptual ability and spatial ability.
5. Effect of culture and language is observed.	Effect of culture and language is not observed.	Effect of culture and language is not observed.
6. May be used for individual and group testing.	May be used for individual and group testing.	Only for individual test only.
7. It can be used on literate subjects only.	It can be used on illiterate and literate both.	It can be used for both literate, illiterate, children and adult.
8. It is easy to construct, administer and score.	It is difficult to construct but easy to administer and score.	It is difficult of construct, administer and score.

NOTES

Some important tests of Intelligence

There are many tests for measuring intelligence. As discussed above, they can be of three types namely verbal, non-verbal and performance types. Some important tests of intelligence are given as follows:

- Stanford–Binet test of intelligence
- Wechsler–Belleve intelligence test
- Thurstone's test of primary mental abilities
- Non-verbal test of intelligence (also known as Reven's Progressive Matrices)
- Culture free intelligence test
- Bhatia–battery performance test

Note: To study and for using any of the intelligence test, consult any psychological test library or psychology lab. Read the manual of the test and try to use it to have first hand experience.

CHECK YOUR PROGRESS

1. How are intelligence tests classified on the basis of administration?
2. What is a performance intelligence test?

3.3 ATTITUDE AND PERSONALITY TESTS

Let us now study attitude and personality tests.

3.3.1 Attitude

L.L. Thurstone states that an attitude 'is the degree of positive and negative affect associated with some psychological object'. Attitude refers to the predisposition of an individual to evaluate some aspect of his world, including ideologies, objects, symbols, people, etc. Attitude is the bent of mind that may be positive, negative, hostile or indifferent. One's attitude is reflected in his behaviour. In a nutshell, attitude includes the sum-total of an individual's inclination, feelings, prejudice or bias, preconceived notions, ideas, fears, threats and convictions about something. For example, somebody's attitude towards Indian culture means his thinking about the culture of India, whether he likes or dislikes Indian culture. Social attitudes are learned from the society through interaction with the social members. Attitude of an individual changes from time to time and it is not an inborn quality, rather it is acquired.

Anastasia states, 'an attitude is often defined as a tendency to react favourably towards a designated class of stimuli, such as a national or racial group, a custom or an institution. Thus defined, attitude cannot be directly observed but must be inferred from overt behaviour, both verbal and non-verbal'.

Freeman states 'an attitude is a dispositional readiness to respond to certain situation, persons or objects in a consistent manner which has been learned and has

become one's typical mode of response. An attitude has a well-defined object of reference. For example, one's view regarding a class of food or drink (such as fish and liquors), sport, maths or democrats are attitudes'.

NOTES

Characteristics and types of attitude

In this section, we discuss the characteristics and forms of attitude:

Characteristics

The characteristics of attitude are as follows:

- Attitudes are not inborn but acquired.
- Attitudes differ from culture to culture.
- Attitudes are integrated into an organized system.
- Attitudes are less consistent, they can be modified.
- Attitude is a position towards outer objects, either 'for' or 'against'.
- Attitudes are socially learned.
- Attitudes are formed due to the process of differentiation.
- Attitudes are complex.
- Attitudes have readiness to respond.
- Attitudes have a time dimension.
- Attitudes have a duration factor.
- Attitudes are predisposition towards objects.
- Attitudes are observed from overt behaviour.
- An attitude has a well-defined object of reference.
- It implies a subject–object relationship.

Types

The predispositions towards different things create different attitude. Attitude of a person varies from situation to situation, time to time, etc. Some people have positive attitude towards religious thoughts and morality, but some has negative attitude towards it. In general, attitude is of two kinds: (i) positive attitude and (ii) negative attitude. For example, Gandhi was interested in non-violence, so he had positive attitude towards non-violence. On the other hand, he was against violence, so he had negative attitude towards violence. Besides this kind of classification of attitude, attitude may be classified into different categories like:

- Acquisitive attitude
- Play attitude
- Scientific attitude
- Business attitude
- Artistic attitude
- Religious attitude

NOTES

The people who have interest in acquisition of something all the time, they have acquisitive attitude. At each and every time, they search for something. Play attitude refers to the interest in playing. The people with playing attitude, always give importance upon playing. Some people are more prone towards science and technology from the childhood, they involve themselves in small scientific activities. They have scientific bent of mind and this is known as 'scientific attitude'. Some people are very much interested in commerce and business activities. This refers to their business attitude. People with skill in art and craft, have artistic attitude. They have their interest in artistic plays, arts and crafts etc. Religious attitude refers to the positive bent of mind to religious activities. The people with religious attitude, believe in different forms of God, and they are known as theists. But the people who have negative bent of mind towards God and religion, are known as atheists. They do not like to involve themselves in religious activities.

Determinants of attitude and attitude testing

Attitude is not an inborn quality. At the time of birth, child's mind is like a clean slate in which the society and environment write everything. Interacting with parents, peer group, siblings, neighbours, society and school mate, child develops his attitude towards an individual or object or anything else. So there are certain factors which determine the attitude of human beings. The determinants are:

1. *Cultural or social determinant:* A child acquires everything from the culture or society, to which he belongs. The cultural ethics, social norms, ceremonies, the religious and moral beliefs of the particular society are acquired by the child. His attitude towards something is framed according to that social or cultural framework. For example, tribal child receives all the aspects of his tribal culture. His language, way of living, thinking, education, all the aspects will be according to the tribal society he is concerned with. For example, the interior tribal people have their attitude that, the outside people are alien to them and they destroy their culture and civilization. So, the child of this tribe, also develops this kind of attitude towards the outsiders.
2. *Psychological determinants:* One's psychology determines his attitude. The person who is kind enough and merciful has sympathetic attitude towards the poor. Emotional and personal experiences, social perception etc., contribute towards development of attitude.
3. *Functional determinants:* Functional determinants are also known as situational determinants. Sometimes, the situation builds attitude in human minds. For example, when somebody is rebuked and mistrusted by his friends in a certain situation, his attitude suddenly changes towards his friends. So, situational factors are very much responsible for attitude development.

Purposes of Attitude Testing and Measurement of Attitude

Purposes of Attitude Testing

Behind every action, there is purpose. When we go for testing attitude of people or students, obviously there are some purposes behind it. In the teaching–learning situation, attitude testing has an important role to play. So the purposes of attitude testing are:

- To assess the entry attitude of the students towards a particular course.
- To assess how far the desirable attitudes have been developed in the students during the course and after the completion of the course.
- To help the students to develop positive attitude towards certain things.
- To help the students in their career plan.
- To help the management to make its administration and supervision a qualitative one.
- To help the teacher to overcome their weakness in the teaching–learning situation.
- To help the students to check their undesirable behaviour.

Measurement of Attitude

Attitude is a subjective concept which is not absolute, rather relative. So, when test is prepared for testing attitude, certain dimensions are to be kept in mind. The dimensions are:

- Direction
- Degree
- Intensity

From the direction point of view, there are two kinds of directions: (i) positive and (ii) negative. When an individual has positive bent of mind towards something, it is known as 'positive attitude' and when he has negative bent of mind towards something it is known as 'negative attitude'. Every student's attitude should be measured in relation to his teaching–learning situation.

Every attitude has its degree. For example, a person who sings occasionally has less degree of positive attitude towards singing in comparison to the person whose profession is singing. So at the time of measuring attitude, the degree of predisposition should be taken into consideration.

Attitudes also have an intensity dimension. At a high degree of intensity, some kind of behaviours are motivated towards a greater extent. So all these dimensions should be kept in mind at the time of attitude testing.

The methods to be followed for the measurement of attitude are:

(i) Thurstone Scale: This scale was developed by Thurstone. Thurstone's attitude scale is known as equal-appearing interval scale. In this scale, both favourable and unfavourable statements regarding a particular topic are reflected in an eleven point scale. The respondent is supposed to check the point for each item according to his attitude. The median of the judged locations for an item is its scale value. The scale position are very much a function of the judges who are chosen.

(ii) Likert Scale: This scale was developed by Likert. All the items of this scale are followed by five options. The respondents are supposed to point out the option they like. The decisions are either favourable or unfavourable on the object, or person. Judges are not appointed for this scale, and this scale is

NOTES

known as 'five-point scale'. Likert type scale is less time consuming and economic. Its approach is more empirical because, it deals with the respondent's score directly rather than employing judges. The sum of the item credits is the total score of the individuals, which is interpreted in terms of empirically established norms.

Example of Likert type scale:

1. 'Science is the	*Scales*	5	4	3	2	1
soul of present	*Values*	SA	A	U	D	SD
day society'.						

Values:

SA	–	Strongly Agree	5
A	–	Agree	4
U	–	Undecided	3
D	–	Disagree	2
SD	–	Strongly Disagree	1

Limitations of attitude testing

Attitude testing has certain limitations which cannot be avoided. The limitations are:

- Attitude is a subjective concept, so it is very difficult to measure attitude quantitatively.
- Attitude is such a complex affair that it cannot be represented by any single numerical index.
- Attitude is learned not inborn. So it varies from situation to situation and time to time.
- In most of cases, it is seen that there is difference between verbally expressed attitudes and attitudes reflected in behaviour.

3.3.2 Personality Tests

The concept of personality is subjective in nature. It is very difficult to assess or measure subjective concepts objectively. Still, psychologists have tried to measure the personality of human beings through different tools and techniques. In the primitive society, physical strength was the norm of personality Measurement; during the *Vedic* period, the memorization of *Vedas* was the norm of personality measurement. Later, astrology, palmistry, physiognomy and phrenology were considered as the measures of personality.

The methods used for assessment of personality may be categorized as subjective, objective and projective techniques. It is very difficult to bring a watertight compartment among all these assessment techniques. Some of the important techniques are discussed here.

Concept of personality

According to G.W. Allport, 'personality is the dynamic organization within individual of those psycho-physical systems that determine his unique adjustment to his

environment'. Eysenck defined personality as 'the more or less stable and enduring organization of a person's character, temperament, intellect and physique, which determine his unique adjustment to the environment'. According to Watson, 'personality is the sum of activities that can be discovered by actual observations over a long period of time to give reliable informations'. For Morton Prince, personality is 'the sum total of all the biological innate dispositions, impulses, tendencies, appetites and instincts of the individual and the dispositions and tendencies acquired by experience'.

In a nutshell, personality refers to the external appearance and internal qualities of an individual. It is something unique to everyone, and it is the result of the interaction of heredity and environment. Personality refers to individual's unique and relatively stable patterns of behaviour, thoughts and feelings. We cannot draw a watertight compartment between personality and all the psychological traits. Personality is a summative approach which assesses all the integrative qualities of an individual. Learning and acquisition of experiences in every platform of life contribute towards growth and development of personality.

Characteristics of personality

The characteristics of personality are as follows:

- Personality is built by heredity and environment.
- There is individual difference in personality.
- Personality determines one's adjustment to his environment.
- Personality emerges from the interaction of psycho-biological organism.
- Personality may be intrinsic or extrinsic.
- Personality is the reflection of all the psychological and physical traits of an individual.
- Personality can be assessed.
- Personality means man within the man.
- Personality refers to social adaptability.
- Personality is a dynamic organization.
- Behaviour is the reflection of personality.
- Personality permits a prediction about an individual.
- Personality is more or less stable in nature.
- Personality exhibits self consciousness.
- Personality includes all the behaviour patterns: conative, cognitive and affective.
- Personality includes conscious, semi-conscious and unconscious activities.
- Learning and acquisition of experiences contribute towards growth and development of personality.
- Personality should not be taken as synonymous with one's character and temperament.
- Personality of an individual is directed towards some specific ends.

NOTES

Determinants of personality

An individual is the by-product of his heredity and environment. Many believe that heredity plays a major role towards personality development, but in reality, no one factor can be given credit of influencing personality. Personality is the by-product of many factors which are discussed below.

(i) Heredity: In most of the cases, children are more similar to their parents and siblings. Heredity influences physique, motor-sensory equipment and level of intelligence. Physical appearance of an individual contributes a lot for his personality. So, heredity does play an important role in the development of personality.

(ii) Environment: Here, we will discuss about three types of environment: (a) Physical environment, (b) Social environment, and (c) Cultural environment. These factors make an individual to 'acquire' a personality.

(a) Physical environment: Physical environment refers to the physical, geographical and geological structure of the area where an individual lives. In case of the cold countries, people are white; but in hot countries, people's complexion is black. The people of North East of India are different from the people of South India in their colour and physical appearance. This difference is mostly due to the physical environment.

(b) Social environment: Social environment of an individual includes all social agents like parents, siblings, peer groups, school, teachers, neighbour etc. All these factors play their role in the development of personality of an individual. Children from homes where morality, honesty, spiritualism, religiousness are given importance, are different from the children from the homes of poverty, family disorder, formal relationship among all etc. A child's maximum time is devoted in school with his teachers and friends. Teachers are the role models before the children. The principle of discipline, living cooperatively, respect to teachers, feeling brotherhood and sisterhood in uniform dress, all these are acquired by schools. Teacher's open mindedness, democratic look, enthusiastic and industriousness, leave a mark upon a child which develops his personality. Interpersonal relationship among the members of a society are important means which help in the development of certain social personality characteristics. All these factors help in the development of personality.

(c) Cultural environment: Cultural values moulds the personality of people belonging to that culture. A child internalizes the values, ideas, beliefs, norms and customs of a culture through the interaction with this culture and the society. Every society has its own cultural heritage, and this cultural heritage transmits from generation to generation successfully which is known as 'enculturation'. The personality of the people of Eastern society is different from the personality of the people of Western society, it is due to the cultural change. Margaret Mead conducted a study on the adolescents of Somoa, a primitive culture. She concluded that the cultural pattern of a society influence to a great extent the personalities of individuals.

Subjective and objective techniques

NOTES

1. **Observation:** Through observation in real-life situation, the personality of an individual can be known. Observation is of two types: (a) participant observation, and (b) non-participant observation. In case of participant observation, the observer becomes a part of the group or becomes a part of the life of the individual whose behaviour is observed. But in case of non-participant observation, the observer remains away from the group or from the individual. Observation is a first-hand information, so it is more valid and reliable. But sometimes the observer's biasness affects the assessment. Some people also hesitate to reflect their natural behaviour when they know that somebody is observing them. Behaviour is a subjective concept, and it varies from situation to situation. But, observation has a great role in assessing personality.
2. **Interview:** Interview is a technique of eliciting response from an interviewee by a couple of interviewers. In the words of Maccoby and Maccoby, 'interview refers to a face-to-face verbal interchange in which the interviewers attempt to elicit information or expression of opinion or belief from the interviewees'. Within a limited period of time, the interviewers ask questions to the interviewees and elicit responses from them. Interview may be of two types: (a) structured interview and (b) unstructured interview. In structured interview, some questions are pre-fixed to be asked. The type of behaviour to be assessed is also pre-planned. But in unstructured interview, the questions are situational. Open interrogation takes place between the interviewers and interviewees. The scope of unstructured interview is more than structured interview because in the former case the interviewers are not restricted to a particular set of predetermined questions. Interview techniques have their own limitations as well. Skilled interviewers are necessary and this technique is time consuming. It also suffers from the personal biasness of the interviewers.
3. **Checklist:** A checklist consists of some items and the subject is supposed to respond to the items according to his interest. From the responds, the personality of the subject is assessed by the psychologists. For example, in the list given below, the subject is asked to check the items which are applicable to him.

   Example,

   - I want to work whole day.
   - I want to lead a comfortable life.
   - I want rest whole day.
   - I want to pray whole day.

   Examples of standardized checklists are: (a) Money problems checklist, and (b) Student problem checklist made by Central Bureau of Educational and Vocational Guidance.
4. **Rating scale:** Rating scale was developed by Galton in 1883. According to Van Dalen, 'a rating scale ascertains the degree, intensity or frequency of a variable'. In the words of Good and Scate, 'the rating scale typically directs

NOTES

attention to different parts or aspects of the thing to be evaluated, doesn't have as many items or categories as the checklist or score card'. Through this device, a rater can record judgments of another person or his own based on the traits defined by the scale. It covers much wider range of natural behaviour than any practicable battery of personality tests. There are different types of rating scales like: (a) Numerical scale, (b) Graphic scale, (c) Cumulative points scale, (d) Standard scale, and (e) Forced-choice rating.

5. **Personality inventory:** The tool was developed by Woodworth in 1919. But at that time, it was known as 'personal data sheets'. In personality inventory, an effort is made to estimate the presence and strength of each specified trait through a number of items representing a variety of situations in which the individuals' generalized mode of responding may be sampled. Personality inventory is of two types: (a) uni-dimensional personality inventory, and (b) multi-dimensional personality inventory.

   Personality is of multi-dimensional quality. That is why the multi-dimensional personality inventory is used in most of the cases. The reliability of personality inventories ranges from very low to satisfactory coefficient 0.8, depending on the traits being measured. Personality inventories are particularly useful in group trends, i.e., in differentiating between groups of adjusted and maladjusted, rather than among individuals.

   Minnesota Multiphasic Personality Inventory (MMPI) is mostly used in the assessment of personality. It was developed in 1940 its first manual was used in 1943. It is available in individual card form and group booklet form. It consists of total 550 items and is used on the individuals above the age of 16. MMPI consists of three answers: True, False and Can't say; and it is mainly used in clinical and diagnostic centres. The time required for administration of MMPI is 90 minutes and its reliability ranges from 0.50 to 0.90.

Projective techniques

The term 'project' was used for the first time by Sigmund Freud in the field of psychology. Projection, according to Freud means externalizing of conflicts or other internal conditions that give rise to conscious pain and anxiety. 'Projection' refers to encourage an individual to use his imagination as per his interest. Projective techniques are those which draw out responses from the unconscious. It enables an individual to reflect his internal feelings like aptitudes, attitudes, values, philosophies etc. All the projective techniques are dealt with complex mental processes which reflect the whole personality of an individual.

The characteristics of projective techniques are as follows:

- Unstructured and ambiguous materials are used in projective techniques.
- It studies the unconscious mind of individuals.
- It studies the total personality of individuals.
- It evokes multi-dimensionality of responses.
- There is freedom to respond on the part of the respondents, when responding the techniques.

NOTES

- The answers to projective techniques are not right or wrong.
- The purpose of projective techniques are disguised.
- The validity and reliability of projective techniques are more.
- Projective techniques have holistic approach in assessing personality.
- These techniques are specially useful for children.

Projective techniques also suffer from certain limitations as follows:

- Projective techniques are subjective.
- Experts are required for use of these techniques.
- It is difficult to construct, administer and evaluate projective techniques.
- It lacks objectivity in evaluation.
- It lacks norms.
- Interpretation of the test is time consuming.
- It is very difficult to measure its validity and reliability.
- In most of the cases, directions for the administration of the projective techniques are not standardized.
- The evaluation of projective techniques varies from evaluator to evaluator.

Some of the projective techniques are explained below.

1. *Thematic Apperception Test (TAT)*

This technique was developed by Morgan and Murray in 1935. Thematic Apperception Test is called a story-telling test because the respondents are to tell stories according to the pictures shown to them. TAT pictures are administered individually as well as collectively to group of individuals. The test material consists of 31 cards, 30 reflecting various pictures and one blank card. All the pictures are related to real-life situations through which an individual passes. All the pictures are given in two sessions and the number of pictures depends upon age, sex and aspect of personality to be assessed. After the administration of the test, the examinees are interviewed, and asked why they wrote such type of stories. The interpretation of the stories takes place with the help of these following points:

- The hero
- Theme of the story
- The end of the story
- Punishment for crime
- Defence and confidence
- Emotional expressions etc.

2. *Rorschach Inkblot Test*

This technique was developed by Swiss Psychiatrist Herman Rorschach in 1921. The test consists of 10 symmetrical ink blots on 10 separate cards of 11 × 9 inches. Five of the ink blots are black and white, and others are multicoloured.

NOTES

All the inkblots are completely unstructured. All the cards are given to the respondent at a time in a sequence. Here, the subject is given more time and is permitted to give as many responses as he likes. After this, the examiners generally conduct inquiry into the details of the subject's response on individual cards. For the scoring purpose, the responses are given specific symbols and entry is made into four columns: location, contents, originality, and determinants.

The reliability of Rorschach Inkblot test is 0.75, and the validity ranges from 0.5 to 0.8. The test requires trained and skillful persons for scoring and interpreting it, and therefore the work must be done only by experienced and trained psychologists.

Psychologically, projection is an unconscious process, whereby an individual attributes certain thoughts, attitudes, emotions or characteristics to other persons or objects in the environment. He expresses his needs to others in an environment and draws incorrect inferences from experience. Projective technique provides the subject with a semi-structured or unstructured stimulus situation, giving him an opportunity to impose upon it his own private needs and his own perceptions and interpretations. Projective techniques require the subject to describe or interpret objects other than himself, and these are based on the hypothesis that an individual's response to an unstructured stimulus is influenced by his needs, motives, fears, expectations and concerns. Interpretations are drawn on the basis of the completeness provided by the testee and the type of stimulus presented to him.

Several forms of projective technique are pictures, inkblots, incomplete sentences, word associations, one's own writings and drawings etc., which are intended to elicit responses that will reveal the respondent's personality structure (feelings, values, motives, characteristics, modes of adjustment and complexes). He is said to project the inner aspects of his personality through his interpretations; and creations his needs, wants, desires, aversions, fears and anxieties get reflected involuntarily.

Projective techniques are the means to dig out the intricacies of human behaviour which are below the surface and are incapable of exposure by subjective and objective techniques such as observation, peer ratings and self-rating questionnaires etc.

Most commonly known projective techniques are the Rorschach Inkblot Test and Thematic Apperception Test (TAT). Rorschach Inkblot Test provides an insight into the nature of disorder on the one hand, and hidden feelings, emotions and desires on the other, which the testee would not have normally desired to be known to others. It consists of a set of 10 inkblots: five made in shades of black and grey only, two contain bright patches of red in addition to shades of grey, and the remaining three employ various colours. In this, the subject is shown one blot at a time and is asked what he sees in it.

Thematic Apperception Test (TAT) consists of the pictures of people in various settings. It has a set of 30 cards and one blank card which are used in various combinations as per the age and sex of the subject. In TAT, the subject is told to make up a story about each picture in turn. In actual clinical practice, a set of 10 selected cards is used for the particular case. TAT is helpful in identifying the presence

or absence, as well as strength, of one or more motives to be used for diagnosing the problems. Other techniques include Children Apperception Test (CAT), Draw-a-Man Test, Word Association Test, etc. Artistic products of structured type or completely unstructured, like finger paintings, can also be used as projective techniques.

One of the major advantages of projective techniques is that they are not difficult to administer. These can be used for persons of all ages, ethnic groups and intelligence levels. Projective techniques are ingenious efforts to measure personality variables. They are said to measure the 'whole personality'. The projective techniques are, however, difficult to interpret for which sustained training is necessary. Despite their widespread use by clinical and school psychologists, projective techniques are relatively un-standardized ordinary measure of reliability that are difficult to obtain. Yet their popularity may be attributed to 'faith validity' and the mistakes assumed by the users.

Use of projective technique

As clarified earlier, projective techniques are useful in bringing out covert behaviour patterns, though indirectly, which are either not revealed through the use of subjective and objective techniques, or are concerned with very deep rooted motives and emotions which the respondent would not like to be exposed to. Children and adults unwilling to discuss their problem may be administered projective techniques. Some of the situations most appropriate for the use of projective techniques are:

- Identifying the subject's real concerns, his conception of himself, and the way he views his human environment.
- Diagnosing the behavioural problems of students.
- Studying the personality of children who are unable or unwilling to discuss their problems directly.
- Verification of the patterns of personality emerging out of subjective or objective techniques.
- Studying the home and school adjustment of adolescent.
- Studying personality or adjustment patterns of more sensitive subjects, or of those who tend to give biased responses on tests.
- Clinical usages.
- Testing illiterate persons of different language backgrounds.

Use of projective techniques by teachers

Projective techniques by their very nature require properly trained personnel to administer and interpret them. It is necessary to be aware of projective techniques and to be able assess if there is any utility in employing them to study various kinds of personality disorders and maladjustments among students. Teachers should not casually venture into using projective techniques unless they are properly trained for them. These techniques should be used only by clinical psychologists or school psychologists who have had intensive training in the use of these methods. In view

NOTES

of the complexity associated with the application of projective techniques, these should be employed in case of students showing the signs of very severe emotional problems. It should be noted that projective techniques are most successfully used for eliciting suppressed desires, feelings, ambitions, attitudes and emotions, guilt and complexes etc.

Limitations of personality assessment

The demerits of personality assessment are:

- Personality is a subjective and relative concept, so it is very difficult to assess it.
- Personality assessment cannot be purely objective.
- Sometimes, the result of personality assessment affects the individuals.
- Sometimes, the result of personality assessment is different from the real-life situation.
- Personality assessment varies from technique to technique.
- Personality assessment varies from evaluator to evaluator.

CHECK YOUR PROGRESS

3. How does L.L. Thurstone define attitude?
4. What is Thurstone scale?
5. What is the Rorschach inkblot test?

3.4 EXAMINATION SYSTEM: CURRENT STRATEGIES

Several methods of assessment and evaluation of learners are in use these days, such as unit tests, half-yearly examinations, annual examinations, semester system, board examinations and entrance tests.

Unit tests

Unit tests are prepared and conducted to check the understanding level and to know the problems of the students at initial level. In these tests, the subject matter is selected from a specific content taught in a limited time period. The subject matter is significantly pertinent to the objectives to attain in that period. The unit can be an entire module. Each unit test is totally independent from other tests to be conducted in the session.

Benefits of unit tests

There are many benefits of unit tests, such as:

- Objectives are clearer and well defined.
- Follow-up after evaluation is simpler.

- Students get opportunities for improvement.
- Each unit gets proper attention and importance
- Students' weakness can be diagnosed and rectified at very early stage.
- As the main examination is broken into small ones, students need to prepare a small portion of content at a time, which helps in effective learning.
- Understanding the objectives of teaching is facilitated.
- Students learn new skills.

Limitation of unit tests

The limitations of unit tests may be summarized as follows:

- They require efficient, hardworking and trained teachers
- If they are not systematically arranged, there may arise confusion and discouragement among the student.
- Proper feedback and diagnosing problems is not so easy.
- All the unit tests may not follow the same pattern. Hence, evaluation may not be reliable.

Half-Yearly and Annual Examinations

Besides unit tests, many schools conduct half-yearly and annual examinations. Half-yearly examinations, which cover almost half of the syllabus, are conducted after the first unit test. The aim behind it is to evaluate the student at slightly bigger level. It again helps students as well as teachers to find out the problems. It determines whether the mid-term objectives have been achieved or not. The other big advantage of the half-yearly examinations is that it provides the teachers an opportunity to assess their teaching methodologies. For example, if many students fail to perform well in a half-yearly examination, it means that there is some serious flaw in the teaching methodologies, which require teachers' attention. If the number of such students is less, then remedial teaching classes could be arranged for them.

Annual examination—also called summative examination— is the final examination. It decides whether the skills and abilities (academics and co-curricular) aimed to be achieved in the beginning of the session have been achieved or not. In the success of this examination, unit tests and mid-term examinations help a lot. On the basis of unit tests and mid-term examinations, students having problems in learning are identified and special classes are arranged for them so that by the end of the session they can perform their best.

Board Examinations and Entrance Test

The board examination system has broadly been classified into the following two levels in India.

1. State level boards

The state level boards are administered by each state. They are independent of the central boards. They design their own syllabus and standards. They facilitate the

learning process by generally following their state language as a medium of instruction. Not only the medium of instruction is a local language but textbooks and other study material are also available in state's local language.

NOTES

2. Central level boards

At the central level, there are two boards in India. One of them is the Central Board of Secondary Education (CBSE). The other all India board is the Council for Indian School Certificate Examination (CISCE, popularly called ISCE).

(i) CBSE

The CBSE was constituted in 1962. It has been set up to achieve the following interlinked objectives:

- To prescribe conditions of examinations and conduct public examination at the end of Class X and XII
- To grant qualifying certificates to successful candidates of the affiliated schools
- To fulfil the educational requirements of those students whose parents were employed in transferable jobs
- To prescribe and update the course of instructions of examinations
- To affiliate institutions for the purpose of examination and raise the academic standards of the country

The prime focus of the Board is on:

- Innovations in teaching-learning methodologies by devising students friendly and students centred paradigms
- Reforms in examinations and evaluation practices
- Skill learning by adding job-oriented and job-linked inputs
- Regularly updating the pedagogical skills of the teachers and administrators by conducting in service training programmes, workshops, etc

The jurisdiction of the Board is extensive and stretches beyond the national geographical boundaries. As a result of the reconstitution, the erstwhile 'Delhi Board of Secondary Education' was merged with the Central Board and thus all the educational institutions recognized by the Delhi Board also became a part of the Central Board. Subsequently, all the schools located in the Union Territory of Chandigarh. Andaman and Nicobar Island, Arunachal Pradesh, the state of Sikkim, and now Jharkhand, Uttaranchal and Chhattisgarh have also got affiliation with the Board. From 309 schools in 1962 the Board today has 8979 schools, including 141 schools in 21 countries. There are 897 Kendriya Vidyalayas, 1761 Government Schools, 5827 Independent Schools, 480 Jawahar Novodaya Vidyalayas and 14 Central Tibetean Schools.

In order to execute its functions effectively Regional Offices have been set up by the Board in different parts of the country to be more responsive to the affiliated schools. The Board has regional offices in Allahabad, Ajmer, Chennai, Guwahati, Panchkula and Delhi. Schools located outside India are looked after by regional office Delhi. The headquarters constantly monitors the activities of the

Regional Offices, although sufficient powers have been vested with the Regional Offices. Issues involving policy matters are, however, referred to the head office. Matters pertaining to day-to-day administration, liaison with schools, pre and post examination arrangements are all dealt with by the respective regional offices.

NOTES

(ii) CICSE

The Council for Indian School Certificate Examination (CICSF) has been constituted to secure suitable representation of governments responsible for schools (which are affiliated to it) in their states/territories; the Inter-State Board for Anglo-Indian Education; the Association of Indian Universities; the Association of Head of Anglo-Indian Schools, the Indian Public School Conference; the Association of Schools for the ISC Examination and eminent educationists.

The objects of the Council is educational, and includes the promotion of science, literature, the fine arts and the diffusion of useful knowledge by conducting school examination through the medium of English. The Council exists solely for educational purposes and not for purposes of profit. It is committed to serving the nation's children, through high quality educational endeavours, empowering them to contribute towards a humane, just and pluralistic society, promoting introspective living, by creating exciting learning opportunities, with a commitment to excellence.

National Institute of Open Schooling (NIOS)

NIOS is an 'Open School' to cater to the needs of a heterogeneous group of learners up to pre-degree level. It was started as a project with in-built flexibilities by the Central Board of Secondary Education (CBSE) in 1979. In 1986, the National Policy on Education suggested strengthening of Open School System for extending open learning facilities in a phased manner at secondary level all over the country as an independent system with its own curriculum and examination leading to certification. Consequently, the Ministry of Human Resource Development (MHRD), Government of India, set up the National Open School (NOS) in November 1989. The pilot project of CBSE on Open School was amalgamated with NOS. The National Open School (NOS) was vested with the authority to register, examine and certify students registered with it up to pre-degree level courses. In July 2002, the Ministry of Human Resource Development amended the nomenclature of the organization from the National Open School (NOS) to the National Institute of Open Schooling (NIOS) with a mission to provide relevant continuing education at school stage, up to pre-degree level through Open Learning system to prioritized client groups as an alternative to formal system, in pursuance of the normative national policy documents and in response to the need assessments of the people, and through it to make its share of contribution to:

- Universalization of education
- Greater equity and justice in society
- Evolution of a learning society

The NIOS provides opportunities to interested learners by making available the following Courses/Programmes of Study through open and distance learning (ODL) mode.

NOTES

- Open Basic Education (OBE) Programme for 14+ years age group, adolescents and adults at A, B and C levels that are equivalent to classes III, V and VIII of the formal school system.
- Secondary Education Course
- Senior Secondary Education Course
- Vocational Education Courses/Programmes
- Life Enrichment Programmes

3.4.1 Examination Reforms

NCERT's General Guidelines for Evaluation

A publication of the NCERT entitled *Reforming Examinations: Some Emerging Concepts* (1978) lists the following generalisations about evaluations:

1. Evaluation is a function of the learner and instruction and, therefore, good evaluation is one which is done by the teacher, of the taught as an individual.
2. Evaluation provides quality control at every stage of the teaching learning process and, therefore, evaluation would be treated as an integral part of the teaching-learning process.
3. Since evaluation provides feedback about the rate of the pupil's learning and the effectiveness of instruction, evaluation should be done unitwise after teaching every unit.
4. As the purpose of teaching is learning by students, focus of teachers' evaluation should be on improvement of pupils' achievement and not on judging their achievement. Therefore, diagnostic testing and remedial teaching should go side by side.
5. Pupils' achievement is the outcome of the integrated process of learning within a given set of conditions. Evaluation of pupils' learning should, therefore, be also integrated with regard to both the process and product of learning,
6. Keeping in view the obsolesence and explosion of knowledge, the new curriculum stresses the learnability aspect more than the knowledge aspect. Emphasis in evaluation should therefore, shift from testing of rote memory to that of problem solving abilities and attitude development.
7. For appraisal of the total development of the learner, it is essential that evaluation should not be limited to scholastic achievement alone, but it should encompass all aspects of pupils development. As such evaluation techniques will have to be extended beyond written and practical examinations to include oral testing, observations, checklists, rating scales and interviews.
8. Since independent learning by students is considered an important method of learning in the new curriculum self-assessment by pupils of their own learning should be practised in the evaluation system of an institution, so that cooperative assessment of the teacher and the learner is encouraged.

NOTES

9. Since it is impossible to achieve 100 percent or even near to 100 percent reliability of the various tools used for evaluating pupils, it is desirable that students should be classified broadly into 5 to 7 grades rather than using 101 point scale as at
10. As every pupil learns at his own rate he should be judged in terms of his own capacities and goals and not in terms of the standards of his class, institution or the board of secondary education. As such passing of a student in all the subjects at a time cannot be considered essential.
11. The grades of every pupil indicate only his level of performance which may be satisfactory or unsatisfactory in terms of his own standard. Thus, a grade howsoever low it may be, cannot be taken as failure but as an indicator of his present level of achievement.
12. Given more time and proper remedial teaching a student can improve his achievement Therefore, a student should get the opportunity of improving his grade in one or more subjects, if fie so desires.
13. The more accurately and meaningfully the evidence about pupils' growth in different aspects of his development is reported to students, teachers, parents, employers and institutes recording of pupil performance in various areas of development is a prerequisit to every evaluation programme.

Evaluation of Non-Scholastic Aspects of Pupil's Growth

A national seminar organised by the NCERT from 8 to 12 October, 1979 made the following recommendations on making effective evaluation of Non-scholastic aspects of pupil's growth.

1. Educationists the world over are expressing a grave concern over the bookish nature of the present day education dominated by examinations which tends to evaluate mostly the cognitive outcomes of education only. This is particularly true of the Indian scene. The seminar strongly recommends that educational policy-makers, planners, administrators and teachers should take a serious note of this lacuna and take proper steps to ensure that adequate attention is paid to the development and evaluation of both the cognitive and non-cognitive aspects of pupil growth. This is possible only when the schools pay adequate attention to the development and scientific evaluation of both these aspects in an integrated and balanced manner.
2. Since the foundations of personality are laid in early childhood, it is very important that developmental evaluation of non-cognitive aspects of growth should start as early as possible. The Education Departments and the Boards of Secondary Education should ensure that proper facilities are provided in the educational institutions so as to make possible the development and evaluation of these aspects of pupil growth. For this purpose, it would be necessary to develop systematic and realistic programmes for various types of institutions and to ensure that they are sincerely executed. The importance of cocurricular activities in the fostering of non-scholastic objectives of

NOTES

education is too well- known to need reiteration by this seminar. It is, therefore, recommended that at least six periods per week must be provided for cocurricular activities in all secondary and higher secondary schools. The proportion could be even large in primary schools. The planning and execution of these activities need to be done in a manner that participation of each and every pupil in some worthwhile activity is duly made possible.

3. Appreciating the role of education in shaping the personality of the child the seminar recommends that evaluation in schools should be used more for developmental purposes than for simply reporting, promoting from class-to-class and certification. This is particularly true to the non-scholastic aspects of growth, scientific assessment of which poses serious theoretical and practical problems. The seminar, therefore, recommends that the evaluation done in these areas should be used for formative: purposes only and should not be quantified to be used for promotion from class-to-class.
4. The seminar considered in detail the role of class-room instructional programmes in the development of non-scholastic aspects of growth. It was agreed that in each subject area specific objectives of the psychomotor and effective domains should also be identified and included in the evaluation scheme. Similar exercise may be done in the cocurricular area as well. In addition, programmes conducive to the development of these objectives should also be identified and executed in both the areas in a systematic manner.
5. Training of teachers in the development and evaluation of non-scholastic aspects of pupil growth is very crucial, more so because of the elusive and non-tangible nature of these aspects. The Departments of Education, Boards of Secondary Education and the Universities should undertake in-service programmes on an extensive scale and also, incorporate to relevant topics and activities in the preservice programmes. Availability of adequate facilities for training in this area should be made an essential condition for recognition of training institutions at all levels.
6. The role of the supervisor in the proper execution of educational programmes in India cannot be ignored. For successful implementation of any innovative project, involvement of administrators and the head-masters is a must and therefore orientation of these categories of staff in the supervision of work in the non-scholastic aspects of growth should receive proper emphasis in their training and orientation programmes.
7. The present state of affairs in education in India is an inevitable outcome of quantitative expansion at the cost of quality control. To restore education to its pristine glory it is necessary that serious notice is taken by the Central and the State Governments and the local communities of the existing inadequate facilities in most of the schools and ernest efforts are made to improve them early so as to make possible a satisfactory implementation of educational programmes. The wide gap between the stated objectives of education and their actual implementation in schools for want of adequate facilities need to be appreciated early enough and remedial measures taken accordingly.

NOTES

8. Instructional materials and manuals make a significant contribution to the successful implementation of educational programmes. There is a great need for the development of such materials in the forms of teacher's guides, students' guides etc. and their circulation to schools. The Department and the Boards of Secondary Education should get such materials developed with the help of Teachers colleges and other experts in adequate quantities for the guidance of teachers and students. The role of mass media in this area need also be recognized.
9. To give due importance to the development and evaluation of non-scholastic aspects of growth in the school time-table, it is necessary that the periods devoted by teachers to this work should be counted in their work-load.
10. For an effective implementation of any programme of development and evaluation of non-scholastic aspects of growth it is essential that good work done by teachers in this area should receive due recognition for which the Department and the Boards can undertake a variety of steps such as holding seminar reading programmes, issuing of special certificates of recognition to outstanding teachers and providing for the assessment of the work in this area in the confidential report forms of teachers.
11. The need for research and development activities in the non-scholastic area of pupil growth is too obvious to be dilated upon. The National Council of Educational Research and Training, the State Institutes of Education, Colleges and the Departments of Education in the Universities should appreciate this need and undertake Systematic and Scientific research in this area. A beginning in this respect has been made by the National Council of Educational Research and Training which needs further intensification at all levels.

Trends in Examinations and Evaluation

As a result of reform movement in examinations and evaluation, the following major trends are disconcernible:

- Use of more scientific and systematic methods instead of arbitrary methods.
- Stress on performance in both academic and non-academic areas
- Continuous evaluation in place of periodic examinations
- Use of a variety of techniques and tools instead of a few techniques and tools of evaluation
- Wider uses of test results in place limited uses
- Emphasis on improvement of achievement rather than measure-ment of achievement
- Treatment of testing in relation to other elements of the curriculum rather than treating testing in isolation
- Preferring grades to marks
- Awarding subjectwise grades in place of overall grades
- Clearing the examination in parts and not at one stroke

NOTES

- Improving grades through subsequent attempts
- Providing opportunities for re-evaluation
- Spot evaluation at specified centres in place of marking by examiners at their residences
- Deployment of special squads for checking unfair means
- Declaring the use of unfair means as an offence
- More flexible and purposeful form of question papers than the traditional one
- Apart from testing memorisation, questions include for evaluating abilities like understanding, critical thinking, application of knowledge and skills etc
- More coverage of the syllabus
- Inclusion of a variety of questions in the question paper
- A large number of questions in place of a few questions
- Replacement of overall options to internal options
- Specifically worded questions in place of large questions
- More objective scoring
- Use of multiple sets of question papers

Use of Question Bank

One of the main shortcomings of the present education system is the lack of appropriate questions asked in the examination. It does not meet the teaching-learning process properly. The person who prepares the questions basically depends on his choice and questioning skill. These questions are prepared on basic knowledge and many times it also has been observed that most of the questions are repeated; only language is changed. These questions are not able to evaluate knowledge, understanding or applicability domain of the students. As a result, students prepare only few questions for the examination, which does not give the complete evaluation of student's knowledge.

Keeping all these points in mind, the idea of preparing a question bank has been generated. A question bank is a set of questions on a subject used either for study review or for drawing questions used in an examination. It is readymade collection of the questions that helps students and teachers in teaching learning process. It is prepared for critical evaluation. It is a means of determining the presence, quality and criticality of the subject. Thus, it can be said that the question bank is a database of questions that can be shared among various courses. How questions meet specific criteria to create assessments is important.

A good question bank:

- Tests student understanding of a subject
- Guides the student to learn a lesson deeply
- Develops questioning skill among students and higher level thinking

In a question bank, a number of questions are prepared from a single unit of a subject, maximum number of questions those can be prepared from a unit are prepared. There may be many types of question papers:

- Bank of objective type questions
- Bank of short answer questions
- Bank of essay type (descriptive) questions
- Bank of miscellaneous questions

These questions are prepared on the basis of difficulty level and discrimination power to make these question banks more reliable and logical.

To prepare a question bank, you must:

1. **Ensure that there are all types of questions-** Objective type questions, short-answer questions, long-answer questions, questions requiring illustrations, diagrams etc. The main thing is that the question bank must be exhaustive as far as the content coverage is concerned
2. **Ensure that there are all level of questions-** There must be all levels of questions in a question bank. These levels pertain to complexity; from very simple to very complex questions, in order to prepare the student fully for examination
3. **Ensure questions from previous years' question papers-** Sometimes, examiners setting the papers use cryptic language to form questions to confuse students. Students should know how to interpret the cryptic questions and answer correctly. This can be achieved by offering them such questions from old exam papers.
4. **Provide answers at the end-** A good question bank always contains answers at the end. Otherwise there is no way for the student to confirm if he/she has answered correctly. For mathematical problems or Physics numerical questions, there need not be the entire solution, but just the final answer.

Differential Vs Uniform Evaluation

In evaluation situations, we tend to overlook the fact that there are individual differences among students. We feel satisfied with using a uniform yardstick for all students at a particular point of time. We do not usually take cognizance of the fact that the rate of learning of different children may be different, the learning situations may vary from class to class and student to student and resource inputs may not be the same. Evaluation has to take care of and cater to the needs of such extensive and intensive differences and promote self-paced learning by providing for individual differences instead of prescribing uniform group evaluation.

Full Vs Partial Evaluation

Full evaluation must take into account all the three aspects i.e. cognitive, affective and psycho-motor. The traditional system of examination evaluates only cognitive abilities. Even in cognitive area external examinations often to do not go beyond the

evaluation of memorisation or convergent thinking. These drawbacks can be overcome through evaluation at the school level.

3.4.2 Open Book Examination

NOTES

While developing assessment tools for classroom, certain issues should be addressed in relation to the following:

- **Purpose and impact:** How should the assessment be used and how should it impact the instruction and the selection of curriculum?
- **Validity and fairness:** Does it calculate what it intends to calculate?
- **Reliability:** Is the data that is to be collected reliable across applications within the classroom, school and district?
- **Significance:** Does it address content and skills, which are to be valued by and reflect current thinking in the field?
- **Efficiency:** Is the method of assessment consistent with the time available in the classroom setting?

There is a huge range of assessments that are accessible for use in reforming science assessment in the classroom. These types of assessments comprise strategies that are traditional as well as alternative. The diverse types of alternative assessments can be used with a range of science content and process skills, including the following general targets:

- Declarative knowledge
- Conditional knowledge
- Procedural knowledge
- Application knowledge
- Problem solving
- Critical thinking
- Documentation
- Understanding

Assessment can be divided into three stages:

1. **Baseline assessment:** It establishes the 'starting point' of a student's understanding.
2. **Formative assessment:** It provides information to help guide the instruction throughout the unit.
3. **Summative assessment:** It informs both the student and the teacher about the level of conceptual understanding and performance capabilities that a student achieves.

The huge range of targets and skills that can be addressed in classroom assessment needs the use of many assessment formats. Some formats, and the stages of assessment in which they most likely occur are shown in Table 3.3.

NOTES

Table 3.3 *Assessment Formats*

Format	Nature/Purpose	Stage
Baseline assessments	Oral and written responses based on individual experience Assess prior knowledge	Baseline
Paper and pencil tests	Multiple choice, short answer, essay, constructed response, written reports Assess students acquisition of knowledge and concepts	Formative
Embedded assessments	Assess an aspect of student learning in the context of the learning experience	Formative
Oral reports	Require communication by the student that demonstrates scientific understanding	Formative
Interviews	Assess individual and group performance before, during and after a science experience	Formative
Performance tasks	Require students to create or take an action related to a problem, issue or scientific concept	Formative and summative
Checklists	Monitor and record anecdotal information	Formative and summative
Investigative projects	Require students to explore a problem or concern stated either by the teacher or the students	Summative
Extended or unit projects	Require the application of knowledge and skills in an open-ended setting	Summative
Portfolios	Assist students in the process of developing and reflecting on a purposeful collection of student-generated data	Formative and summative

Oral exams

An oral exam is a chance for one to demonstrate their knowledge, their presentation/speaking skills as well as their ability to communicate. These exams can be formal or informal, but all exams should be considered formal exchanges for making a good impression. For both types, you should listen carefully to the question and then answer directly.

NOTES

Formal exams have a list of questions in a ready format. The criteria for evaluation are usually set in a right/wrong format, and can at times be competitive. For this kind of exam, if you wish to add 'related' or qualified information, first ask permission as a basic courtesy.

Informal exams are mostly more open and your responses can be usually longer and the evaluations can be more subjective. Answers are generally less exact (right/wrong) and more value is added for problem solving analysis and method, as well as interpersonal communication and presentation.

Written tests

Written tests are those tests that are given on paper or on a computer. A test taker who takes a written test might respond to precise items by writing or typing inside a given space of the test or on a separate form or document.

A test developer's choice of what style or format to use while developing a written test is generally random agreed that there is no solitary invariant standard for testing. As a result, these tests might comprise only one test item format or might have a combination of different test item formats.

Some common written test formats are:

- **MCQs or Multiple Choice Questions-** A statement or question is given with five or six options and the respondent has to choose the correct answer or answers, depending on the correct answers included. In general, besides the correct answer/s, at least three incorrect answers are included to confuse the respondent.
- **Fill-in-the-blanks-** A statement is given with some key words missing and the respondent needs to provide the missing words.
- **Short-answer questions-** The respondent is asked to write brief answers, which may include a definition, a short description, an example, a list, or to draw a figure and so on.
- **Long-answer questions-** The respondent is asked to write descriptive, essay type answers that require him or her to analyse and apply the concepts learned.

Open book examination

In an open book exam you are assessed on understanding instead on recall as well as memorization.

You will be expected to do the following:

- Apply material to new situations
- Analyse elements and relationships
- Synthesize or structure
- Evaluate using your material as evidence
- Access to content varies by instructor

The exam can be taken home or in the classroom with questions that are seen or unseen before exam time. You should not underestimate the preparation

needed for an open book exam and your time will be limited. So the key is proper organization in order to quickly find data, quotes, examples and/or arguments that you use in your answers.

NOTES

Preparation

- Keep current on readings and assignments in class
- Prepare brief and concise notes on ideas and concepts that are being tested
- Select carefully what you intend to bring with you to the exam, and note anything significant about what you do not
- Include your own comments on the information that will provide value to your arguments, and show that you have considered this through
- Anticipate with model questions but not model answers
- Instead challenge yourself with how would you answer questions and what options or resources you might need to think.

Weightage in content

Table 3.4 points towards the diverse features of the content that are to be tested and the weightage that is to be given to these aspects.

Table 3.4 *Features of the Content that is to be Tested and Weighed*

Sl. No	Content	Marks	Percentage
1	Sub topic - 1	15	60
2	Sub topic - 2	10	40
Total		**25**	**100**

Reasons for stating objectives and aims

The statement of educational objectives and aims has numerous advantages. Some of them are as follows:

- To assist teachers in planning the curriculum
- To convey the educational intent of the course to students and to the faculty
- To help recognize the resources needed to undertake the instruction process
- To provide a basis for evaluating the course, and a basis for the assurance of quality

Aims, objectives and learning outcomes

All educational objectives can be understood in the following ways:

1. What the teacher intends to do?
2. What does it intend the student to learn, or what he or she will be able to learn by gaining learning experience?

NOTES

Earlier, such purposes have often been defined in terms of the teacher's activity. This no longer suffices because the aims of teaching are required to be defined in terms of the ultimate purpose, which is student learning.

It is, therefore, a consensus that for each teaching activity, there should be two kinds of statement of intent:

Objectives (or learning outcomes as intended)

A number of specific statements explicating what it intends for the student to have learnt or be able to do, as a result of the learning experience.

Hierarchy of aims and objectives

There are many situations in which aims and objectives can be spelt out. The success and attainment of higher level aims will be dependent on the achievement of lower level aims. The important question to ask is: 'How will the aim and objective for this individual teaching session help achieve the overall aims and objectives of the course?'

Allocation of time

While group assignments help teachers save time in some areas (e.g., grading final projects), they may lead to use time in other areas (e.g., time required up front to reckon suitable project topics, interact with external clients, divide students in groups; time during the semester to meet with and regulate student groups; time at the end of the semester to determine the contributions made by individual team members.)

Exam blueprint

Once you understand the learning objectives and question types for your exam, you should compose an exam blueprint. An exam blueprint may be structured in terms of a chart that represents the number of questions that you think is desirable in your exam within each topic and objective level. The blueprint states the learning objectives and skills that the learners are to be tested for and the relative importance that must be given to each. The blueprint also ensures that you cover the desired topics for your assessment. Table 3.5 shows an example blueprint.

After you have composed your blueprint, start by writing questions that correspond with the learning objective level for each topic area.

A blueprint specifies what the test should cover rather than describing what the course syllabus covers. A test blueprint should have the test title, the fundamental purpose of the test, the features of the curriculum that are covered by the test, an indicator of the students who will take the test, the kinds of tasks that will be used in the test (and how these tasks will correspond with other pertinent evidence to be collected), the uses of the evidence provided by the test, the conditions under which the test will be given (time, place, who will prepare the test, who will score the answer sheets, how accuracy of scoring will be ensured, whether students will be able to consult books (or use calculators) while attempting the test, and any precautions that are to be taken so as to ensure that the responses

are only the work of the student attempting the test), and the balance of the questions.

Table 3.5 *Exam Blueprint*

	Topic I	Topic II	Topic III	Topic IV	Total
Knowledge	1	2	1	1	5 (12.5%)
Comprehension	2	1	2	2	7 (17.5%)
Application	4	4	3	4	15 (37.5%)
Analysis	3	2	3	2	10 (25%)
Synthesis		1		1	2 (5%)
Evaluation			1		1 (2.5%)
Total	10 (25%)	10 (25%)	10 (25%)	10 (25%)	40

NOTES

3.4.3 Semester System

The semester system is basically a division of an academic year in two parts in which an educational institution aims to complete the predefined course. Examinations are conducted after the completion of one semester. The second semester is totally independent of the first.

Characteristics of semester system

The characteristics of the semester system may be summarized as follows:

- A semester system divides the whole academic year into two terms, approximately 16 to 18 weeks each.
- It is a very proactive system, engaging both the faculty and the students in one or the other academic activity throughout the year.
- It does not allow the students to feel free most of the time as it does not allow them to study only when examination is just round the corner as is the case in the annual examination system.
- It keeps the students engaged throughout the year. However, it reduces examination pressure also by dividing the year syllabus into parts.
- It assesses the students at short intervals and hence enhances their learning skills by constantly informing them about their weaknesses.

NOTES

Limitations

The limitations of the semester system are as follows:

- Sometimes it may lead to create confusion among the students.
- Teachers may face the problem of increased frequency of examinations and extended work load.
- Sometimes it may create a confusion among the students as to how things are going to happen.

CHECK YOUR PROGRESS

6. What is NIOS?
7. What are the characteristics of a good question bank?
8. What are the three stages of assessment?

3.5 SUMMARY

- Intelligence of an individual or group of individuals is tested through various types of tests which are called as intelligence test.
- Intelligence test can be studied by classifying on the basis of administration in the following ways:
  (i) Individual intelligence test
  (ii) Group intelligence test
- Intelligence test can be studied by classifying on the basis of the nature of the test
  (i) Verbal intelligence test
  (ii) Non-verbal intelligence test
  (iii) Performance intelligence test
- There are many tests for measuring intelligence. They can be of three types namely verbal, non-verbal and performance types.
- Some important tests of intelligence are given as follows:
  - Stanford–Binet test of intelligence
  - Wechsler–Belleve intelligence test
  - Thurstone's test of primary mental abilities
  - Non-verbal test of intelligence (also known as Reven's Progressive Matrices)
  - Culture free intelligence test
  - Bhatia–battery performance test
- L.L. Thurstone states that an attitude 'is the degree of positive and negative affect associated with some psychological object'.

NOTES

- Attitude may be classified into different categories like:
  - o Acquisitive attitude
  - o Play attitude
  - o Scientific attitude
  - o Business attitude
  - o Artistic attitude
  - o Religious attitude
- The determinants of attitude are:
  - o Cultural or social determinant
  - o Psychological determinants
  - o Functional determinants
- The methods to be followed for the measurement of attitude are:
  - (i) Thurstone Scale
  - (ii) Likert Scale
- According to G.W. Allport, 'personality is the dynamic organization within individual of those psycho-physical systems that determine his unique adjustment to his environment'.
- The term 'project' was used for the first time by Sigmund Freud in the field of psychology. Projection, according to Freud means externalizing of conflicts or other internal conditions that give rise to conscious pain and anxiety.
- Several methods of assessment and evaluation of learners are in use these days, such as unit tests, half-yearly examinations, annual examinations, semester system, board examinations and entrance tests.

3.6 KEY TERMS

- **Attitude:** It refers to a predisposition or a tendency to respond positively or negatively towards a certain idea, object, person, or situation
- **Personality:** It refers to the combination of characteristics or qualities that form an individual's distinctive character.
- **Projection:** According to Freud, projection means the externalizing of conflicts or other internal conditions that give rise to conscious pain and anxiety.
- **Question bank:** It refers to a set of questions on a subject used either for study review or for drawing questions used in examination.

3.7 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The classification of intelligence tests on the basis of administration is as follows:
   - (i) Individual intelligence tests
   - (ii) Group intelligence tests

NOTES

2. In a performance intelligence test, the subject is required to handle or manipulate objects in such a way as to complete a specified task.
3. L.L. Thurstone states that an attitude is the degree of positive and negative affect associated with some psychological object.
4. Thurstone's attitude scale is known as equal-appearing interval scale. In this scale, both favourable and unfavourable statements regarding a particular topic are reflected in an eleven point scale. The respondent is supposed to check the point for each item according to his attitude. The median of the judged locations for an item is its scale value. The scale position are very much a function of the judges who are chosen.
5. The Rorschach inkblot test was developed by Swiss Psychiatrist Herman Rorschach in 1921. The test consists of 10 symmetrical ink blots on 10 separate cards of 11 × 9 inches. Five of the ink blots are black and white, and others are multi-coloured. All the cards are given to the respondent at a time in a sequence. After this, the examiners generally conduct inquiry into the details of the subject's response on individual cards. For the scoring purpose, the responses are given specific symbols and entry is made into four columns: location, contents, originality, and determinants.
6. NIOS is an 'Open School' to cater to the needs of a heterogeneous group of learners up to pre-degree level. It was started as a project with in-built flexibilities by the Central Board of Secondary Education (CBSE) in 1979.
7. A good question bank:
   - Tests student understanding of a subject
   - Guides the student to learn a lesson deeply
   - Develops questioning skill among students and higher lever thinking
8. Assessment can be divided into three stages:
   1. Baseline assessment: It establishes the 'starting point' of a student's understanding.
   2. Formative assessment: It provides information to help guide the instruction throughout the unit.
   3. Summative assessment: It informs both the student and the teacher about the level of conceptual understanding and performance capabilities that a student achieves.

3.8 QUESTIONS AND EXERCISES

Short-Answer Questions

1. Differentiate between individual intelligence test and group intelligence test.
2. Discuss how attitude can be measured.
3. What are some of the methods that are in use to evaluate learners?
4. What are open-book examinations? How can one prepare for them?
5. What is a semester system? Discuss its characteristics.

Long-Answer Questions

1. What is attitude? Discuss its characteristics and types.
2. What is personality? What are its characteristics? Discuss its determinants.
3. Discuss some of the projective techniques of examining personality.
4. Discuss some of the major trends in examinations and evaluations seen in recent times.

NOTES

3.9 FURTHER READING

Bloom, B. S. *Taxonomy of Objectives of Cognitive Domains*. New York: Longmans Green &Co.

Ebel, R.L. *Essentials of Educational Measurement.* New Jersey: Englewood Cliff.

Grunlund, N. E. *Measurement and Evaluation in Teaching* (3rd Ed). New York: Mcmillan Publishing Co.

Julian, C., Stanley and Hopkins. *Educational and Psychological Measurement and Evaluation*. New Delhi: Prentice Hall of India Pvt. Ltd.

Lindermon, Richard, H. *Educational Measurement*. Mumbai: B. S. Taraporevala Sons & Co.

NOTES

UNIT 4 STATISTICS IN MEASUREMENT AND EVALUATION-I

Structure

4.0 INTRODUCTION

In the previous unit, you learnt about psychological testing. In this unit, the discussion will turn towards statistics in measurement and evaluation. Statistics is science of collecting and analysing numerical data in large quantities, particularly for the purpose of inferring proportions in a whole from those in a representative sample.

One can analyse quantitative data through techniques such as measures of central tendency. Measures of central tendency are of various types, such as arithmetic mean, mode and median. This is also commonly known as simply the mean. Even though average, in general, means any measure of central location, when we use the word average in our daily routine, we always mean the arithmetic average.

4.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Discuss how data is interpreted
- Describe how data is represented graphically

- Discuss the different measures of central tendency
- Illustrate how mean, mode, and median can be calculated
- Explain correlation analysis
- Describe normal probability curve

NOTES

4.2 STATISTICAL TREATMENT OF DATA

Meaning, importance and steps involved in processing data

Research does not merely consist of data that is collected. Research is incomplete without proper analysis of the collected data. Processing of data involves analysis and manipulation of the collected data by performing various functions. The data has to be processed in accordance with the outline laid down at the time of developing the research plan. Processing of data is essential for ensuring that all relevant data has been collected to perform comparisons and analyses. The functions that can be performed on data are as follows:

- Editing
- Coding
- Tabulation
- Classification

Usually, experts are of the opinion that the exercise of processing and analysing of data is inter-related. Therefore, the two should be thought as one and the same thing. It is argued that analysis of data generally involves a number of closely-related operations, which are carried out with the objective of summarizing the collected data and organizing it in such a way that they are able to answer the research questions associated with it.

However, in technical terms, processing of data involves data representation in a way that it is open to analysis. Similarly, the analysis of data is defined as the computation of certain measures along with searching for the patterns of relationship that may exist among data groups.

Editing of data

Editing of data involves the testing of data collection instruments in order to ensure maximum accuracy. This includes checking the legibility, consistency and completeness of the data. The editing process aims at avoiding equivocation and ambiguity. The collected raw data is also examined to detect errors and omissions, if any. A careful scrutiny is performed on the completed questionnaires and schedules to assure that the data has the following features:

- Accuracy
- Consistency
- Unity
- Uniformity
- Effective arrangement

NOTES

The stages at which editing should be performed can be classified as follows:

- **Field Editing**: This involves reviewing the reporting forms, by the investigator, that are written in an abbreviated or illegible form by the informant at the time of recording the respondent's responses. Such type of editing must be done immediately after the interview. If performed after some time, such editing becomes complicated for the researcher, as it is difficult to decipher any particular individual's writing style. The investigator needs to be careful while field editing and restrain the researcher from correcting errors or omission by guesswork.
- **Central Editing**: This kind of editing involves a thorough editing of the entire data by a single editor or a team of editors. It takes place when all the schedules created according to the research plan have been completed and returned to the researcher. Editors correct the errors such as data recorded in the wrong place or the data recorded in months when it should be recorded in weeks. They can provide an appropriate answer to incorrect or missing replies by reviewing the other information in the schedule. At times, the respondent can be contacted for clarification. In some cases, if the answer is inappropriate or incomplete and an accurate answer cannot be determined on any basis, then the editor should delete or remove that answer from the collected data. He/She can put a note as 'no answer' in this case. The answers that can be easily deciphered as wrong should be dropped from the final results.

Besides using the above-stated methods according to the data source, the researcher should also keep in mind the following points while editing:

- Familiarity with the instructions given to interviewers and coders
- Know-how of editing instructions
- Single line striking for deleting of an original entry
- Standardized and distinctive editing of data
- Initialization of all answers that are changed

Coding of data

Coding of data can be defined as representing the data symbolically using some predefined rules. Once data is coded and summarized, the researcher can analyse it and relationships can be found among its various categories.

Checklist for coding

This enables the researcher to classify the responses of the individuals according to a limited number of categories or classes. Such classes should possess the following important characteristics:

- Classes should be appropriate and in accordance to the research problem under consideration.
- They must include a class for every data element.
- There should be a mutual exclusivity, which means that a specific answer can be placed in one and only one cell of a given category set.

- The classes should be one-dimensional. This means that every class is defined in terms of only one concept.

Significance of Coding

NOTES

Coding of data is necessary for its efficient analysis. Codiing facilitates reduction of data from a variety to a small number of classes. Thus, only that information which is important and critical for analysis is retained in the research. Coding decisions are usually taken at the designing stage of the questionnaire. This makes it possible to pre-code the questionnaire choices, which in turn, is helpful for computer tabulation.

However, in case of hand coding, some standard method should be used. One such method is to code in the margin with a coloured pencil. The other method is to transcribe data from the questionnaire to a coding sheet. Whatever method is adopted, you should ensure that coding errors are altogether eliminated or reduced to a minimum level.

Classification of Data

Research studies involve extensive collection of raw data and usage of the data to implement the research plan. To make the research plan easier, the data needs to be classified in different groups for understanding the relationship among the different phases of the research plan. Classification of data involves arrangement of data in groups or classes on the basis of some common characteristics. The methods of classification can be divided under the following two headings:

- Classification according to attributes
- Classification according to class intervals

Figure 4.1 shows the categories of data.

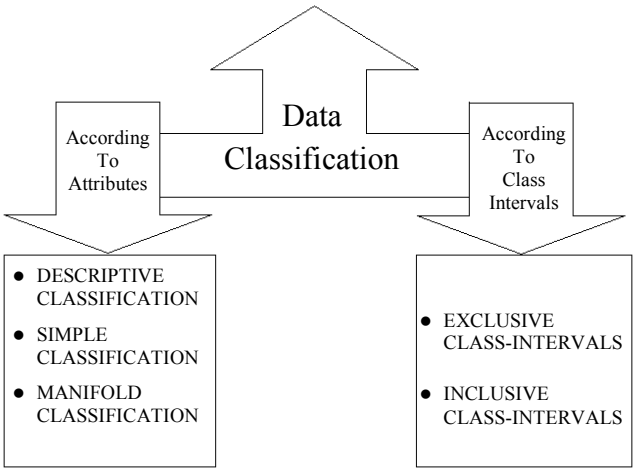


Fig. 4.1 *Data Classification*

Classification of data according to attributes

Data is classified on the basis of similar features as follows:

- **Descriptive classification:** This classification is performed according to the qualitative features and attributes which cannot be measured quantitatively.

NOTES

These features can be either present or absent in an individual or an element. The features related to descriptive classification of attributes can be literacy, sex, honesty, solidarity, etc.

- **Simple classification:** In this classification the elements of data are categorized on the basis of those that possess the concerned attribute and those that do not.
- **Manifold classification:** In this classification two or more attributes are considered simultaneously and the data is categorized into a number of classes on the basis of those attributes. The total number of classes of final order is given by 2^n, where n = number of attributes considered.

Classification of data according to class intervals

Classifying data according to the class intervals is a quantitative phenomenon. Class intervals help categorize the data with similar numerical characteristics, such as income, production, age, weight, etc. Data can be measured through some statistical tools like mean, mode, median, etc. The different categories of data according to class intervals are as follows:

- **Statistics of variables:** This term refers to the measurable attributes, as these typically vary over time or between individuals. The variables can be discrete, i.e., taking values from a countable or finite set, continuous, i.e., having a continuous distribution function, or neither. This concept of variable is widely utilized in the social, natural and medical sciences.
- **Class intervals:** They refer to a range of values of a variable. This interval is used to break up the scale of the variable in order to tabulate the frequency distribution of a sample. A suitable example of such data classification can be given by means of categorizing the birth rate of a country. In this case, babies aged zero to one year will form a group; those aged two to five years will form another group, and so on. The entire data is thus categorized into several numbers of groups or classes or in other words, class intervals. Each class interval has an upper limit as well as a lower limit, which is defined as 'the class limit.' The difference between two class limits is known as class magnitude. Classes can have equal or unequal class magnitudes.

  The number of elements, which come under a given class, is called the frequency of the given class interval. All class intervals, with their respective frequencies, are taken together and described in a tabular form called the frequency distribution.

Problems related to classification of data

The problems related to classification of data on the basis of class intervals are divided into the following three categories:

(i) **Number of classes and their magnitude:** There are differences regarding the number of classes into which data can be classified. As such, there are no pre-defined rules for the classification of data. It all depends upon the skill

NOTES

and experience of the researcher. The researcher should display the data in such a way that it should be clear and meaningful to the analyst.

As regards the magnitude of classes, it is usually held that class intervals should be of equal magnitude, but in some cases unequal magnitudes may result in a better classification. It is the researcher's objective and judgement that plays a significant role in this regard. In general, multiples of two, five and ten are preferred while determining class magnitudes. H.A. Sturges suggested the following formula to determine the size of class interval:

$$i = R/(1 + 3.3 \log N)$$

where,

i = size of class interval.

R = Range (difference between the values of the largest element and smallest element among the given elements).

N = Number of items to be grouped.

Sometimes, data may contain one or two or very few elements with very high or very low values. In such cases, the researcher can use an open-ended interval in the overall frequency distribution. Such intervals can be expressed below two years; or twelve years and above. However, such intervals are not desirable, yet cannot be avoided.

(ii) **Choice of class limits:** While choosing class limits, the researcher must determine the mid-point of a class interval. A mid-point is, generally, derived by taking the sum of the upper and lower limit of a class and then dividing it by two. The actual average of elements of that class interval should remain as close to each other as possible. In accordance with this principle, the class limits should be located at multiples of two, five, ten, twenty and hundred and such other figures. The class limits can generally be stated in any of the following forms:

- **Exclusive type class intervals**: These intervals are usually stated as follows:
  - 10–20
  - 20–30
  - 30–40
  - 40–50

  These intervals should be read in the following way:
  - 10 and under 20
  - 20 and under 30
  - 30 and under 40
  - 40 and under 50

  In the exclusive type of class intervals, the elements whose values are equal to the upper limit of a class are grouped in the next higher class. For example, an item whose value is exactly thirty would be put in 30–40-class interval and not in 20–30-class interval. In other words, an exclusive type of class interval is that in which the upper limit of a class interval is

excluded and items with values less than the upper limit, but not less than the lower limit, are put in the given class interval.

- **Inclusive Type Class Intervals**: These intervals are normally stated as follows:
  - 11–20
  - 21–30
  - 31–40
  - 41–50

  This should be read as follows:
  - 11 and under 21
  - 21 and under 31
  - 31 and under 41
  - 41 and under 51

  In this method, the upper limit of a class interval is also included in the concerning class interval. Thus, an element whose value is twenty will be put in 11–20-class interval. The stated upper limit of the class interval 11–20 is twenty but the real upper limit is 20.999999 and as such 11–20 class interval really means eleven and under twenty-one. When data to be classified happens to be a discrete one, then the inclusive type of classification should be applied. But when data happens to be a continuous one, the exclusive type of class intervals can be used.

(iii) **Determining the frequency of each class**: The frequency of each class can be determined using tally sheets or mechanical aids. In tally sheets, the class groups are written on a sheet of paper and for each item a stroke (a small vertical line) is marked against the class group in which it falls. The general practice is that after every four small vertical lines in a class group, the fifth line for the element falling in the same group is indicated as a diagonal line through the above said four lines. This enables the researcher to perform the counting of elements in each one of the class groups. Table 4.1 displays a hypothetical tally sheet.

Table 4.1 *A Tally Sheet*

Income groups (Rupees)	Tally mark	Number of families (Class frequency)
Below 600	𝍸 𝍸 𝍸 III	15
601-900	𝍸 𝍸 I	9
901-1300	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 I	25
1301-1500	𝍸 𝍸 𝍸 IIII	16
1501 and above	𝍸 𝍸 II	10
Total		75

In case of large inquiries and surveys, class frequencies can be determined by means of mechanical aids, i.e., with the help of machines. Such machines function, either manually or automatically and run on electricity. These machines can sort out

cards at a speed of around 25,000 cards per hour. Although this method increases the speed, it is an expensive method.

Tabulation of data

NOTES

In simple terms, tabulation means placing the results and data collected from research in a tabular form.

Methods of tabulation

Tabulation can be done either manually or mechanically using various electronic devices. Several factors like the size and type of study, cost considerations, time pressures and availability of tabulating machines decide the choice of tabulation. Relatively large data requires computer tabulation. Manual tabulation is preferred in case of small inquiries, when the number of questionnaires is small and they are of relatively short length. The different methods used in hand tabulation are as follows:

- **Direct tally method**: This method involves simple codes, which the researcher can use to directly tally data with the questionnaire. The codes are written on a sheet of paper called tally sheet and for each response, a stroke is marked against the code in which it falls. Usually, after every four strokes against a particular code, the fifth response is indicated by drawing a diagonal or horizontal line through the strokes. These groups are easy to count and the data is sorted against each code conveniently.
- **List and tally method**: In this method, code responses may be transcribed into a large worksheet, allowing a line for each questionnaire. This facilitates listing of a large number of questionnaires in one worksheet. Tallies are then made for each question.
- **Card sort method**: This is the most flexible hand tabulation method, where the data is recorded on special cards that are of convenient sizes and shapes and have a series of holes. Each hole in the card stands for a code. When the cards are stacked, a needle passes through a particular hole representing a particular code. These cards are then separated and counted. In this way, frequencies of various codes can be found out by the repetition of this technique.

Significance of tabulation

Tabulation enables the researcher to arrange data in a concise and logical order. It summarizes the raw data and displays the same in a compact form for further analysis. It helps in the orderly arrangement of data in rows and columns. The various advantages of tabulation of data are as follows:

- A table saves space and reduces descriptive and explanatory statements to the minimum.
- It facilitates and eases the comparison process.
- Summation of elements and detection of omissions and errors becomes easy in a tabular description.
- A table provides a basis for various statistical computations.

NOTES

Checklist for tables

A table should communicate the required information to the reader in such a way that it becomes easy for him/her to read, comprehend and recall information when required. Certain conventions have to be followed during tabulation of data. These are as follows:

- All tables should have a clear, precise and adequate title to make them intelligible enough without any reference to the text.
- Tables should be featured with clarity and readability.
- Every table should be given a distinct number to facilitate an easy reference.
- The table should be of an appropriate size and tally with the required information.
- Headings for columns and rows should be in bold font letters. It is a general rule to include an independent variable in the left column or the first row. The dependent variable is contained in the bottom row or the right column.
- Numbers should be displayed such that they are neat and readable.
- Explanatory footnotes, if any, regarding the table should be placed directly beneath the table, along with the reference symbols used in the table.
- The source of the table should be indicated just below the table.
- The table should contain thick lines to separate data under one class from the data under another class and thin lines to separate the different subdivisions of the classes.
- All column figures should be properly aligned.
- Abbreviations should be avoided in a table to the best possible extent.
- If data happens to be large, then it should not be crowded in a single table. It makes the table unwieldy and inconvenient.

Tabulation can also be classified as complex and simple. The former type of tabulation gives information about one or more groups of independent variables, whereas, the latter shows the division of data in two or more categories.

Use of statistical tools for analysis

A researcher needs to be familiar with different statistical methods so as to be able to use the appropriate method in his research study. There are certain basic statistical methods that can be classified into the following three groups:

- Descriptive statistics
- Inferential statistics
- Measures of central tendency and dispersion

Descriptive statistics

According to Smith, descriptive statistics is the formulation of rules and procedures where data can be placed in a useful and significant order. The foundation of applicability of descriptive statistics is the need for complete data presentation. The most important and general methods used in descriptive statistics are as follows:

- **Ratio**: This indicates the relative frequency of the various variables to one another.

NOTES

- **Percentage**: Percentages (%) can be derived by multiplying a ratio with 100. It is thus a ratio representing a standard unit of 100.
- **Frequency table**: It is a means to tabulate the rate of recurrence of data. Data arranged in such a manner is known as distribution. In case of a large distribution tendency, larger class intervals are used. This facilitates the researcher to acquire a more orderly system.
- **Histogram**: It is the graphical representation of a frequency distribution table. The main advantage of graphical representation of data in the form of histogram is that data can be interpreted immediately.
- **Frequency polygon**: It is used for the representation of data in the form of a polygon. In this method, a dot that represents the highest score is placed in the middle of the class interval. A frequency polygon is derived by linking these dots. An additional class is sometimes added in the end of the line with the purpose of creating an anchor.
- **Cumulative frequency curve**: The procedure of frequency involves adding frequency by starting from the bottom of the class interval, and adding class by class. This facilitates the representation of the number of persons that perform below the class interval. The researcher can derive a curve from the cumulative frequency tables with the purpose of reflecting data in a graphical manner.

Inferential statistics

Inferential statistics enable researchers to explore unknown data. Researchers can make deductions or statements using inferential statistics with regard to the broad population from which samples of known data has been drawn. These methods are called inferential or inductive statistics. These methods include the following common techniques:

- **Estimation**: It is the calculated approximation of a result, which is usable, even if the input data may be incomplete or uncertain. It involves deriving the approximate calculation of a quantity or a degree. For example, drawing an estimate of cost of a project; or deriving a rough idea of how long the project will take.
- **Prediction**: It is a statement or claim that a particular event will surely occur in future. It is based on observation, experience and a scientific reasoning of what will happen in the given circumstances or situations.
- **Hypothesis Testing**: Hypothesis is a proposed explanation whose validity can be tested. Hypothesis testing attempts to validate or disprove preconceived ideas. In creating hypothesis, one thinks of a possible explanation for a remarked behaviour. The hypothesis dictates that the data selected should be analysed for further interpretations.

There are also two chief statistical methods based on the tendency of data to cluster or scatter. These methods are known as measures of central tendency and measures of dispersion. You will learn about these in the subsequent section.

NOTES

4.2.1 Interpretation of Data

Analysis of data is the process of transformation of data for the purpose of extracting some useful information, which in turn facilitates the discovery of some useful conclusions. Finding conclusions from the analysed data is known as interpretation of data. However, if the analysis is done, in the case of experimental data or survey, then the value of the unknown parameters of the population and hypothesis testing is estimated.

Analysis of data can be either descriptive or inferential. Inferential analysis is also known as statistical analysis. The descriptive analysis is used to describe the basic features of the data in a study, such as persons, work groups and organizations. The inferential analysis is used to make inferences from the data, which means that we are trying to understand some process and make some possible predictions based on this understanding.

Types of analysis

The various types of analyses are as follows:

- **Multiple Regression Analysis**: This analysis is used to predict the single dependent variable by a set of independent variables. In multiple regression analysis, independent variables are not correlated.
- **Multiple Discriminant Analysis**: In multiple discriminant analysis, there is one single dependent variable, which is very difficult to measure. One of the main objectives of this type of analysis is to understand the group differences and predict the likelihood that an entity, i.e., an individual or an object belongs to a particular class or group based on several metric independent variables.
- **Canonical Analysis**: It is a method for assessing the relationship between variables. This analysis also allows you to investigate the relationship between the two sets of variables.

Univariate, Bivariate and Multivariate Analyse

There are also many types of analyses performed according to the variance that exists in data. They are carried out to check if the differences among three or more variables are so significant that data has to be evaluated statistically. There are three types of such analyses; namely, univariate, bivariate and multivariate analyses. These types are discussed as follows:

- **Univariate Analysis**: In this analysis, only a single variable is taken into consideration. It is usually the first activity pursued while analysing data. It is performed with the purpose of describing each variable in terms of mean, median or mode, and variability. Examples of such analysis are averages or a set of cases that may come under a specific category amidst a whole sample.
- **Bivariate Analysis**: This type of analysis examines the relationship between two variables. It tries to find the extent of association that exists between these variables. Thus, a bivariate analysis may help you, for example, to find whether the variables of irregular meals and migraine headaches are

NOTES

associated. It may also help to find the extent to which they may be associated. Here, two variables are thus statistically measured simultaneously.

- **Multivariate Analysis**: This type of analysis involves observation and analysis of three or more than three statistical variables at a time. Such an analysis is performed using statistical tests or even a tabular format. Thus, for example, using the multivariate analysis method, you can study the variables of age, educational qualification and annual income of a given set of population at the same time.

Usually, these types of analyses are more convenient when performed in a tabular format. Multivariate analysis involves, using a cross-classification or contingency table. Such a table is made of two columns and two rows, showing the frequencies of two variables that are displayed in rows and columns. This is more popularly known as constructing the bivariate table. Traditionally, the independent variable is displayed in columns and the dependent one in rows. A multivariate table, if related to the same data, is the result of combining the bivariate tables. In this case, each bivariate table is known as a partial table. Usually, a multivariate table is created with the purpose of explaining or replicating the primary relationship that is found in the bivariate table. Table 4.2(a) and (b) shows an example of a bivariate table and a multivariate table.

Table 4.2 (a) *Bivariate Table*

	1991	**1992**	**1993**
Percentage of students failed	33 per cent	38 per cent	42 per cent
Percentage of students passed	67 per cent	62 per cent	58 per cent

Table 4.2 (b) *Multivariate Table*

	1991	**1992**	**1993**
	First Attempt	Second Attempt	Third Attempt
Percentage of students who passed in Maths	27 per cent	35 per cent	–
Percentage of students who passed in English	53 per cent	60 per cent	44 per cent

Although the data in both tables is related, as the variable of attempts is distinct, the multivariate table has been displayed separately in this example. However, you should note that the tables have dealt simultaneously with two or more variables of the data.

Data interpretation

Data interpretation refers to the identification of trends in different variables. The researcher uses statistics for this purpose. The researcher is required to be familiar with the knowledge of the scales of measurement. This enables him/her to choose

the appropriate statistical method for his/her research project. The scales of measurement facilitate the allotment of numerical values to characteristics adhering to any specific rules. This measurement is also related to such levels of measurement of data like nominal, ordinal and internal and ratio levels. These levels can be explained as follows:

NOTES

- **Nominal measurement**: The nominal measurement assigns a numerical value to a specific characteristic. It is the fundamental form of measurement. The nominal measurement calculates the lowest level of data available for measurement.
- **Ordinal measurement**: This type of measurement involves allotting a specific feature to a numerical value in terms of a specific order. The ordinal scale displays the way in which the entity is measured. The ordinal scale of measurement is used to calculate and derive data pertaining to the median, percentage, rank order, correlations and percentile.
- **Interval measurement**: A researcher can depict the difference between the first aspect of a data and its another aspect using this level of measurement. The interval scale of measurement is useful for the researcher in several ways. It can be applied to calculate arithmetic mean, averages, standard deviations and to determine the correlation between different variables.
- **Ratio measurement**: In this method, there are fixed proportions (ratios) between the numerical and the amount of the characteristics that it represents. A researcher should remember while measuring the ratio levels that, there exists a fixed zero point. The ratio level of measurement facilitates researchers in determining if the aspects of ratio measurement possess any certain characteristic. Almost any type of arithmetical calculations can be executed using this scale of measurement.

The most important feature of any measuring scale is its reliability and validity, which is explained as follows:

- **Reliability**: This term deals with accuracy. A scale measurement can be said to be reliable, when it exactly measures only that which it is supposed to measure. In other words, when the same researcher repeats a test, i.e., with a different group but resembling the original group, he/she should get the same results as the former.
- **Validity**: According to Leedy, validity is the assessment of the soundness and the effectiveness of the measuring instrument. There are several types of validity, which include the following:
  - **Content Validity**: It deals with the accuracy with which an instrument measures the factors or content of the course or situations of the research study.
  - **Prognostic Validity**: It depends on the possibility to make judgements from results obtained by the concerned measuring instrument. The judgement is future oriented.
  - **Simultaneous Validity**: This involves comparison of one measuring instrument with another; one that measures the same characteristic and is available immediately.

NOTES

CHECK YOUR PROGRESS

1. What are the functions that can be performed on data?
2. How can coding of data be defined?
3. What is descriptive statistics?
4. When is univariate analysis performed?

4.3 FREQUENCY DISTRIBUTION AND GRAPHIC REPRESENTATION OF DATA

Data is referred to as recorded description of the observations of an individual or group of individuals. Examples of data may be the marks of students achieved on any test, recorded observation regarding the availability of various types of flowers in the park, recorded rates of various types of vegetables available in the local market, prices of different types of laptop made by various companies, etc.

Data may be collected from the following two important sources:

1. **External sources:** It may be divided into the following three categories:
   (i) **Primary sources:** These are the sources, which are original in nature, i.e., the data is collected from the real object regarding which the research is going on.
   (ii) **Secondary sources:** These are the sources that use data collected through the primary sources.
   (iii) **Tertiary sources:** These sources filter data from secondary sources.
2. **Internal sources:** The data collected or created by the organization itself are called internal sources. Large organizations generally create its own data. For example, the Central government generates or collects large volume of data on population through census, data related to enrolment of students in various levels of schooling, literacy rate at states and national levels and so on. All reports submitted by various committees and commissions of government constitute internal data.

Nature of Data

Data may be numerical like 2, 4, 150 and 6, and so on, or categorical, for example, republican, democrat, etc. As there is no hierarchy in the categories, it is not ordinal in nature. Numerical data may be discrete or continuous in nature. The discrete data is created either from a finite number or a countable number of possible values. On the other hand, continuous data is created from infinite number of possible values, which may be presented in the form of points on a continuous scale having no gaps or interruptions on it.

In other words, it may be said that the discrete data can be counted whereas continuous data cannot be counted. Some of the examples of discrete data are

NOTES

numbers of cars or motorcycles in the parking, fatal airline accidents occurred in India, number of enrolled students in a course, cows grazing in a field, etc. As all these can be counted, these data can be called discrete data. Examples of continuous data are weight of a person, height of an individual or a child, speed of a car, temperature measured by thermometer, quantity of water coming out of tape, etc. All these data cannot be measured with absolute precision.

Numerical data may be classified into cardinal and ordinal. In cardinal data, the absolute difference between the given two numbers is meaningful (e.g., 10 miles, 8 miles, 6 miles, 2 miles). Notice that there is a difference of two miles between the first two numbers i.e. 10 miles and 8 miles. It may be said that 10 miles is one and a quarter times more than the 8 miles, similarly, 6 miles is 2 times more that 2 miles. The ordinal data gives a comparative picture among its various values. In this case, the absolute difference between two numbers can only convey order and not the absolute difference as that of cardinal data. For example, P is either larger than Q, or equal to Q or less than Q.

Scales of measurement

Measurement is the process of assigning numbers or symbols or grades to objects or events in a well-designed manner. There are four scales of measurements, listed as follows:

1. Nominal scale
2. Ordinal scale
3. Interval scale
4. Ratio scale

1. **Nominal scale:** Statistically, this scale is the lowest measurement scale or level used by us in data analysis. In nominal scale, we simply place data into various categories. It has no order or structure. For example, the Yes/No or True/False scale used in data collection and analysis is nominal scale in which there are two categories. It does not have any order and there is no distance between these two categories, i.e., YES and NO as well as between True and False. The statistics, which we can apply with the data collected or available in nominal scales are put in the non-parametric group. Examples that come into the category of non-parametric group are mode, cross tabulation with chi-square, etc.
2. **Ordinal scale:** This scale comes next to the nominal scale in terms of power of measurement. Ranking is the simplest ordinal scale. In ordinal scale, place the given data into various categories in terms of rank or power or considering some particular trait or value or quality. For example, when you rank the given five types of fruits from most tasty to least tasty depending on our own taste or depending upon some pre- determined criteria, you are basically creating an ordinal scale of preference with regard to the taste of given fruits. There is never an absolute or objective distance between any two points or rank in this scale. It is subjective in nature and varies from person to person. Ordinal scales interpret gross order only and not the relative positional

NOTES

distances between or among various ranks. This scale uses non-parametric statistics for analysis and interpretation of the data. Some of the examples of statistics used for this scale are as follows:

a. Median and mode
b. Rank order correlation
c. Non-parametric analysis of variance

3. **Interval scale:** This scale assumes and keeps an equal distance between two consecutive points or elements being presented on the scales of measurement. This is the scale in which the interpretation of differences in the distance along the scale is possible and comprehendible. It is different from an ordinal scale. In ordinal scale, you can only present differences in order of the various categories or points but not the differences in the degree of the order. In interval scales, you can define by metrics such as logarithms. In all these cases, the distances between various points on the sale are not equal but they are strictly definable based on the metric system used. In interval scale, parametric statistical techniques are used for data analysis. Some of the examples of the statistics used are as follows:
   - Mean
   - Standard deviation
   - Correlation – r
   - Regression
   - Analysis of variance (ANOVA)
   - Factor analysis
   - Full range of advanced multivariate and modeling techniques

   Note: You can use non-parametric techniques with interval as well as ratio data, but non-parametric tests are less powerful than the parametric tests.

4. **Ratio scale:** The ratio scale is the top level scale of measurement. It is generally not available in the field of social science research. One of the most important salient features of ratio scale is that it has a true zero point. The simplest and easily comprehensible example of ratio scale is the measurement of distance or length (disregarding any philosophical points about the method of defining and identifying zero distance). It may be differentiated from that of the interval and ratio scale in an easy way. For example, the centigrade scale of temperature has a zero point, which is arbitrary in nature. Fahrenheit scale has an equivalent of this point at -32^{o} (0^{o} in centigrade scale is equal to -32^{o} in Fahrenheit scale). Hence, even though temperature appears to be in ratio scale, in reality, it is an interval scale. Similar to the interval scale data, ratio scale data also uses parametric statistical techniques. The examples of these techniques are as follows:
   - Mean
   - Standard deviation
   - Correlation – r
   - Regression

- Analysis of variance (ANOVA)
- Factor analysis
- Full range of advanced multivariate and modeling techniques

NOTES

4.3.1 Presentation of Data in Sequence: Grouping, Tabulation and Graphical Representation

After the collection of data, it requires to be presented in a systematic form so that it can be managed properly. Raw data gives no full meaning until it is arranged properly; hence, it is very important to present data in arranged form. There are two methods of presenting data. These are:

1. Grouping or tabular presentation
2. Graphical and diagrammatic presentation

Grouping and Tabulation of Data

Grouping and tabular presentation of data refers to the presentation of data in tables in a systematic organization of groups, i.e., in rows and columns.

Raw data

The mass of data in its original form as collected from the field or through experiment is called as raw data. This is the data that have not been arranged in any numerical form. It is simply collected data and have not been treated at all. Each single item in the raw data may be in the form of score, value, number, grade, observation, etc. An example of raw data is given here. This data depicts the scores of students of class IX on a test in English language having maximum marks of 20.

13, 12, 13, 12, 11, 14, 12, 11, 12, 15, 14, 10, 14, 13, 12, 13, 13, 12, 10, 11, 11, 12, 13, 14, 11, 12, 13, 14, 15 and 12.

Array

When the mass of collected raw data is arranged in an order of magnitude, may be in increasing or decreasing order, it is called as an array. The preceding raw data is arranged in the following array:

10, 10, 11, 11, 11, 11, 11, 12, 12, 12, 12, 12, 12, 12, 12, 12, 13, 13, 13, 13, 13, 13, 13, 14, 14, 14, 14, 14, 15, and 15.

Frequency distribution

The commonest way of organizing raw data is to form a frequency distribution. It may be done in the form of tally mark and frequencies (Table 4.3) and grouped frequency distribution (Table 4.4).

In tally marks and frequency table (Table 4.4) just fill the numbers in column 1 in ascending or descending order. Fill tally marks in the second column for all numbers or scores equivalent to the numbers it appears in the distribution. Tally mark for the score five may be shown by drawing four lines and crossing it. Write the total frequency of that score in the third column. This type of frequency distribution is used when the data is small in size.

NOTES

Table 4.3 *Tally Marks and Frequencies*

Column 1 Score	Column 2 Tally Mark	Column 3 Frequency (f)
10	//	2
11	/////	5
12	///// ////	9
13	///// //	7
14	/////	5
15	//	2
		Total N = 30

When the amount of data is largle, it is useful to group data into classes or class intervals or categories as shown in Table 4.4.

Table 4.4 *Grouped Frequency Distribution*

Marks Obtained (Class Interval)	**Frequency (f)**
10-11	7
12-13	16
14-15	7
	Total N = 30

The chief advantage of grouped frequency is that it gives a clear picture of overall distribution of scores or values. Classes should not be either too many or too few. Ideally it should vary from 5 to 20. Too many class intervals tend to destroy the patterns of distribution; on the other hand, too few classes will destroy much of the detail present in the raw data.

Classification of data

It is the process of arranging the data according to some common characteristics possessed by the item constituting the data. The main purpose of classifying data is to condense it.

In Table 4.4, the second class interval or category is 12-13. The end numbers of this class are 12 and 13 which are called the class limits for this class. 12 is the lower class limit whereas 13 is the upper class limit. Similarly, in all class intervals, there are always lower and upper class limits.

Class boundaries

In Table 4.4, the second class interval 12-13 theoretically includes the scores 11.5 and 13.5. These scores are called lower class boundary and upper class boundary, respectively, for the second class. For any class interval, the class boundaries may be determined by adding the upper class limit of one class to the lower class limit of the next class and dividing the sum by 2.

Example 4.1: For the second class in Table 4.4,

lower class boundary = (11+12)/2 =23/2 = 11.5

The upper class boundary = (13+14)/2 = 27/2 =13.5

Width or class size of a given class limit or class interval

Width or class size is the difference between the lower class boundary and the upper class boundary.

Width or class size of class limit = Upper class boundary – Lower class boundary

For the second class interval in Table 4.4, width or class size = 13.5-11.5 = 2.

Number of classes

The number of the classes for the given data is determined on the basis of:

- The number of items in the series. If the items are large, the number of classes will be more and if the items are few, the numbers of classes will be few.
- The range is the difference between the highest and lowest value in the distribution. Higher the range, more is the number of class interval.

Actual class interval

Actual class interval is the size of class interval in a frequency distribution, which is the distance from one lower limit to the next lower limit. Stated limits and actual limits for Table 4.4 is presented in Table 4.5.

Table 4.5 *Actual or real class limit*

Stated limit	Real limit
10-11	9.5-11.5
12-13	11.5-13.5
14-15	13.5-15.5

Midpoints of class intervals

A midpoint of class interval is the point exactly halfway between the upper and lower real limits and can be calculated by dividing the sum of the upper and lower limits by 2. For example,

Mid point of class interval 14-15 = (14+15)/2=29/2=14.5

Or (13.5 + 15.5)/2 = 29/2 = 14.5

Cumulative frequency Distribution

NOTES

At times, the investigator needs to know the number of observations or scores below a certain specific limit or point in the distribution. This is done with the help of a cumulative frequency distribution. To develop a cumulative frequency distribution just add one more column to Table 4.5 and give the heading to this column as cumulative frequency as shown in Table 4.6. While filling the cumulative frequency of any class interval just add all the frequencies of previous class intervals in the frequency of this particular class interval. The cumulative frequency of the last class interval will always be equal to the total size of the data or total frequency or N.

Table 4.6 *Cumulative Frequency Distribution*

Marks Obtained (Class Interval)	Frequency (f)	Cumulative frequency (Cf)
10-11	7	7
12-13	16	23
14-15	7	30
	Total N = 30	

Graphical Representation of Data

Data can also be presented through charts and graphs to make it more attractive and informative and easily understandable. The data presented in this form are also visually more dramatic than frequency distribution. These means of presenting data are useful for conveying an impression of the overall shape of a distribution and highlighting any clustering of cases or values in a particular range of scores.

Types of graphical representation of data: There are various types of graphical representation of data; some of them are as follows:

- Histogram
- Frequency polygon
- Frequency curve
- Cumulative Prequency or ogive
- **Histogram:** Histogram is a diagram used to present a frequency distribution. It consists of a set of elongated rectangles, which represent frequencies of

NOTES

the various classes. If all the classes are of the same width, then all the rectangles will also have the same width. The frequencies are then represented by the heights of the rectangles.

Using histogram is appropriate for the data or frequencies given in continuous interval scale. The following steps may be followed to construct a histogram from a frequency distribution.

Step 1: Plot the given class interval or scores along the X axis using the real class limits.

Step 2: Plot frequencies along the Y axis.

Step 3: For each category given in the frequency distribution, construct a bar with heights corresponding to the number of cases in the category and with width corresponding to the real limits of the class intervals.

Step 4: Label each axis of the graph properly.

Step 5: Give the title or heading to the graph.

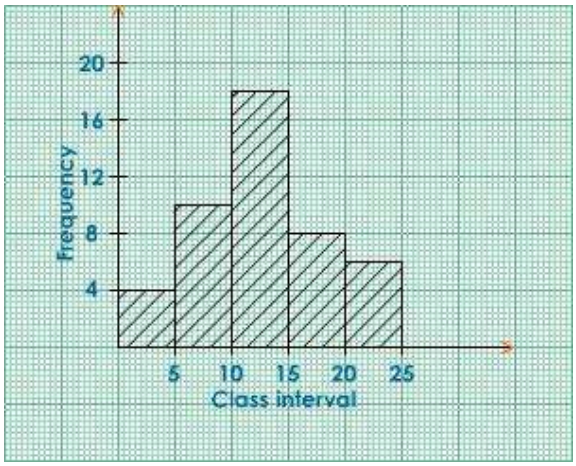
- **Frequency polygon:** This is another way of representing frequency distribution in the form of a graph. To present frequency distribution first of all draw histogram as discussed earlier and then join the mid point of all the tops of the rectangles. A frequency polygon is essentially a line graph of the frequency distribution.
- **Frequency curve:** A frequency curve is a smooth and free hand curve for the graphical representation of the frequency distribution. The objective of drawing it is to eliminate the accidental variables that may be present in the data. Frequency curve is not a closed curve. It is drawn in the same way as that of the frequency polygon except drawing the rectangle in the graph. Just draw points where the class interval or its midpoints and its frequency meet and join all points with a line.
- **Cumulative frequency curve or ogives:** For the construction of cumulative frequency curve, first develop the cumulative frequency distribution. The cumulative frequency curves or ogives are drawn along X and Y axis in which cumulative frequency is shown on Y axis and the class interval is shown on the X axis. A line graph in the form of curve is plotted connective the cumulative frequencies corresponding to the upper boundaries of the classes. Some examples for graphical representation have been given here. For convenience, graph paper has been used in certain examples.

Example 4.2: Construct a histogram for the following frequency distribution:

Class Interval	Frequency
0 - 5	4
5 - 10	10
10 - 15	18
15 - 20	8
20 - 25	6

NOTES

Solution: Histogram is constructed as follows:

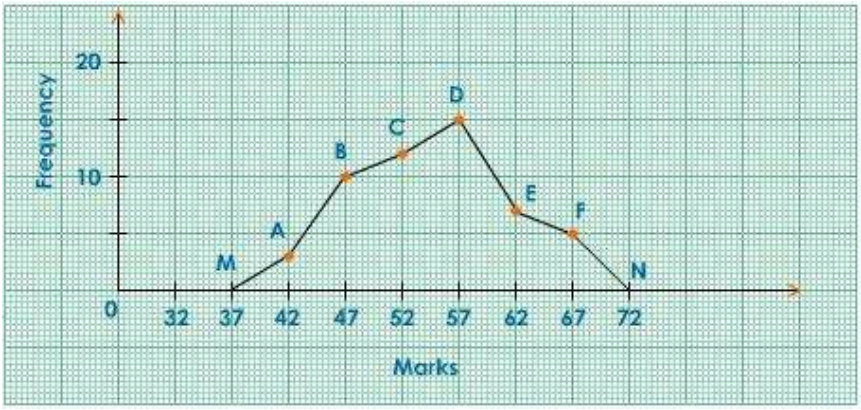


Histogram of Given Frequency

Example 4.3: The marks of 52 students are recorded here. Draw a frequency polygon for this data.

Class	Mid-mark	Frequency
40 - 44	42	3
45 - 49	47	10
50 - 54	52	12
55 - 59	57	15
60 - 64	62	7
65 - 69	67	5

Solution: Frequency polygon is as follows:



Frequency Polygon for Marks

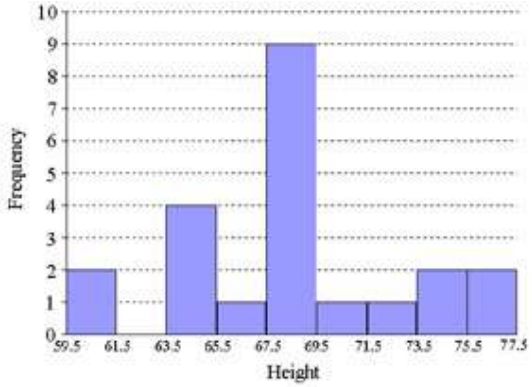
Example 4.4: Suppose the heights of twenty students in inches are:

60, 68, 69, 64, 68, 67, 68, 69, 77, 69, 69, 72, 69, 65, 65, 68, 64, 71, 74, 74

The variable is height. The sample size is $n = 20$. The mean and the median are both about 68.5 inches. The standard deviation is about 3.9 inches. Draw a frequency histogram.

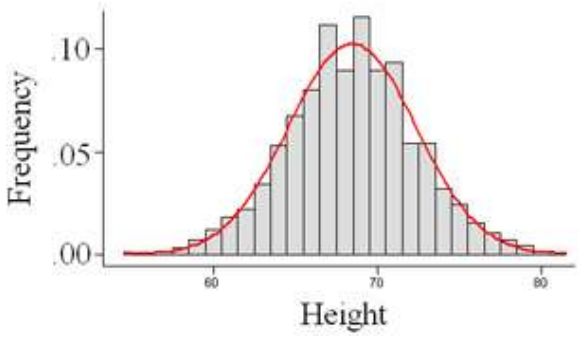
NOTES

Solution: The following is a frequency histogram for the height data:



Frequency Histogram for the Height Data

By drawing a smooth curve over the preceding histogram, the following frequency curve is obtained:



Frequency Curve for the Height Data

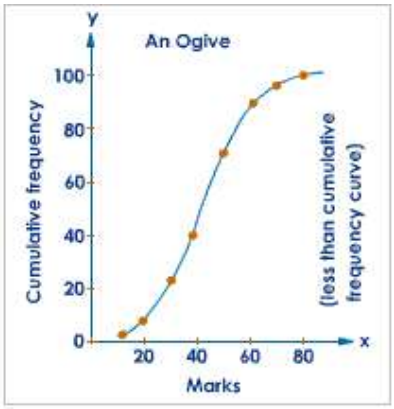
Example 4.5: Draw a 'less than' ogive curve for the following data:

Marks	Frequency	Cumulative Frequency
0 - 10	2	2
10 - 20	8	10
20 - 30	12	22
30 - 40	18	40
40 - 50	28	68
50 - 60	22	90
60 - 70	6	96
70 - 80	4	100

Solution: Plot the points with coordinates having abscissae as actual limits and ordinates as the cumulative frequencies: (10, 2), (20, 10), (30, 22), (40, 40), (50, 68),

NOTES

(60, 90), (70, 96) and (80, 100) as the coordinates of the points. Joining the points plotted by a smooth curve forms the ogive.



'Less than' Ogive Curve

CHECK YOUR PROGRESS

5. What is meant by measurement?
6. What is array?
7. How will you differentiate between numerical and cardinal data?

4.4 MEASURES OF CENTRAL TENDENCY AND VARIABILITY

There are several commonly used measures of central tendency, such as arithmetic mean, mode and median. These values are very useful not only in presenting the overall picture of the entire data but also for the purpose of making comparisons among two or more sets of data.

As an example, questions like 'How hot is the month of June in Delhi?' can be answered, generally by a single figure of the average for that month. Similarly, suppose we want to find out if boys and girls at age 10 years differ in height for the purpose of making comparisons. Then, by taking the average height of boys of that age and average height of girls of the same age, we can compare and record the differences.

While arithmetic mean is the most commonly used measure of central location, mode and median are more suitable measures under certain set of conditions and for certain types of data. However, each measure of central tendency should meet the following requisites :

1. It should be easy to calculate and understand.
2. It should be rigidly defined. It should have only one interpretation so that the personal prejudice or bias of the investigator does not affect its usefulness.

NOTES

3. It should be representative of the data. If it is calculated from a sample, then the sample should be random enough to be accurately representing the population.
4. It should have sampling stability. It should not be affected by sampling fluctuations. This means that if we pick 10 different groups of college students at random and compute the average of each group, then we should expect to get approximately the same value from each of these groups.
5. It should not be affected much by extreme values. If few very small or very large items are present in the data, they will unduly influence the value of the average by shifting it to one side or other, so that the average would not be really typical of the entire series. Hence, the average chosen should be such that it is not unduly affected by such extreme values.

Meaning of the Measures of Central Tendency

If the progress scores of the students of a class are taken and they are arranged in a frequency distribution, we may sometime find that there are very few students who either score very high or very low. The marks of most of the student will lie somewhere between the highest and the lowest scores of the whole class. This tendency of a group about distribution is named as central tendency and typical score that lies in between the extremes and shared by most of the students is referred to as a measure of central tendency. Tate in 1955 defined the measures of central tendency as, '*A sort of average or typical value of the items in the series and its function is to summarize the series in terms of this average value*.'

The most common measures of central tendency are:

1. Arithmetic Mean or Mean
2. Median
3. Mode

Let us consider the three measures of central tendency.

I. Arithmetic Mean

This is also commonly known as simply the mean. Even though average, in general, means any measure of central location, when we use the word average in our daily routine, we always mean the arithmetic average. The term is widely used by almost every one in daily communication. We speak of an individual being an average student or of average intelligence. We always talk about average family size or average family income or Grade Point Average (GPA) for students, and so on.

Calculating arithmetic mean (*M*): The simplest but most useful measure of central tendency is the arithmetic mean. It can be defined as the sum of all the values of the items in a series divided by the number of items. It is represented by the symbol *M*.

NOTES

Calculation of mean in the case of ungrouped data

Let $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9$ and X_{10} be the scores obtained by 10 students on an achievement test. Then the arithmetic mean or mean score of the group of these ten students can be calculated as:

$$M = X_1 + X_2 + X_3 + X_4 + X_5 + \ldots + X_{10}/10$$

The formula for calculating the mean of an ungrouped data is as follows:

$$M = \sum X/N$$

Where, $\sum X$ stands for the sum of sores or values of the items and N for the total number in a series or group.

Calculation of mean in the case of grouped data (Data in the form of Frequency Distribution)

General Method: In a frequency distribution where all the frequencies are greater than one, the mean is calculated by the formula:

$$\text{M} = \sum fX / N$$

Where, X represents the mid-point of the class interval, f its respective frequency and N the total of all frequencies.

Short-Cut Method: Mean for the grouped data can be computed easily with the help of following formula:

$$M = A + \sum fx'/N \times i$$

Where,

A = Assumed mean

i = Class interval

f = Respective frequency of the mid-values of the class intervals

N = Total Frequency

x' = $X - A / i$

Combined Mean: If the arithmetic averages and the number of items in two or more related groups are known, the combined (or composite) mean of the entire group can be obtained by the following formula:

$$\overline{\overline{X}} = \left[\frac{n_1 \overline{x}_1 + n_2 \overline{x}_2}{n_1 + n_2} \right]$$

The advantage of combined arithmetic mean is that, one can determine the over, all mean of the combined data without having to going back to the original data.

An Example:

Find the combined mean for the data given below

$n_1 = 10, x_1 = 2, n_2 = 15, x_2 = 3$

Solution:

$$\overline{\overline{X}} = \left[\frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}\right]$$

$$= \left[\frac{10 \times 2 + 15 \times 3}{10 + 15}\right]$$

$$= \frac{20 + 45}{25}$$

$$= 2.6$$

NOTES

For discussion purposes, let us assume a variable X which stands for some scores, such as the ages of students. Let the ages of 5 students be 19, 20, 22, 22 and 17 years. Then variable X would represent these ages as follows:

X: 19, 20, 22, 22, 17

Placing the Greek symbol σ(Sigma) before $\bar{X}$ would indicate a command that all values of X are to be added together. Thus:

$\sigma\bar{X} = 19 + 20 + 22 + 22 + 17$

The mean is computed by adding all the data values and dividing it by the number of such values. The symbol used for sample average is $\overline{X}$ so that:

$$\overline{X} = \frac{19 + 20 + 22 + 22 + 17}{5}$$

In general, if there are n values in the sample, then

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n}$$

In other words,

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}, \qquad i = 1,\ 2\ ...\ n.$$

The above formula states, add up all the values of X_i where the value of i starts at 1 and ends at n with unit increments so that $i = 1, 2, 3, ... n$.

If instead of taking a sample, we take the entire population in our calculations of the mean, then the symbol for the mean of the population is μ (mu) and the size of the population is N, so that:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}, \qquad i = 1,\ 2\ ...N.$$

NOTES

If we have the data in grouped discrete form with frequencies, then the sample mean is given by:

$$\overline{X} = \frac{\Sigma f(X)}{\Sigma f}$$

Where, Σf = Summation of all frequencies' n
$\Sigma f(X)$ = Summation of each value of X multiplied by its corresponding frequency (f).

Example 4.6: Let us take the ages of 10 students as follows:

19, 20, 22, 22, 17, 22, 20, 23, 17, 18

This data can be arranged in a frequency distribution as follows:

(X)	(f)	$f(X)$
17	2	34
18	1	18
19	1	19
20	2	40
22	3	66
23	1	23
	Total = 10	200

In the above case we have $\Sigma f = 10$ and $\Sigma f(X) = 200$, so that:

$$\overline{X} = \frac{\Sigma f(X)}{\Sigma f}$$

$$= 200/10 = 20$$

Characteristics of the mean

The arithmetic mean has three interesting properties. These are:

1. The sum of the deviations of individual values of X from the mean will always add up to zero. This means that if we subtract all the individual values from their mean, then some values will be negative and some will be positive, but if all these differences are added together then the total sum will be zero. In other words, the positive deviations must balance the negative deviations. Or symbolically:

$$\sum_{i=1}^{n}(X_i - \overline{X}) = 0, i = 1, 2, \ldots n.$$

2. The second important characteristic of the mean is that it is very sensitive to extreme values. Since the computation of the mean is based upon inclusion of all values in the data, an extreme value in the data would shift the mean towards it, thus making the mean unrepresentative of the data.
3. The third property of the mean is that the sum of squares of the deviations about the mean is minimum. This means that if we take differences between individual values and the mean and square these differences individually and then add these squared differences, then the final figure

will be less than the sum of the squared deviations around any other number other than the mean. Symbolically, it means that:

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 = \text{Minimum}, i = 1, 2, \ldots n.$$

NOTES

Advantages of mean

The following are the various advantages of mean:

1. Its concept is familiar to most people and is intuitively clear.
2. Every data set has a mean, which is unique and describes the entire data to some degree. For instance, when we say that the average salary of a professor is ₹ 25,000 per month, it gives us a reasonable idea about the salaries of professors.
3. It is a measure that can be easily calculated.
4. It includes all values of the data set in its calculation.
5. Its value varies very little from sample to sample taken from the same population.
6. It is useful for performing statistical procedures, such as computing and comparing the means of several data sets.

Disadvantages of mean

The following are the various disadvantages of mean:

1. It is affected by extreme values, and hence, not very reliable when the data set has extreme values especially when these extreme values are on one side of the ordered data. Thus, a mean of such data is not truly a representative of such data. For instance, the average age of three persons of ages 4, 6 and 80 years gives us an average of 30.
2. It is tedious to compute for a large data set as every point in the data set is to be used in computations.
3. We are unable to compute the mean for a data set that has open-ended classes either at the high or at the low end of the scale.
4. The mean cannot be calculated for qualitative characteristics, such as beauty or intelligence, unless these can be converted into quantitative figures such as intelligence into IQs.

II. Median

The median is a measure of central tendency and it appears in the centre of an ordered data. It divides the list of ordered values in the data into two equal parts so that half of the data will have values less than the median and half will have values greater than the median.

If the total number of values is odd, then we simply take the middle value as the median. For instance, if there are 5 numbers arranged in order, such as 2, 3, 3, 5, 7, then 3 is the middle number and this will be the median. However, if the total number of values in the data is even, then we take the average of the middle two values. For instance, let there be 6 numbers in the ordered data such as 2, 3, 3, 5, 7,

NOTES

8, then the average of middle two numbers which are 3 and 5 would be the median, which is

$$\text{Median} = \frac{(3+5)}{2} = 4$$

In general, the median is $\frac{n+1}{2}$ th observation in the ordered data.

The median is a useful measure in the sense that it is not unduly affected by extreme values and is specially useful in open ended frequencies.

Calculating Median (M_d): If the items of a series are arranged in ascending or descending order of magnitude, the measure or value of the central item in the series is termed as median. The median of a distribution can thus be said as the point on the score scale below which half (or 50 per cent) of the scores fall. Thus median is the score or the value of that central item which divides the series into two equal parts. Therefore, it should be understood that the central item itself is not the median. It is only the measure or value of the central item that is known as the median. For example, if we arrange in ascending or descending order the marks of 5 students, then the marks obtained y the third student from either side will be termed as median of the scores of the group of students under consideration.

Computation of median for ungrouped data

The following two situations could arise:

1. **When *N* (No. of Items in a Series) is Odd:** In this case where *N* is odd (not divisible by 2), the median can be computed by the following formula:

   M_d = The measure or value of the $(N + 1)/2$ th item.

2. **When *N* (No. of Items in a Series) is Even:** In this case where *N* is even (divisible by 2), the median can be determined by the following formula:

   M_d = The value of the $(N/2)$ th item + The value of $[(N/2) + 1]$ th item/ 2

Calculation of median for grouped data (In the Form of Frequency Distribution)

If the data is available in the form of a frequency distribution like below, then calculation of median first requires the location of median class.

Scores	*f*
65-69	1
60-64	3
55-59	4
50-54	7
45-49	9
40-44	11
35-39	8

30-34	4
25-29	2
20-24	1
	$N = 50$

NOTES

Actually, median is the measure or score of the central item. Therefore, it is needed to locate the central item. It may be done through the formulae given earlier in case of ungrouped data for the odd and even values of N (total frequencies). Here, in the present distribution, N (= 50) is even. Therefore, median will fall somewhere between the score of 25th and 26th items in the given distribution. In the given frequency distribution table, if we add frequencies either above or below we may see that the class interval designated as 40-44 is to be labeled as the class where the score representing median will fall.

After estimating the median class, the median of the distribution may be interpolated with the help of following formula:

$$M_d = L + [\ (N/2) - F\ /f] \times i$$

Where,

L = Exact lower limit of the median class

F = Total of all frequencies before in the median class

f = Frequency of the median class

i = Class interval

N = Total of all the frequency

By applying the above formula, we can compute the median of the given distribution in the following way:

$$M_d = 39.5 + (50/2) - 15 / 11 \times 5 = 39.5 + 10/11 \times 5$$
$$= 39.5 + 50/11 = 39.5 + 4.55 = 44.05$$

Advantages of median

The following are the advantages of median:

1. Median is a positional average and hence the extreme values in the data set do not affect it as much as they do to the mean.
2. Median is easy to understand and can be calculated from any kind of data, even for grouped data with open-ended classes.
3. We can find the median even when our data set is qualitative and can be arranged in the ascending or the descending order, such as average beauty or average intelligence.
4. Similar to mean, median is also unique meaning that there is only one median in a given set of data.
5. Median can be located visually when the data is in the form of ordered data.
6. The sum of absolute differences of all values in the data set from the median value is minimum meaning that it is less than any other value of central tendency in the data set, which makes it more central in certain situations.

NOTES

Disadvantages of median

The following are the disadvantages of median:

1. The data must be arranged in order to find the median. This can be very time consuming for a large number of elements in the data set.
2. The value of the median is affected more by sampling variations. Different samples from the same population may give significantly different values of the median.
3. The calculation of median in case of grouped data is based on the assumption that the values of observations are evenly spaced over the entire class interval and this is usually not so.
4. Median is comparatively less stable than the mean, particularly for small samples, due to fluctuations in sampling.
5. Median is not suitable for further mathematical treatment. For instance, we cannot compute the median of the combined group from the median values of different groups.

III. Mode

The mode is another form of average and can be defined as the most frequently occurring value in the data. The mode is not affected by extreme values in the data and can easily be obtained from an ordered set of data. It can be useful and more representative of the data under certain conditions and is the only measure of central tendency that can be used for qualitative data. For instance, when a researcher quotes the opinion of an average person, he is probably referring to the most frequently expressed opinion which is the modal opinion. In our example of ages of 10 students as:

19, 20, 22, 22, 17, 22, 20, 23, 17 and 18

The mode is 22, since it occurs more often than any other value in this data.

Calculating Mode (M_0): Mode is defined as the size of a variable which occurs most frequently. It is the point on the score sale that corresponds to the maximum frequency of the distribution. In any series, it is the value of the item which is most characteristics or common and is usually repeated the maximum number of times.

Computation of mode for ungrouped data

Mode can easily be computed merely by looking at the data. All that one has to do is to find out the score which is repeated maximum number of times.

For example, suppose we have to find out the value of mode from the following scores of students:

25, 29, 24, 25, 27, 25, 28, 25, 29

Here the score 25 is repeated maximum number of times and thus, value of the mode in this case is 25.

Computation of mode for grouped data

When data is available in the form of frequency distribution, the mode is computed from the following formula:

Mode (M_0) = 3 M_d – 2M

Where, M_d is the median and M is the mean of the given distribution. The mean as well as the median of the distribution are first computed and then, with the help of the above formula, mode is computed.

NOTES

Another method for grouped data

Mode can be computed directly from the frequency distribution table without calculating mean and median. For this purpose, we can use the following formula:

$$M_0 = L + f_1/ (f_1 + f_{-1}) \times i$$

Where,

L = Lower limit of the model class (the class in which mode maybe supposed to lie)

i = Class interval

f_1 = Frequency of the class adjacent to the modal class for which lower limit is greater than that for the modal class.

f_{-1} = Frequency of the class adjacent to the modal class for which the lower limit is less than that for the modal class.

Advantages of mode

The following are the advantages of mode:

1. Similar to median, the mode is not affected by extreme values in the data.
2. Its value can be obtained in open-ended distributions without ascertaining the class limits.
3. It can be easily used to describe qualitative phenomenon. For instance, if most people prefer a certain brand of tea then this will become the modal point.
4. Mode is easy to calculate and understand. In some cases it can be located simply by observation or inspection.

Disadvantages of mode

The following are the disadvantages of mode:

1. Quite often, there is no modal value.
2. It can be bi-modal or multi-modal or it can have all modal values making its significance more difficult to measure.
3. If there is more than one modal value, the data is difficult to interpret.
4. A mode is not suitable for algebraic manipulations.
5. Since the mode is the value of maximum frequency in the data set, it cannot be rigidly defined if such frequency occurs at the beginning or at the end of the distribution.
6. It does not include all observations in the data set, and hence, less reliable in most of the situations.

NOTES

Weighted arithmetic mean

In the computation of arithmetic mean we had given equal importance to each observation in the series. This equal importance may be misleading if the individual values constituting the series have different importance as in the following example:

The Raja Toy shop sells

Toy Cars at	₹ 3 each
Toy Locomotives at	₹ 5 each
Toy Aeroplanes at	₹ 7 each
Toy Double Decker at	₹ 9 each

What shall be the average price of the toys sold, if the shop sells 4 toys, one of each kind?

Mean Price, i.e., $\bar{x} = \frac{\sum x}{4} = \text{Rs}\,\frac{24}{4} = ₹\,6$

In this case the importance of each observation (Price quotation) is equal in as much as one toy of each variety has been sold. In the above computation of the arithmetic mean this fact has been taken care of by including 'once only' the price of each toy.

But if the shop sells 100 toys: 50 cars, 25 locomotives, 15 aeroplanes and 10 double deckers, the importance of the four price quotations to the dealer is **not equal** as a source of earning revenue. In fact their respective importance is equal to the number of units of each toy sold, i.e.,

The importance of Toy Car	50
The importance of Locomotive	25
The importance of Aeroplane	15
The importance of Double Decker	10

It may be noted that 50, 25, 15, 10 are the quantities of the various classes of toys sold. It is for these quantities that the term 'weights' is used in statistical language. Weight is represented by symbol 'w', and Σw represents the sum of weights.

While determining the 'average price of toy sold' these weights are of great importance and are taken into account in the manner illustrated below:

$$\bar{x} = \frac{w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4}{w_1 + w_2 + w_3 + w_4} = \frac{\sum wx}{\sum w}$$

When w_1, w_2, w_3, w_4 are the respective weights of x_1, x_2, x_3, x_4 which in turn represent the price of four varieties of toys, viz., car, locomotive, aeroplane and double decker, respectively.

$$\bar{x} = \frac{(50\times3)+(25\times5)+(15\times7)+(10\times9)}{50+25+15+10}$$

$$= \frac{(150)+(125)+(105)+(90)}{100} = \frac{470}{100} = ₹\,4.70$$

The table below summarizes the steps taken in the computation of the weighted arithmetic mean.

$$\Sigma w = 100; \quad \Sigma wx = 470$$

$$\bar{x} = \frac{\sum wx}{\sum w} = \frac{470}{100} = 4.70$$

NOTES

The weighted arithmetic mean is particularly useful where we have to compute the *mean of means.* If we are given two arithmetic means, one for each of two different series, in respect of the *same variable*, and are required to find the arithmetic mean of the combined series, the weighted arithmetic mean is the only suitable method of its determination.

Weighted Arithmetic Mean of Toys Sold by the Raja Toy Shop

Toys	*Price Per Toy* ₹ x	*Number Sold* w	*Price × Weight* xw
Car	3	50	150
Locomotive	5	25	125
Aeroplane	7	15	105
Double Decker	9	10	90
		$\Sigma w = 100$	$\Sigma xw = 470$

Example 4.7: The arithmetic mean of daily wages of two manufacturing concerns A Ltd. and B Ltd. is ₹ 5 and ₹ 7, respectively. Determine the average daily wages of both concerns if the number of workers employed were 2,000 and 4,000, respectively.

Solution: (*a*) Multiply each average (viz., 5 and 7) by the number of workers in the concern it represents.

(*b*) Add up the two products obtained in (*a*) above

(*c*) Divide the total obtained in (*b*) by the total number of workers.

Weighted Mean of Mean Wages of A Ltd. and B Ltd.

Manufacturing Concern	*Mean Wages* x	*Workers Employed* w	*Mean Wages × Workers Employed* wx
A Ltd.	5	2,000	10,000
B Ltd.	7	4,000	28,000
		$\sum w = 6{,}000$	$\sum wx = 38{,}000$

$$\bar{x} = \frac{\sum wx}{\sum w}$$

$$= \frac{38{,}000}{6{,}000}$$

$$= ₹\ 6.33$$

The above mentioned examples explain that 'Arithmetic Means and Percentage' are not original data. They are derived figures and their importance is relative to the original data from which they are obtained. This relative importance must be taken into account by weighting while averaging them (means and percentage).

Different positional numbers

The position of value in statistics is determined using specific methods for a given set of data or observations. The following are the popular common measures of positions:

NOTES

- **Percentiles:** Percentiles are those values which divide a given data set into hundred equal parts. It is the value of a variable below which certain per cent of observations fall. For instance, the 25th percentile is the value below which 25 per cent of the observations occur. The 25th percentile is also referred as the first quartile, the 50th percentile as the median or second quartile and the 75th percentile as the third quartile.
- **Quartiles:** It segments the data in four regions and is commonly used to measure the position of value in statistics. It is a number and not a range of values.
- **Standard Scores:** It is also termed as *Z*-values, *Z*-scores, normal scores and standardized variables. It is a dimensionless quantity and can be calculated using the following formula:

$$Z = (X - \mu)/\sigma$$

Measures of position values

We have defined the median as the value of the item which is located at the centre of the array, we can define other measures which are located at other specidied points. Thus, the *N*th *percentile* of an array is the value of the item such that *N* per cent items lie *below* it. Clearly then the *N*th percentile P_n of grouped data is given by,

$$P_n = l + \frac{\frac{nN}{100} - C}{f} \times i$$

where, *l* is the lower limit of the class in which *nN*/100th item lies, *i* its width, *f* its frequency, *C* the cumulative frequency upto (but not including) this class, and *N* is the total number of items.

We similarly define the *N*th *decile* as the value of the item below which (*nN*/10) items of the array lie. Clearly,

$$D_n = P_{10n} = l + \frac{\frac{nN}{10} - C}{f} \times i \quad (1)$$

The other most commonly referred to measures of location are the quartiles. Thus, *n*th quartile is the value of the item which lie at the *n*(*N*/4)th item. Clearly Q_2, the second quartile is the median. For grouped data,

$$Q_n = P_{25n} = l + \frac{\frac{nN}{4} - C}{f} \times i \quad (2)$$

Some measures other than measures of central tendency are often employed when summarizing or describing a set of data where it is necessary to divide the data into equal parts. These are positional measures and are called quantiles and consist of quartiles, deciles and percentiles. The quartiles divide the data into four equal parts. The deciles divide the total ordered data into ten equal parts and percentiles divide

NOTES

the data into 100 equal parts. Consequently, there are three quartiles, nine deciles and 99 percentiles. The quartiles are denoted by the symbol Q so that Q_1 will be such point in the ordered data which has 25 per cent of the data below and 75 per cent of the data above it. In other words, Q_1 is the value corresponding to $\left(\frac{n+1}{4}\right)$th ordered observation. Similarly, Q_2 divides the data in the middle, and is also equal to the median and its value Q_2 is given by:

Q_2 = The value of $2\left(\frac{n+1}{4}\right)$th ordered observation in the data.

Similarly, we can calculate the values of various deciles. For instance,

$D_1 = \left(\frac{n+1}{10}\right)$th observaton in the data

$D_7 = 7\left(\frac{n+1}{10}\right)$th observation in the ordered data.

Percentiles are generally used in the research area of education where people are given standard tests and it is desirable to compare the relative position of the subject's performance on the test. Percentiles are similarly calculated as:

$P_7 = 7\left(\frac{n+1}{100}\right)$th observation in the ordered data.

$P_{69} = 69\left(\frac{n+1}{100}\right)$th observation in the ordered data.

Quartiles

The formula for calculating the values of quartiles for grouped data is given as follows.

$Q = L + (j/f)C$

Where,

Q = The quartile under consideration.

L = Lower limit of the class interval which contains the value of Q.

j = The number of units we lack from the class interval which contains the value of Q, in reaching the value of Q.

f = Frequency of the class interval containing Q.

C = Size of the class interval.

Let us assume we took the data of the ages of 100 students and a frequency distribution for this data has been constructed as shown.

NOTES

The frequency distribution is as follows:

Ages (*CI*)	*Mid-point* (*X*)	(*f*)	*f*(*X*)	$f(X)^2$
16 and upto 17	16.5	4	66	1089.0
17 and upto 18	17.5	14	245	4287.5
18 and upto 19	18.5	18	333	6160.5
19 and upto 20	19.5	28	546	10647.0
20 and upto 21	20.5	20	410	8405.0
21 and upto 22	21.5	12	258	5547.0
22 and upto 23	22.5	4	90	2025.0
		Totals = 100	1948	38161

In our case, in order to find Q_1, where Q_1 is the cut-off point so that 25 per cent of the data is below this point and 75 per cent of the data is above, we see that the first group has 4 students and the second group has 14 students making a total of 18 students. Since Q_1 cuts off at 25 students, it is the third class interval which contains Q_1. This means that the value of L in our formula is 18.

Since we already have 18 students in the first two groups, we need 7 more students from the third group to make it a total of 25 students, which is the value of Q_1. Hence, the value of (j) is 7. Also, since the frequency of this third class interval which contains Q_1 is 18, the value of (f) in our formula is 18. The size of the class interval C is given as 1. Substituting these values in the formula for Q, we get

$$Q_1 = 18 + (7/18)1$$
$$= 18 + 0.38 = 18.38$$

This means that 25 per cent of the students are below 18.38 years of age and 75 per cent are above this age.

Similarly, we can calculate the value of Q_2, using the same formula. Hence,

$$Q_2 = L + (j/f)C$$
$$= 19 + (14/28)1$$
$$= 19.5$$

This also happens to be the median.

By using the same formula and the same logic we can calculate the values of all deciles as well as percentiles.

Geometric mean

If α, β, γ are in GP, then β is called a *geometric mean* between α and γ, written as GM.

If $a_1, a_2, ...,a_n$ are in GP, then $a_2, ..., a_{n-1}$ are called *geometric means* between a_1 and a_n.

Thus, 3, 9, 27 are three geometric means between 1 and 81.

Non-zero quantities $a_1, a_2, a_3, ..., a_n,....$, each term of which is equal to the product of preceding term and a constant number, form a *Geometrical Progression* (written as G.P.).

NOTES

Thus, all the following quantities are in G.P.

(a) $1, 2, 4, 8, 16, ...$

(b) $3, -1, \frac{1}{3}, \frac{-1}{9}, \frac{1}{27}, ...$

(c) $1, \sqrt{2}, 2, 2\sqrt{2}, ...$

(d) $a, \frac{a}{b}, \frac{a}{b^2}, \frac{a}{b^3}, ...,$ where $a \neq 0, b \neq 0.$

(e) $1, \frac{1}{5}, \frac{1}{25}, \frac{1}{125}, ...$

The constant number is termed as the *common ratio* of the G.P.

The *n*th Term of a G.P.

Let first term be a and r, the common ratio, By definition the G.P. is $a, ar, ar^2, ...$

$$1\text{st term} = a = ar^0 = ar^{1-1}$$

$$2\text{nd term} = ar = ar^1 = ar^{2-1}$$

... ...

In general, nth term $= ar^{n-1}$.

In examples of the preceding section, we compute 5th, 7th, 3rd, 11th and 8th term of (a), (b), (c), (d) and (e) respectively.

In (a) Ist term is 1 and common ratio = 2.

Hence, 5th term $= ar^4 = 1.2^4 = 16.$

In (b) $a = 3, r = \frac{-1}{3}$, hence, 7th term $= ar^6 = 3\left(\frac{-1}{3}\right)^6 = \frac{1}{243}.$

In (c) $a = 1, r = \sqrt{2}$, hence, 3rd term $= ar^2 = 2.$

In (d) Ist term $= a, r = \frac{1}{b}$, hence, 11th term $= ar^{10} = \frac{a}{b^{10}}.$

In (e) $a = 1, r = \frac{1}{5}$, hence, 8th term $= ar^2 = \frac{1}{5^7} = \frac{1}{78125}.$

Sum of First *n* Terms of a G.P.

Let $a, ar, ar^2, ...$ be a given G.P. and let S_n be the sum of its first n terms.

Then, $S_n = a + ar + ar^2 + ... + ar^{n-1}.$

This gives that $rS_n = ar + ar^2 + ... + ar^{n-1} + ar^n$

Subtracting, we get, $S_n - r\,S_n = a - ar^n = a\,(1 - r^n)$

In case $r \neq 1$, $S_n = \frac{a\left(1-r^n\right)}{\left(1-r\right)}$

In case $r = 1$, $S_n = a + a + a + ... + a$ (n times)

$= na.$

Thus, sum of n terms of a G.P. is $\frac{a(1-r^n)}{1-r}$ provided $r \neq 1$.

In case $r = 1$, sum of G.P. is na.

NOTES

Example 4.8: Find the sum of the first 14 terms of a G.P.

$$3, 9, 27, 81, 243, 729, ...$$

Solution: In this case $a = 3$, $r = 3$, $n = 14$.

$$\text{So,} \quad S_n = \frac{a(1-r^n)}{1-r} = \frac{3(1-3^{14})}{1-3}$$

$$= \frac{3}{2}(3^{14} - 1).$$

Example 4.9: Find the sum of first 11 terms of a G.P. given by

$$1, -\frac{1}{2}, \frac{1}{4}, -\frac{1}{8} \ldots, \ldots$$

Solution: Here, $a = 1$, $r = -\frac{1}{2}$, $n = 11$.

$$\text{So,} \quad S_n = \frac{a(1-r^n)}{1-r} = \frac{1\left[1-\left(-\frac{1}{2}\right)^{11}\right]}{1+\frac{1}{2}}$$

$$= \frac{2^{11}+1}{3\times 2^{10}} = \frac{683}{1024}.$$

To insert *n* geometric means between two given numbers *a* and *b*

Let $G_1, G_3, \ldots G_n$ be n geometric means between a and b. Thus, $a, G_1, G_2, \ldots G_n, b$ is a GP, b being $(n+2)$th term $= ar^{n+1}$, where r is the common ratio of GP

$$\text{Thus,} \quad b = ar^{n+1} \Rightarrow r = \left(\frac{b}{a}\right)^{\frac{1}{n+1}}$$

$$\text{So,} \quad G_1 = ar = a\left(\frac{b}{a}\right)^{\frac{1}{n+1}} = \left(a^n b\right)^{\frac{1}{n+1}}$$

$$G_2 = ar^2 = a\left(\frac{b}{a}\right)^{\frac{2}{n+1}} = \left(a^{n-1}b^2\right)^{\frac{1}{n+1}}$$

... .

$$G_n = ar^{n-1} = a\left(\frac{b}{a}\right)^{\frac{n-1}{n+1}} = \left(a^2 b^{n-1}\right)^{\frac{1}{n+1}}$$

Example 4.10: Find 7 GM's between 1 and 256.

Solution: Let $G_1, G_2, \ldots G_7$, be 7 GM's between 1 and 256

Then, 256= 9th term of GP,

$= 1.\, r^8$, where r is the common ratio of the GP

NOTES

This gives that, $r^8 = 256 \Rightarrow r = 2$

Thus,
$$G_1 = ar = 1.2 = 2$$
$$G_2 = ar^2 = 1.4 = 4$$
$$G_3 = ar^3 = 1.8 = 8$$
$$G_4 = ar^4 = 1.16 = 16$$
$$G_5 = ar^5 = 1.32 = 32$$
$$G_6 = ar^6 = 1.64 = 64$$
$$G_7 = ar^7 = 1.128 = 128$$

Hence, required GM's are 2, 4, 8, 16, 32, 64, 128.

Example 4.11: Sum the series $1 + 3x + 5x^2 + 7x^3 + ...$ up to n terms, $x \neq 1$.

Solution: Note that nth term of this series $= (2n - 1)\, x^{n-1}$

Let $S_n = 1 + 3x + 5x^2 + ... + (2n - 1)\, x^{n-1}$

Then, $xS_n = x + 3x^2 + ... + (2n - 3)\, x^{n-1} + (2n - 1)\, x^n$

Subtracing, we get

$$S_n(1 - x) = 1 + 2x + 2x^2 + ... + 2x^{n-1} - (2n - 1)\, x^n$$

$$= 1 + 2x \left(\frac{1 - x^{n-1}}{1 - x} \right) - (2n - 1)\, x^n$$

$$= \frac{1 - x + 2x - 2x^n - (2n - 1)\, x^n (1 - x)}{1 - x}$$

$$= \frac{1 + x - 2x^n - (2n - 1)\, x^n + (2n - 1)\, x^{n+1}}{1 - x}$$

$$= \frac{1 + x - (2n + 1)\, x^n + (2n - 1)\, x^{n+1}}{1 - x}$$

Hence, $$S = \frac{1 + x - (2n + 1)\, x^n + (2n - 1)\, x^{n+1}}{(1 - x)^2}$$

Example 4.12: If in a GP $(p + q)$th term $= m$ and $(p - q)$th term $= n$, then find its pth and qth terms.

Solution: Suppose that the given GP be $a, ar, ar^2, ar^3, \ldots$

By hypothesis, $(p + q)$th term $= m = ar^{p+q-1}$

$(p - q)$th term $= n = ar^{p-q-1}$

Then, $$\frac{m}{n} = r^{2q} \Rightarrow r = \left(\frac{m}{n} \right)^{1/2q}$$

Hence, $$m = a \left(\frac{m}{n} \right)^{(p+q-1)/2q} \Rightarrow a = m^{(q-p+1)/2q}\, n^{(p+q-1)/2q}$$

Thus, $p\text{th term} = ar^{p-1} = m^{1/2}\, n^{1/2} = \sqrt{mn}$

$$q\text{th term} = ar^{q-1} = m^{\frac{2q-p}{2q}}\, n^{\frac{p}{2q}}$$

NOTES

Example 4.13: Sum the series $5 + 55 + 555 + \ldots$ up to n terms.

Solution: Let $S_n = 5 + 55 + 555 + \ldots.$

$$S_n = 5\,(1 + 11 + 111 + \ldots.)$$

$$= \frac{5}{9}\,(9 + 99 + 999 + \ldots)$$

$$= \frac{5}{9}\,[(10-1) + (100-1) + (1000-1) + \ldots]$$

$$= \frac{5}{9}\,[(10 + 10^2 + 10^3 + \ldots + 10^n) - (1 + 1 + \ldots. n \text{ terms})]$$

$$= \frac{5}{9}\,[(10 + 10^2 + 10^3 + \ldots + 10^n) - n]$$

$$= \frac{5}{9}\left[\frac{10(1-10^n)}{1-10} - n\right]$$

$$= \frac{5}{9}\left[\frac{10(10^n - 1)}{9} - n\right]$$

$$= \frac{50}{81}(10^n - 1) - \frac{5n}{9}$$

Example 4.14: If a, b, c, d are in GP, prove that $a^2 - b^2$, $b^2 - c^2$ and $c^2 - d^2$ are also in GP.

Solution: Since, $\frac{b}{a} = \frac{c}{b} = \frac{d}{c} = k$ (say)

we have, $b = ak,\ c = bk,\ d = ck$

i.e., $b = ak,\ c = ak^2,\ d = ak^3$

Now, $(b^2 - c^2)^2 = (a^2k^2 - a^2k^4)^2$

$= a^4k^4(1 - k^2)^2$

Also, $(a^2 - b^2)\,(c^2 - d^2) = (a^2 - a^2k^2)\,(a^2k^4 - a^2k^6)$

$= a^4(1 - k^2)\,(k^4 - k^6)$

$= a^4k^4\,(1 - k^2)^2$

Hence, $(b^2 - c^2) = (a^2 - b^2)\,(c^2 - d^2)$

This gives that, $a^2 - b^2, b^2 - c^2, c^2 - d^2$ are in GP.

Example 4.15: Three numbers are in GP. Their product is 64 and sum is $\frac{124}{5}$. Find them.

Solution: Let the numbers be $\frac{a}{r}, a, ar$

Since, $\frac{a}{r} + a + ar = \frac{124}{5}$ and $\frac{a}{r} \times a \times ar = 64$,

we have, $a^3 = 64 \Rightarrow a = 4$

NOTES

This gives, $\frac{4}{r} + 4 + 4r = \frac{124}{5}$

$\Rightarrow \quad \frac{1}{r} + 1 + r = \frac{31}{5}$

$\Rightarrow \quad \frac{r^2 + 1}{r} = \frac{26}{5}$

$\Rightarrow \quad 5r^2 + 5 = 26r$

$\Rightarrow \quad 5r^2 - 26r + 5 = 0$

$\Rightarrow \quad 5r^2 - 25r - r + 5 = 0$

$\Rightarrow \quad 5r(r-5) - 1(r-5) = 0$

$\Rightarrow \quad (r-5)(5r-1) = 0$

$\Rightarrow \quad r = \frac{1}{5} \quad \text{or} \quad 5$

In either case, numbers are $\frac{4}{5}$, 4 and 20.

Example 4.16: If *a, b, c* are in GP and $a^x = b^y = c^z$, prove that

$$\frac{1}{x} + \frac{1}{z} = \frac{2}{y}$$

Solution: *a*, *b*, *c* are in GP, $b^2 = ac$

But, $b^y = a^x \quad \Rightarrow \quad a = b^{y/x}$

and, $b^y = c^z \quad \Rightarrow \quad c = b^{y/z}$

So, we get $b^2 = b^{y/x} \cdot b^{y/z}$

$$= b^{y\left(\frac{1}{x} + \frac{1}{z}\right)}$$

$\Rightarrow \quad 2 = y\left(\frac{1}{x} + \frac{1}{z}\right)$

$\Rightarrow \quad \frac{1}{x} + \frac{1}{z} = \frac{2}{y}$

Example 4.17: Sum to *n* terms the series

$$0.7 + 0.77 + 0.777 + \ldots$$

Solution: Given series,

$$= 0.7 + 0.77 + 0.777 + \ldots \text{ up to } n \text{ terms}$$

$$= 7\,(0.1 + 0.11 + 0.111 + \ldots \text{ up to } n \text{ terms})$$

$$= \frac{7}{9}\,(0.9 + 0.99 + 0.999 + \ldots \text{ up to } n \text{ terms})$$

$$= \frac{7}{9}\left[\left(1 - \frac{1}{10}\right) + \left(1 - \frac{1}{10^2}\right) + \left(1 - \frac{1}{10^3}\right) + \ldots\right]$$

$$= \frac{7}{9}\left[n - \frac{1}{10} - \frac{1}{10^2} - \ldots \text{ up to } n \text{ terms}\right]$$

NOTES

$$= \frac{7}{9}\left[n - \frac{\frac{1}{10}(1 - 1/10^n)}{1 - \frac{1}{10}} \right]$$

$$= \frac{7}{9}\left[n - \frac{1}{9}\left(1 - \frac{1}{10^n}\right) \right]$$

$$= \frac{7}{9}\left[n - \frac{1}{9}\left(1 - \frac{1}{10^n}\right) \right]$$

Example 4.18: The sum of three numbers in GP is 35 and their product is 1000. Find the numbers.

Solution: Let the numbers be $\frac{\alpha}{r}$, α, αr

The product of $\frac{\alpha}{r} \times \alpha \times \alpha r = 1000$

$$\alpha^3 = 1000$$

$$\Rightarrow \quad \alpha = 10$$

So, the numbers are $\frac{10}{r}$, 10, $10r$

The sum of these numbers = 35

$$\Rightarrow \quad \frac{10}{r} + 10 + 10r = 35$$

$$\Rightarrow \quad \frac{2}{r} + 2r = 5$$

$$\Rightarrow \quad 2r^2 - 5r + 2 = 0$$

$$\Rightarrow \quad (2r - 1)(r - 2) = 0$$

$$\Rightarrow \quad r = 2 \quad \text{or} \quad \frac{1}{2}$$

$r = 2$ gives the numbers as 5, 10, 20

$r = \frac{1}{2}$, gives the numbers as 20, 10, 5, the same as the first set.

Hence, the required numbers are 5, 10 and 20.

Example 4.19: The sum of the first eight terms of a GP (of real terms) is five times the sum of the first four terms. Find the common ratio.

Solution: Let the GP be a, ar, ar^2, . . .

$$S_8 = \text{Sum of first eight terms} = \frac{a(1 - r^8)}{1 - r}$$

$$S_4 = \text{Sum of first four terms} = \frac{a(1 - r^4)}{1 - r}$$

By hypothesis, $\quad S_8 = 5S_4 \Rightarrow \frac{a(1 - r^8)}{1 - r} = \frac{5a(1 - r^4)}{1 - r}$

NOTES

$\Rightarrow \quad 1 - r^8 = 5(1 - r^4)$

$\Rightarrow \quad (1 - r^4)(1 + r^4) = 5(1 - r^4)$

In case, $r^4 - 1 = 0$ we get, $r^2 - 1 = 0 \Rightarrow r = \pm 1$

(Note that $r^2 + 1 = 0 \Rightarrow r$ is imaginary)

Now, $r = 1 \Rightarrow$ The given series is $a + a + a + \ldots$

but, $S_8 = 8a$ and $S_4 = 4a$

So, $S_8 \neq 5S_4$

In case $r = -1$, we get, $S_8 = 0$ and $S_4 = 0$, hence the hypothesis is satisfied.

Suppose now, $r^4 - 1 \neq 0$, then $1 + r^4 = 5$

$\Rightarrow \quad r^4 = 4 \Rightarrow r^2 = 2 \quad (r^2 \neq -2)$

$\Rightarrow \quad r = \pm\sqrt{2}$

Hence, $r = -1$ or $\pm\sqrt{2}$

Example 4.20: If S is the sum, P the product of n term of G.P. and R the sum of reciprocals of n terms in GP, then prove that

$$P^2R^n = S^n.$$

Solution: Let $a, ar, ar^2, \ldots$ be the given GP

Then, $S = a + ar + ar^2 + \ldots$ up to n terms

$$= \frac{a(1 - r^n)}{1 - r} \qquad \ldots(1)$$

$$P = a \cdot ar \cdot ar^2 \ldots . ar^{n-1}$$

$$= a^n \, r^{1 + 2 + 3 + \ldots + (n-1)}$$

$$= a^n \, r^{\frac{(n-1)}{2}(2 + n - 2)}$$

$$= a^n r^{\left(\frac{n-1}{2}\right)n} \qquad \ldots(2)$$

$$R = \frac{1}{a} + \frac{1}{ar} + \frac{1}{ar^2} + \ldots \quad \text{up to } n \text{ terms}$$

$$= \frac{\frac{1}{a}\left(1 - \frac{1}{r^n}\right)}{1 - \frac{1}{r}} = \frac{r}{a} \frac{(r^n - 1)}{(r - 1) r^n}$$

$$= \frac{(1 - r^n)}{a(1 - r) r^{n-1}} \qquad \ldots(3)$$

By Equations (2) and (3),

$$P^2R^n = a^{2n} \, r^{n(n-1)} \, \frac{(1 - r^n)^n}{a^n (1 - r)^n \, r^{n(n-1)}}$$

$$= \frac{a^n (1 - r^n)^n}{(1 - r)^n} = S^n, \text{ by (1)}$$

NOTES

Example 4.21: The ratio of the 4th to the 12th term of a GP with positive common ratio is $\frac{1}{256}$. If the sum of the two terms is 61.68, find the sum of series to 8 terms.

Solution: Let the series be $a, ar, ar^2, \ldots,$

$$T_4 = 4\text{th term} = ar^3$$

$$T_{12} = 12\text{th term} = ar^{11}$$

By hypothesis, $\frac{T_4}{T_{12}} = \frac{1}{256}$

i.e., $\frac{ar^3}{ar^{11}} = \frac{1}{256}$

$$\frac{1}{r^8} = \frac{1}{256}$$

$\Rightarrow$ $r^8 = 256$

$\Rightarrow$ $r = \pm 2$

Since r is given to be positive, we reject negative sign.

Again, it is given that

$$T_4 + T_{12} = 61.68$$

i.e., $a(r^3 + r^{11}) = 61.68$

$$a(8 + 2048) = 61.68$$

$$a = \frac{61.68}{2056} = 0.03$$

Hence, $S_8 = \text{Sum to eight terms}$

$$= \frac{a(1 - r^8)}{1 - r} = \frac{a(r^8 - 1)}{r - 1}$$

$$= \frac{(0.03)(256 - 1)}{(2 - 1)} = 0.03 \times 255 = 7.65$$

Example 4.22: A manufacturer reckons that the value of a machine which costs him Rs 18750 will depreciate each year by 20%. Find the estimated value at the end of 5 years.

Solution: At the end of first year the value of machine is

$$= 18750 \times \frac{80}{100} = \frac{4}{5}(18750)$$

At the end of 2nd year it is equal to $\left(\frac{4}{5}\right)^2 (18750)$; proceeding in this manner, the estimated value of machine at the end of 5 years is $\left(\frac{4}{5}\right)^5 (18750)$

$$= \frac{64 \times 16}{125 \times 25} \times 18750$$

$$= \frac{1024}{125} \times 750 = 1024 \times 6$$
$$= 6144 \text{ rupees}$$

NOTES

Example 4.23: Show that a given sum of money accumulated at 20 % per annum, more than doubles itself in 4 years at compound interest.

Solution: Let the given sum be *a* rupees. After 1 year it becomes $\frac{6a}{5}$ (it is increased by $\frac{a}{5}$).

At the end of two years it becomes $\frac{6}{5}\left(\frac{6a}{5}\right) = \left(\frac{6}{5}\right)^2 a$

Proceeding in this manner, we get that at the end of 4th year, the amount will be $\left(\frac{6}{5}\right)^4 a = \frac{1296}{625}\ a$

Now, $\frac{1296}{625} a - 2a = \frac{46}{625} a,$ since *a* is a + ve quantity, so the amount after 4 years is more than double of the original amount.

Example 4.24: If $x = a + \frac{a}{r} + \frac{a}{r^2} + \ldots \infty$

$$y = b - \frac{b}{r} + \frac{b}{r^2} + \ldots \infty$$

and $z = c + \frac{c}{r^2} + \frac{c}{r^4} + \ldots \infty$

Show that $\frac{xy}{z} = \frac{ab}{c}$

Solution: Clearly, $x = \frac{a}{1 - \frac{1}{r}} = \frac{ar}{r-1},$

$$y = \frac{b}{1 - (-1/r)} = \frac{br}{r+1}$$

and, $z = \frac{c}{1 - \frac{1}{r^2}} = \frac{cr^2}{r^2 - 1}$

Now, $\frac{xy}{z} = \frac{ab\,r^2}{(r^2-1)} \Big/ \left(\frac{cr^2}{r^2-1}\right) = \frac{ab}{c}$

Example 4.25: If $a^2 + b^2$, $ab + bc$ and $b^2 + c^2$ are in GP, prove that *a, b, c* are also in GP.

Solution: Since $a^2 + b^2$, $ab + bc$ and $b^2 + c^2$ are in GP, we get,

$$(ab + bc)^2 = (a^2 + b^2)\ (b^2 + c^2)$$

NOTES

$$b^2(a^2 + 2ac + c^2) = a^2b^2 + a^2c^2 + b^4 + b^2c^2$$

$$\Rightarrow \quad 2ab^2c = a^2c^2 + b^4$$

$$\Rightarrow \quad a^2c^2 - 2ab^2c + b^4 = 0$$

$$\Rightarrow \quad (ac - b^2)^2 = 0$$

$$\Rightarrow \quad ac = b^2$$

$\Rightarrow$ a, b, c are in GP.

Harmonic mean

If a, b, c are in HP, then b is called a *Harmonic Mean* between a and c, written as HM.

Harmonical progression

Non zero quantities whose reciprocals are in AP, or Arithmetic Progression are said to be in *Harmonical Progression*, written as HP.

Consider the following examples:

(*a*) $1, \frac{1}{3}, \frac{1}{5}, \frac{1}{7}, \ldots\ldots$

(*b*) $\frac{1}{2}, \frac{1}{5}, \frac{1}{8}, \frac{1}{11}, \ldots\ldots$

(*c*) $2, \frac{5}{2}, \frac{10}{3}, \ldots$

(*d*) $\frac{1}{a}, \frac{1}{a+b}, \frac{1}{a+2b}, \ldots\ldots \quad a, b > 0$

(*e*) $5, \frac{55}{9}, \frac{55}{7}, 11, \ldots\ldots$

It can be easily checked that in each case, the series obtained by taking reciprocal of each of the term is an AP.

To insert *n* harmonic means between *a* and *b*

Let $H_1, H_2, H_3, \ldots, H_n$ be the required Harmonic Means. Then,
$a, H_1, H_2, \ldots, H_n, b$ are in HP

i.e., $\frac{1}{a}, \frac{1}{H_1}, \frac{1}{H_2}, \ldots, \frac{1}{H_n}, \frac{1}{b}$ are in AP

Then, $\frac{1}{b} = (n+2)$th term of an AP

$$= \frac{1}{a} + (n+1)d$$

Where d is the common difference of AP.

This gives, $d = \frac{a-b}{(n+1)ab}$

NOTES

Now, $$\frac{1}{H_1} = \frac{1}{a} + d = \frac{1}{a} + \frac{a-b}{(n+1)\,ab}$$

$$= \frac{nb + b + a - b}{(n+1)\,ab} = \frac{a + nb}{(n+1)\,ab}$$

So, $$\frac{1}{H_1} = \frac{a + nb}{(n+1)\,ab}$$

$\Rightarrow$ $$H_1 = \frac{(n+1)\,ab}{a + nb}$$

Again, $$\frac{1}{H_2} = \frac{1}{a} + 2d = \frac{1}{a} + \frac{2\,(a-b)}{(n+1)\,ab}$$

$$= \frac{nb + b + 2a - 2b}{(n+1)\,ab} = \frac{2a - b + nb}{(n+1)\,ab}$$

$\Rightarrow$ $$H_2 = \frac{(n+1)\,ab}{2a - b + nb}$$

Similarly, $$\frac{1}{H_3} = \frac{1}{a} + 3d = \frac{3a - 2b + nb}{(n+1)\,ab}$$

$\Rightarrow$ $$H_3 = \frac{(n+1)\,ab}{3a - 2b + nb}$$ and so on,

$$\frac{1}{H_n} = \frac{1}{a} + nd = \frac{1}{a} + \frac{n\,(a-b)}{(n+1)\,ab}$$

$$= \frac{nb + b + na - nb}{(n+1)\,ab}$$

$$= \frac{na + b}{(n+1)\,ab} \Rightarrow H_n = \frac{(n+1)\,ab}{na + b}$$

Example 4.26: Find the 5th term of $2, 2\frac{1}{2}, 3\frac{1}{3}, \ldots\ldots$

Solution: Let 5th term be x. Then, $\frac{1}{x}$ is 5th term of corresponding AP $\frac{1}{2}, \frac{2}{5}, \frac{3}{10}, \ldots\ldots$

Then, $$\frac{1}{x} = \frac{1}{2} + 4\left(\frac{2}{5} - \frac{1}{2}\right) = \frac{1}{2} + 4\left(\frac{-1}{10}\right)$$

$\Rightarrow$ $$\frac{1}{x} = \frac{1}{2} - \frac{2}{5} = \frac{1}{10} \Rightarrow x = 10$$

Example 4.27: Insert two harmonic means between $\frac{1}{2}$ and $\frac{4}{17}$.

Solution: Let H_1, H_2 be two harmonic means between $\frac{1}{2}$ and $\frac{4}{17}$.

Thus, $2, \frac{1}{H_1}, \frac{1}{H_2}, \frac{17}{4}$ are in AP Let d be their common difference.

Then, $$\frac{17}{4} = 2 + 3d$$

NOTES

$$\Rightarrow \qquad 3d = \frac{9}{4} \quad \Rightarrow \quad d = \frac{3}{4}$$

Thus,
$$\frac{1}{H_1} = 2 + \frac{3}{4} = \frac{11}{4} \quad \Rightarrow \quad H_1 = \frac{4}{11}$$

$$\frac{1}{H_2} = 2 + 2 \times \frac{3}{4} = \frac{7}{2} \quad \Rightarrow \quad H_2 = \frac{2}{7}$$

Required harmonic means are $\frac{4}{11}, \frac{2}{7}$.

Coefficient of variation

The square of standard deviation, namely σ^2, is termed as variance and is more often specified than the standard deviation. Clearly, it has the same properties as standard deviation.

As is clear, the standard deviation σ or its square, the variance, cannot be very useful in comparing two series where either the units are different or the mean values are different. Thus, a σ of 5 on an examination where the mean score is 30 has an altogether different meaning than on an examination where the mean score is 90. Clearly, the variability in the second examination is much less. To take care of this problem, we define and use a coefficient of variation, *V*. Where,

$$V = \frac{\sigma}{\bar{x}} \times 100$$

This is expressed as percentage.

Example 4.28: The following are the scores of two batsmen A and B in a series of innings:

A	12	115	6	73	7	19	119	36	84	29
B	47	12	76	42	4	51	37	48	13	0

Who is the better run-getter? Who is more consistent?

Solution: In order to decide as to which of the two batsmen, *A* and *B*, is the better run-getter, we should find their batting averages. The one whose average is higher will be considered as a better batsman.

To determine the consistency in batting we should determine the coefficient of variation. The less this coefficient the more consistent will be the player.

A			*B*		
Score x	x	x^2	*Scores* x	x	x^2
12	–38	1,444	47	14	196
115	+65	4,225	12	–21	441
6	–44	1,936	76	43	1,849
73	+23	529	42	9	81
7	–43	1,849	–4	– 29	841
19	–31	961	51	18	324

NOTES

119	+69	4,761	37	4	16
36	–14	196	48	15	225
84	+34	1,156	13	–20	400
29	–21	441	0	–33	1,089
$\sum x = 500$		17,498	$\sum x = 330$		5,462

Batsman *A*:

$$\bar{x} = \frac{500}{10} = 50$$

$$\sigma = \sqrt{\frac{17,498}{10}} = 41.83$$

$$V = \frac{41.83 \times 100}{50}$$

$$= 83.66 \text{ per cent}$$

Batsman *B*:

$$\bar{x} = \frac{330}{10} = 33$$

$$\sigma = \sqrt{\frac{5,462}{10}} = 23.37$$

$$V = \frac{23.37}{33} \times 100$$

$$= 70.8 \text{ per cent}$$

A is a better batsman since his average is 50 as compared to 33 of *B*, but *B* is more consistent since the variation in his case is 70.8 as compared to 83.66 of *A*.

Example 4.29: The following table gives the age distribution of students admitted to a college in the years 1914 and 1918. Find which of the two groups is more variable in age.

Age	*Number of Students in*	
	1914	*1918*
15	–	1
16	1	6
17	3	34
18	8	22
19	12	35
20	14	20
21	13	7
22	5	19
23	2	3
24	3	–
25	1	–
26	–	–
27	1	–

Solution:

Age	*Assumed Mean–21 1914*				*Assumed Mean–19 1918*			
	f	*x′*	*fx′*	*fx′²*	*f*	*x′*	*fx*	*fx′²*
15	0	–6	0	0	1	–4	–4	16
16	1	–5	–5	25	6	–3	–18	54
17	3	–4	–12	48	34	–2	–68	136
18	8	–3	–24	72	22	–1	–22	22

NOTES

19	12	–2	–24	48			–112	
20	14	–1	–14	14				
			–79		35	0	0	0
21	13	0	0	0	20	1	20	20
22	5	1	5	5	7	2	14	28
23	2	2	4	8	19	3	57	171
24	3	3	9	27	3	4	12	48
25	1	4	4	16	147		+103	495
26	0	5	0	0			–9	
27	1	6	6	36				
	63		+28	299				
			–51					

1914 Group:

$$\sigma = \sqrt{\frac{\sum fx'^2}{N} - \left[\frac{\sum (fx')}{N}\right]^2}$$

$$= \sqrt{\frac{299}{63} - \left(\frac{-51}{63}\right)^2}$$

$$= \sqrt{4.476 - 0.655} = \sqrt{4.091}$$

$$= 2.02.$$

$$\bar{x} = 21 + \left(\frac{-51}{63}\right) = 21 - 8 = 20.2$$

$$V = \frac{2.02}{20.2} \times 100$$

$$= \frac{202}{20.2} = 10$$

1918 Group:

$$\sigma = \sqrt{\frac{495}{147} - \left(\frac{-9}{147}\right)^2} = \sqrt{3.3673 - 0.0037}$$

$$= \sqrt{3.3636} = 1.834$$

$$\bar{x} = 19 + \left(\frac{-9}{147}\right)$$

$$= 19 - .06 = 18.94$$

$$V = \frac{1.834}{18.94} \times 100$$

$$= 9.68$$

The coefficient of variation of the 1914 group is 10 and that of the 1918 group 9.68. This means that the 1914 group is more variable, but only barely so.

NOTES

CHECK YOUR PROGRESS

8. How did Tate define measures of central tendency?
9. Define median.
10. What is mode?

4.5 CO-EFFICIENT OF CORRELATION

Correlation analysis is the statistical tool generally used to describe the degree to which one variable is related to another. The relationship, if any, is usually assumed to be a linear one. This analysis is used quite frequently in conjunction with regression analysis to measure how well the regression line explains the variations of the dependent variable. In fact, the word correlation refers to the relationship or interdependence between two variables. There are various phenomena which are related to each other. For instance, when demand of a certain commodity increases, its price goes up and when its demand decreases, its price comes down. Similarly, with age the height of children, with height the weight of children, with money the supply and the general level of prices go up. Such sort of relationships can as well be noticed for several other phenomena. The theory by means of which quantitative connections between two sets of phenomena are determined is called the 'Theory of Correlation'.

On the basis of the theory of correlation, one can study the comparative changes occurring in two related phenomena and their cause–effect relation can be examined. It should, however, be borne in mind that relationships like 'black cat causes bad luck', 'filled up pitchers result in good fortune' and similar other beliefs of the people cannot be explained by the theory of correlation, since they are all imaginary and are incapable of being justified mathematically. Thus, correlation is concerned with relationship between two related and quantifiable variables. If two quantities vary in sympathy, so that a movement (an increase or decrease) in one tends to be accompanied by a movement in the same or opposite direction in the other and the greater the change in one, the greater is the change in the other, the quantities are said to be correlated. This type of relationship is known as correlation or what is sometimes called, in statistics, as covariation.

For correlation, it is essential that the two phenomena should have cause–effect relationship. If such relationship does not exist then one should not talk of correlation. For example, if the height of the students as well as the height of the trees increases, then one should not call it a case of correlation because the two phenomena, viz., the height of students and the height of trees are not even casually related. However, the relationship between the price of a commodity and its demand, the price of a commodity and its supply, the rate of interest and savings, etc. are examples of correlation, since in all such cases the change in one phenomenon is explained by a change in another phenomenon.

NOTES

It is appropriate here to mention that correlation in case of phenomena pertaining to natural sciences can be reduced to absolute mathematical term, e.g., heat always increases with light. However, in phenomena pertaining to social sciences, it is often difficult to establish any absolute relationship between two phenomena. Hence, in social sciences, we must take the fact of correlation being established if in a large number of cases, two variables always tend to move in the same or opposite direction.

Correlation can either be positive or it can be negative. Whether correlation is positive or negative would depend upon the direction in which the variables are moving. If both variables are changing in the same direction, then correlation is said to be positive, but when the variations in the two variables take place in opposite direction, the correlation is termed as negative. This can be explained as follows:

Changes in Independent Variable	*Changes in Dependent Variable*	*Nature of Correlation*
Increase (+)↑	Increase (+)↑	Positive (+)
Decrease (–)↓	Decrease (–)↓	Positive (+)
Increase (+)↑	Decrease (–)↓	Negative (–)
Decrease (–)↓	Increase (+)↑	Negative (–)

Statisticians have developed two measures for describing the correlation between two variables, viz., the coefficient of determination and the coefficient of correlation. These two methods are explained in detail in the following sections.

Methods of computing

The following are the various methods of computing:

I. Coefficient of determination

The **coefficient of determination** (symbolically indicated as r^2, though some people would prefer to put it as R^2) is a measure of the degree of linear association or correlation between two variables, say X and Y, one of which happens to be an independent variable and the other being a dependent variable. This coefficient is based on the following two types of variations:

(a) The variation of the Y values around the fitted regression line, viz., $\sum\left(Y-\hat{Y}\right)^2$, technically known as the unexplained variation.

(b) The variation of the Y values around their own mean, viz., $\sum\left(Y-\overline{Y}\right)^2$, technically known as the total variation.

If we subtract the unexplained variation from the total variation, we obtain what is known as the explained variation, i.e., the variation explained by the line of regression. Thus, Explained Variation = (Total variation) – (Unexplained variation)

$$= \sum\left(Y-\overline{Y}\right)^2 - \sum\left(Y-\hat{Y}\right)^2$$

$$= \sum\left(\hat{Y}-\overline{Y}\right)^2$$

The Total and Explained as well as Unexplained variations can be shown as given in Figure. 4.2.

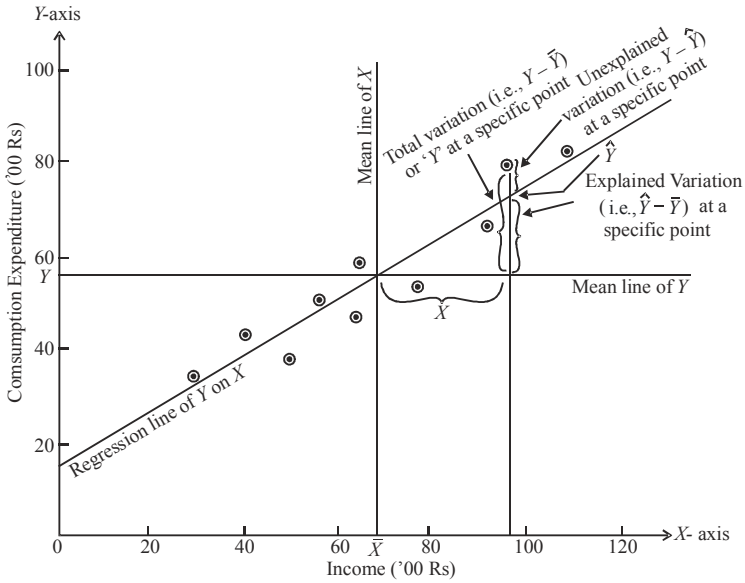


Fig. 4.2 *Diagram Showing Total, Explained and Unexplained Variations*

Coefficient of determination is that fraction of the total variation of *Y* which is explained by the regression line. In other words, coefficient of determination is the ratio of explained variation to total variation in the *Y* variable related to the *X* variable. Coefficient of determination can be algebraically stated as,

$$r^2 = \frac{\text{Explained variation}}{\text{Total variation}}$$

$$= \frac{\Sigma\left(\hat{Y} - \bar{Y}\right)^2}{\Sigma\left(Y - \bar{Y}\right)^2}$$

Alternatively, r^2 can also be stated as,

$$r^2 = 1 - \frac{\text{Explained variation}}{\text{Total variation}}$$

$$= 1 - \frac{\Sigma\left(\hat{Y} - \bar{Y}\right)^2}{\Sigma\left(Y - \bar{Y}\right)^2}$$

Interpreting r^2

The coefficient of determination can have a value ranging from 0 – 1. The value of 1 can occur only if the unexplained variation is 0, which simply means that all the data points in the Scatter diagram fall exactly on the regression line. For a 0 value to occur, $\Sigma(Y - \bar{Y})^2 = \Sigma(Y - \hat{Y})^2$, which simply means that *X* tells us nothing about *Y* and hence there is no regression relationship between *X* and *Y* variables. Values between 0 and 1 indicate the 'Goodness of fit' of the regression line to the sample data. The higher the value of r^2, the better the fit. In other words, the value of r^2 will lie

NOTES

NOTES

somewhere between 0 and 1. If r^2 has a 0 value then it indicates no correlation, but if it has a value equal to 1 then it indicates that there is perfect correlation and as such the regression line is a perfect estimator. However, in most cases, the value of r^2 will lie somewhere between these two extremes of 1 and 0. One should remember that r^2 close to 1 indicates a strong correlation between X and Y, while an r^2 near 0 means there is little correlation between these two variables. r^2 value can as well be interpreted by looking at the amount of the variation in Y, the dependant variable, that is explained by the regression line. Supposing, we get a value of $r^2 = 0.925$ then this would mean that the variations in independent variable (say X) would explain 92.5 per cent of the variation in the dependent variable (say Y). If r^2 is close to 1 then it indicates that the regression equation explains most of the variations in the dependent variable (see Example 4.30).

Example 4.30: Calculate the coefficient of determination (r^2) using the provided data. Calculate and analyse the result.

Observations	1	2	3	4	5	6	7	8	9	10
Income (X) ('00 ₹)	41	65	50	57	96	94	110	30	79	65
Consumption Expenditure (Y) ('00 ₹)	44	60	39	51	80	68	84	34	55	48

Solution:

r^2 can be worked out as follows:

Since, $$r^2 = 1 - \frac{\text{Unexplained variation}}{\text{Total variation}} = 1 - \frac{\sum\left(Y - \hat{Y}\right)^2}{\sum\left(Y - \overline{Y}\right)^2}$$

As, $\sum\left(Y - \overline{Y}\right)^2 = \sum Y^2 = \left(\sum Y^2 - n\overline{Y}^2\right)$, we can write,

$$r^2 = 1 - \frac{\sum\left(Y - \hat{Y}\right)^2}{\sum Y^2 - n\overline{Y}^2}$$

Calculating and putting the various values, we have the following equation:

$$r^2 = 1 - \frac{260.54}{34223 - 10(56.3)^2} = 1 - \frac{260.54}{2526.10} = 0.897$$

Analysis of Result: The regression equation used to calculate the value of the coefficient of determination (r^2) from the sample data shows that, about 90 per cent of the variations in consumption expenditure can be explained. In other words, it means that the variations in income explain about 90 per cent of variations in consumption expenditure.

Observation	1	2	3	4	5	6	7	8	9	10
Income (X) ('00 ₹)	41	65	50	57	96	94	110	30	79	65
Consumption Expenditure (Y) ('00 ₹)	44	60	39	51	80	68	84	34	55	48

NOTES

Properties of correlation coefficient

The coefficient of correlation, symbolically denoted by '*r*', is another important measure to describe how well one variable is explained by another. It measures the degree of relationship between the two casually related variables. The value of this coefficient can never be more than +1 or less than –1. Thus, +1 and –1 are the limits of this coefficient. For a unit change in independent variable, if there happens to be a constant change in the dependent variable in the same direction, then the value of the coefficient will be +1 indicative of the perfect positive correlation; but if such a change occurs in the opposite direction, the value of the coefficient will be –1, indicating the perfect negative correlation. In practical life, the possibility of obtaining either a perfect positive or perfect negative correlation is very remote particularly in respect of phenomena concerning social sciences. If the coefficient of correlation has a zero value then it means that there exists no correlation between the variables under study.

There are several methods of finding the coefficient of correlation, but the following ones are considered important:

(a) Coefficient of correlation by the method of least squares.

(b) Coefficient of correlation using simple regression coefficients.

(c) Coefficient of correlation through product moment method or Karl Pearson's coefficient of correlation.

Whichever of these three methods we adopt, we get the same value of *r*.

II. Method of Least Squares

Under this method, first, the estimating equation is obtained using the least square method of simple regression analysis. The equation is worked out as,

$$\hat{Y} = a + bX_i$$

$$\text{Total variation} = \sum \left(Y - \overline{Y}\right)^2$$

$$\text{Unexplained variation} = \sum \left(Y - \hat{Y}\right)^2$$

$$\text{Explained variation} = \sum \left(\hat{Y} - \overline{Y}\right)^2$$

Then, by applying the following formulae, we can find the value of the coefficient of correlation as,

$$r = \sqrt{r^2} = \sqrt{\frac{\text{Explained variation}}{\text{Total variation}}}$$

$$= \sqrt{1 - \frac{\text{Unexplained variation}}{\text{Total variation}}}$$

$$= \sqrt{1 - \frac{\sum \left(Y - \hat{Y}\right)^2}{\sum \left(Y - \overline{Y}\right)^2}}$$

NOTES

This clearly shows that the coefficient of correlation happens to be the square root of the coefficient of determination.

Short-cut formula for finding the value of '*r*' by the method of least squares can be repeated and readily written as,

$$r = \sqrt{\frac{a\Sigma Y + b\Sigma XY - n\overline{Y}^2}{\Sigma Y^2 - n\overline{Y}^2}}$$

Where,

a = *Y*-intercept

b = Slope of the estimating equation

X = Values of the independent variable

Y = Values of dependent variable

$\overline{Y}$ = Mean of the observed values of Y

n = Number of items in the sample (i.e., pairs of observed data)

The plus (+) or the minus (–) sign of the coefficient of correlation worked out by the method of least squares, is related to the sign of '*b*' in the estimating equation, viz., $\hat{Y} = a + bX_i$. If '*b*' has a minus sign, the sign of '*r*' will also be minus, but if '*b*' has a plus sign, then the sign of '*r*' will also be plus. The value of '*r*' indicates the degree along with the direction of the relationship between the two variables *X* and *Y*.

III. Simple Regression Coefficients

Under this method, the estimating equation of *Y* and the estimating equation of *X* is worked out using the method of least squares. From these estimating equations we find the regression coefficient of *X* on *Y*, i.e., the slope of the estimating equation of *X* (symbolically written as b_{XY}) and this happens to be equal to $r\frac{\sigma_X}{\sigma_Y}$ and similarly, we find the regression coefficient of *Y* on *X*, i.e., the slope of the estimating equation of *Y* (symbolically written as b_{YX}) and this happens to be equal to $r\frac{\sigma_Y}{\sigma_X}$. For finding '*r*', the square root of the product of these two regression coefficients are worked out as[1]

$$r = \sqrt{b_{XY}.b_{YX}}$$

$$= \sqrt{r\frac{\sigma_X}{\sigma_Y}.r\frac{\sigma_Y}{\sigma_X}}$$

$$= \sqrt{r^2} = r$$

As stated earlier, the sign of '*r*' will depend upon the sign of the regression coefficients. If they have minus sign, then '*r*' will take minus sign but the sign of '*r*' will be plus if regression coefficients have plus sign.

IV. Other Measures

Two other measures are often talked about along with the coefficients of determinations and that of correlation. These are as follows:

NOTES

(a) Coefficient of Non-Determination: Instead of using coefficient of determination, sometimes coefficient of nondetermination is used. Coefficient of non-determination (denoted by k^2) is the ratio of unexplained variation to total variation in the Y variable related to the X variable. Algebrically,

$$k^2 = \frac{\text{Unexplained variation}}{\text{Total variation}} = \frac{\Sigma\left(Y - \hat{Y}\right)^2}{\Sigma\left(Y - \overline{Y}\right)^2}$$

Concerning the data of Example 1, coefficient of nondetermination will be calculated as follows:

$$k^2 = \frac{260.54}{2526.10} = 0.103$$

The value of k^2 shows that about 10 per cent of the variation in consumption expenditure remains unexplained by the regression equation we had worked out, viz., $\hat{Y} = 14.000 + 0.616X_i$. In simple terms, this means that variable other than X is responsible for 10 per cent of the variations in the dependent variable Y in the given case.

Coefficient of non-determination can as well be worked out as,

$$k^2 = 1 - r^2$$

Accordingly for Example 1, it will be equal to 1–0.897 = 0.103.

Note: Always remember that $r^2 + k^2 = 1$.

(b) Coefficient of Alienation: Based on k^2, we can work out one more measure, namely the coefficient of alienation, symbolically written as 'k'. Thus, coefficient of alienation, i.e., 'k' $= \sqrt{k^2}$.

Unlike $r + k^2 = 1$, the sum of 'r' and 'k' will not be equal to 1 unless one of the two coefficients is 1 and in this case the remaining coefficients must be zero. In all other cases, 'r' + 'k' > 1. Coefficient of alienation is not a popular measure from a practical point of view and is used very rarely.

V. Karl Pearson's Coefficient

Karl Pearson's method is the most widely used method of measuring the relationship between two variables. This coefficient is based on the following assumptions:

(a) There is a linear relationship between the two variables, which means that a straight line would be obtained if the observed data is plotted on a graph.

(b) The two variables are casually related, which means that one of the variables is independent and the other one is dependent.

(c) A large number of independent causes operate on both the variables so as to produce a normal distribution.

According to Karl Pearson, '*r*' can be worked out as,

NOTES

$$r = \frac{\sum XY}{n\sigma_X \sigma_Y}$$

Where,

$X = (X - \bar{X})$

$Y = (Y - \bar{Y})$

σ_X = Standard deviation of X series and is equal to $\sqrt{\frac{\sum X^2}{n}}$

σ_Y = Standard deviation of Y series and is equal to $\sqrt{\frac{\sum Y^2}{n}}$

n = Number of pairs of X and Y observed

A short-cut formula, known as the Product Moment Formula, can be derived from the above stated formula as,

$$r = \frac{\sum XY}{n\sigma_X \sigma_Y}$$

$$= \frac{\sum XY}{\sqrt{\frac{\sum X^2}{n} \cdot \frac{\sum Y^2}{n}}}$$

$$n = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}}$$

The above formulae are based on obtaining true means (viz., $\bar{X}$ and $\bar{Y}$) first and then doing all other calculations. This happens to be a tedious task, particularly if the true means are in fractions. To avoid difficult calculations, we make use of the assumed means in taking out deviations and doing the related calculations. In such a situation, we can use the following formula for finding the value of '*r*':[2]

(a) In Case of Ungrouped Data:

$$r = \frac{\frac{\sum dX.dY}{n} - \left(\frac{\sum dX}{n} \cdot \frac{\sum dY}{n}\right)}{\sqrt{\frac{\sum dX^2}{n} - \left(\frac{\sum dX}{n}\right)^2} \cdot \sqrt{\frac{\sum dY^2}{n} - \cdot \left(\frac{\sum dY}{n}\right)^2}}$$

$$= \frac{\sum dX.dY - \left(\frac{\sum dX \times \sum dY}{n}\right)}{\sqrt{\sum dX^2 - \frac{(\sum dX)^2}{n}} \sqrt{\sum dY^2 - \frac{(\sum dY)^2}{n}}}$$

NOTES

Where, $\Sigma dX = \Sigma(X - X_A)$ X_A = Assumed average of X

$\Sigma dY = \Sigma(Y - Y_A)$ Y_A = Assumed average of Y

$\Sigma dX^2 = \Sigma(X - X_A)^2$

$\Sigma dY^2 = \Sigma(Y - Y_A)^2$

$\Sigma dX \,.\, dY = \Sigma(X - X_A)\,(Y - Y_A)$

n = Number of pairs of observations of X and Y

(b) In Case of Grouped Data:

$$r = \frac{\frac{\Sigma fdX.dY}{n} - \left(\frac{\Sigma fdX}{n} \cdot \frac{\Sigma fdY}{n}\right)}{\sqrt{\frac{\Sigma fdX^2}{n} - \left(\frac{\Sigma fdX}{n}\right)^2}\sqrt{\frac{\Sigma fdY^2}{n} - \left(\frac{\Sigma fdY}{n}\right)^2}}$$

or

$$r = \frac{\Sigma fdX.dY - \left(\frac{\Sigma fdX.\Sigma fdY}{n}\right)}{\sqrt{\Sigma fdX^2 - \left(\frac{\Sigma fdX}{n}\right)^2}\sqrt{\Sigma fdY^2 - \left(\frac{\Sigma fdY}{n}\right)^2}}$$

Where, $\Sigma fdX.dY = 0\Sigma f\,(X - X_A)\,(Y - Y_A)$

$\Sigma fdX = \Sigma f\,(X - X_A)$

$\Sigma fdY = \Sigma f\,(Y - Y_A)$

$\Sigma fdY^2 = \Sigma f\,(Y - Y_A)^2$

$\Sigma fdX^2 = \Sigma f\,(X - X_A)^2$

n = Number of pairs of observations of X and Y

Probable Error (P.E.) of the Coefficient of Correlation

Probable Error (P.E.) of r is very useful in interpreting the value of r and is worked out for Karl Pearson's coefficient of correlation as,

$$\text{P.E.} = 0.6745\frac{1 - r^2}{\sqrt{n}}$$

If r is less than its P.E., it is not at all significant. If r is more than P.E., there is correlation. If r is more than six times its P.E. and greater than ± 0.5, then it is considered significant. Let us consider Example 2.

NOTES

Example 4.31: From the following data calculate '*r*' between *X* and *Y* applying the following three methods:

(a) The method of least squares.

(b) The method based on regression coefficients.

(c) The product moment method of Karl Pearson.

Verify the obtained result of any one method with that of another.

X	1	2	3	4	5	6	7	8	9
Y	9	8	10	12	11	13	14	16	15

Solution:

Let us develop the following table for calculating the value of '*r*':

X	*Y*	X^2	Y^2	*XY*
1	9	1	81	9
2	8	4	64	16
3	10	9	100	30
4	12	16	144	48
5	11	25	121	55
6	13	36	169	78
7	14	49	196	98
8	16	64	256	128
9	15	81	225	135

$n = 9$

$\sum X = 45$ $\sum Y = 108$ $\sum X^2 = 285$ $\sum Y^2 = 1356$ $\sum XY = 597$

$\therefore \quad \bar{X} = 5; \qquad \bar{Y} = 12$

(a) Coefficient of correlation by the method of least squares is worked out as follows:

First find out the estimating equation,

$$\hat{Y} = a + bX_i$$

Where, $$b = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum X^2 - n\bar{X}^2}$$

$$= \frac{597 - 9(5)(12)}{285 - 9(25)} = \frac{597 - 540}{285 - 225} = \frac{57}{60} = 0.95$$

and $$a = \bar{Y} - b\bar{X}$$

$$= 12 - 0.95(5) = 12 - 4.75 = 7.25$$

Hence, $\hat{Y} = 7.25 + 0.95X_i$

Now '*r*' can be worked out as follows by the method of least squares,

NOTES

$$r = \sqrt{1 - \frac{\text{Unexplained variation}}{\text{Total variation}}}$$

$$= \sqrt{1 - \frac{\Sigma\left(Y - \hat{Y}\right)^2}{\Sigma\left(Y - \overline{Y}\right)^2}} = \sqrt{\frac{\Sigma\left(\hat{Y} - \overline{Y}\right)^2}{\Sigma\left(Y - \overline{Y}\right)^2}}$$

$$= \sqrt{\frac{a\Sigma Y + b\Sigma XY - n\overline{Y}^2}{\Sigma Y^2 - n\overline{Y}^2}}$$

This is as per short-cut formula,

$$r = \sqrt{\frac{7.25(108) + 0.95(597) - 9(12)^2}{1356 - 9(12)^2}}$$

$$= \sqrt{\frac{783 + 567.15 - 1296}{1356 - 1296}}$$

$$= \sqrt{\frac{54.15}{60}} = \sqrt{0.9025} = 0.95$$

(b) Coefficient of correlation by the method based on regression coefficients is worked out as,

$\because$ Regression coefficients of *Y* on *X,*

i.e., $b_{YX} = \frac{\Sigma XY - n\overline{X}\overline{Y}}{\Sigma X^2 - n\overline{X}^2}$

$$= \frac{597 - 9 \times 5 \times 12}{285 - 9(5)^2} = \frac{597 - 540}{285 - 225} = \frac{57}{60}$$

Regression coefficient of *X* on *Y,*

i.e., $b_{XY} = \frac{\Sigma XY - n\overline{X}\overline{Y}}{\Sigma Y^2 - n\overline{Y}^2}$

$$= \frac{597 - 9 \times 5 \times 12}{1356 - 9(12)^2} = \frac{597 - 540}{1356 - 1296} = \frac{57}{60}$$

Hence, $r = \sqrt{b_{YX} \cdot b_{XY}}$

$$= \sqrt{\frac{57}{60} \times \frac{57}{60}} = \frac{57}{60} = 0.95$$

NOTES

(c) Coefficient of correlation by the product moment method of Karl Pearson is worked out as,

$$r = \frac{\sum XY - n\overline{X}\,\overline{Y}}{\sqrt{\sum X^2 - n\overline{X}^2}\sqrt{\sum Y^2 - n\overline{Y}^2}}$$

$$= \frac{597 - 9(5)(12)}{\sqrt{285 - 9(5)^2}\sqrt{1356 - 9(12)^2}}$$

$$= \frac{597 - 540}{\sqrt{285 - 225}\sqrt{1356 - 1296}} = \frac{57}{\sqrt{60}\sqrt{60}} = \frac{57}{60} = 0.95$$

Hence, we get the value of $r = 0.95$. We get the same value by applying the other two methods also. Therefore, whichever method we apply, the results will be the same.

Product Moment Correlation

In statistics, the Pearson Product Moment Correlation Coefficient is a measure of the correlation (linear dependence) between two variables *X* and *Y*, giving a value between +1 and −1 inclusive. It is sometimes referred the PPMCC or PCC or Pearson's *r.* It is widely used in the sciences as a measure of the strength of linear dependence between two variables. It was developed by Karl Pearson from a related idea introduced by Francis Galton in the 1880s.

Pearson's correlation coefficient between two variables is defined as the covariance of the two variables divided by the product of their standard deviations. The form of the definition involves a 'product moment', i.e., the mean (the first moment about the origin) of the product of the mean adjusted random variables; hence the modifier *product moment* in the name.

For a Population

Pearson's correlation coefficient when applied to a population is commonly represented by the Greek letter ρ (rho) and may be referred to as the *population correlation coefficient* or the *population Pearson correlation coefficient*. The formula for ρ is:

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E\left[(X - \mu_X)(Y - \mu_Y)\right]}{\sigma_X \sigma_Y}$$

For a Sample

Pearson's correlation coefficient when applied to a sample is commonly represented by the letter *r* and may be referred to as the *sample correlation coefficient* or the *sample Pearson correlation coefficient*. We can obtain a formula for *r* by substituting estimates of the covariances and variances based on a sample into the above formula. That formula for *r* is:

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}}$$

An equivalent expression gives the correlation coefficient as the mean of the products of the standard scores. Based on a sample of paired data (X_i, Y_i), the sample Pearson correlation coefficient is as follows:

NOTES

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{s_X}\right)\left(\frac{Y_i - \bar{Y}}{s_Y}\right)$$

Where,

$\frac{X_i - \bar{X}}{s_X}, \bar{X}$, and s_X are the standard score, sample mean, and sample standard deviation, respectively.

How to Calculate Product Moment Correlation Coefficient?

The product moment correlation coefficient allows you to work out the linear dependence of two variables (referred to as *X* and *Y*). Let us consider an example, suppose you are the owner of a restaurant. You record the time of every 10th customer stayed in your restaurant (*X* in minutes) and the amount spend (*Y*, in rupees). If it is considered that the longer time the customer stayed the bigger is the amount spend, then this would be a positive correlation. Or it can also be considered in the other way, i.e., the richer the client the lesser time he takes for lunch in restaurant, then this would be a negative correlation. Pearson Product-Moment Correlation Coefficient or PMCC can be calculated to find the correlation in a situation.

Step 1: Remove Incomplete Pairs: After removing incomplete pairs, use only those observations where *both X* and *Y* are known. However, do not exclude observations just because one of the values equals zero.

Step 2: Summarize the Data into the Values needed for the Calculation: These are:

- *n*: The number of data.
- ΣX: The sum of all the *X* values.
- ΣX^2: The sum of the squares of the *X* values.
- ΣY: The sum of all the *Y* values.
- ΣY^2: The sum of the squares of the *Y* values.
- ΣXY: The sum of each *X* value multiplied by its corresponding *Y* value.

Step 3: Calculate S_{XY}, S_{XX} and S_{YY} using these values:

- $S_{XY} = \Sigma XY - (\Sigma XY \div n)$
- $S_{XX} = \Sigma X^2 - (\Sigma X\ \Sigma X \div n)$
- $S_{XY} = \Sigma Y^2 - (\Sigma Y\ \Sigma Y \div n)$

Step 4: Insert these Values into the Equation below to Calculate the Product Moment Correlation Coefficient (*r*): The value should be between 1 and –1.

$$r = \frac{S_{xy}}{\sqrt{S_{xy} S_{yy}}}$$

NOTES

- If a value is close to 1 implies strong positive correlation. The higher the *X*, the higher the *Y*.
- If a value close to 0 implies little or no correlation.
- If a value close to –1 implies strong negative correlation. The higher the *X*, the lower the *Y*.

Rank Correlation

If observations on two variables are given in the form of ranks and not as numerical values, it is possible to compute what is known as rank correlation between the two series.

The **rank correlation**, written as ρ, is a descriptive index of agreement between ranks over individuals. It is the same as the ordinary coefficient of correlation computed on ranks, but its formula is simpler.

$$\rho = 1 - \frac{6\Sigma D_i^2}{n(n^2 - 1)}$$

Here, n is the number of observations and D_i, the positive difference between ranks associated with the individuals i.

Like r, the rank correlation lies between –1 and +1. Consider Examples 3 and 4 for better understanding

Example 4.32: The ranks given by two judges to 10 individuals are as follows:

	Rank given by			
Individual	*Judge I*	*Judge II*	*D*	D^2
	x	y	$= x - y$	
1	1	7	6	36
2	2	5	3	9
3	7	8	1	1
4	9	10	1	1
5	8	9	1	1
6	6	4	2	4
7	4	1	3	9
8	3	6	3	9
9	10	3	7	49
10	5	2	3	9
				$\Sigma D^2 = 128$

Solution:

The rank correlation is given by,

$$\rho = 1 - \frac{6\Sigma D^2}{n^3 - n} = 1 - \frac{6 \times 128}{10^3 - 10} = 1 - 0.776 = 0.224$$

The value of $\rho = 0.224$ shows that the agreement between the judges is not high.

Example 4.33: Consider Example 3 and compute *r* and compare.

Solution:

The simple coefficient of correlation *r* for the previous data is calculated as follows:

x	y	x^2	y^2	xy
1	7	1	49	7
2	5	4	25	10
7	8	49	64	56
9	10	81	100	90
8	9	64	81	72
6	4	36	16	24
4	1	16	1	4
3	6	9	36	18
10	3	100	9	30
5	2	25	4	10
$\Sigma x = 55$	$\Sigma y = 55$	$\Sigma x^2 = 385$	$\Sigma y^2 = 385$	$\Sigma xy = 321$

$$r = \frac{321 - 10 \times \frac{55}{10} \times \frac{55}{10}}{\sqrt{385 - 10 \times \left(\frac{55}{10}\right)^2}\sqrt{385 - 10 \times \left(\frac{55}{10}\right)^2}} = \frac{18.5}{\sqrt{82.5 \times 82.5}} = \frac{18.5}{82.5} = 0.224$$

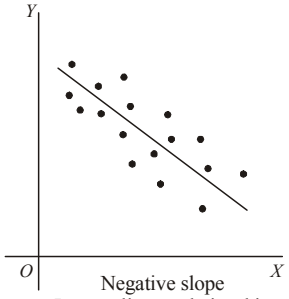
This shows that the Spearman ρ for any two sets of ranks is the same as the Pearson *r* for the set of ranks. However, it is much easier to compute ρ.

Often, the ranks are not given. Instead, the numerical values of observations are given. In such a case, we must attach the ranks to these values to calculate ρ.

Example 4.34: Show by means of diagrams various cases of scatter expressing correlation between *x, y.*

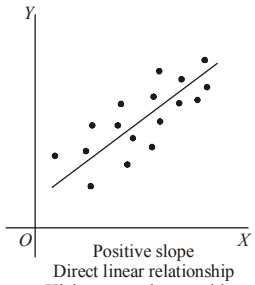
Solution:

(*a*)



Negative slope
Inverse linear relationship
High scatter *r* low, negative

(*b*)

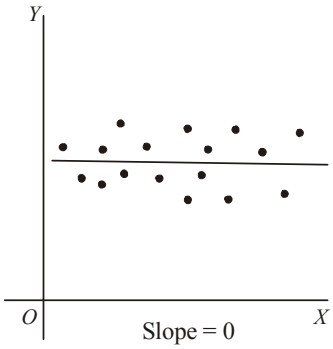


Positive slope
Direct linear relationship
High scatter *r* low, positive

NOTES

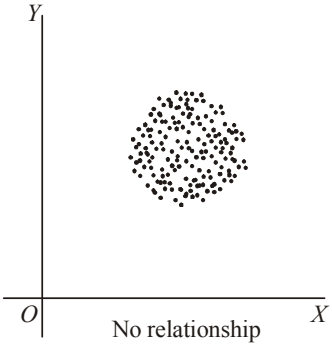
NOTES

(*c*)



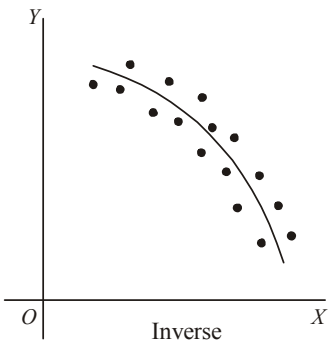
Slope = 0
$r = 0$

(*d*)



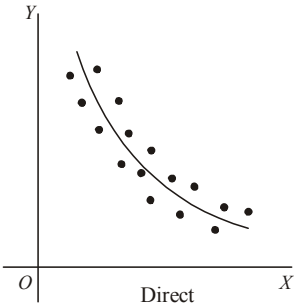
No relationship
$r \simeq 0$

(*e*)



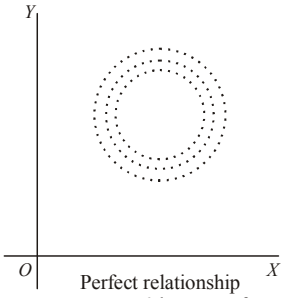
Inverse
curvilinear relationship

(*f*)



Direct
curvilinear relationship

(*g*)



Perfect relationship
But, $r = 0$ because of
non-linear relation

Correlation analysis helps us in determining the degree to which two or more variables are related to each other.

When there are only two variables, we can determine the degree to which one variable is linearly related to the other. Regression analysis helps in determining the pattern of relationship between one or more independent variables and a dependent variable. This is done by an equation estimated with the help of data.

CHECK YOUR PROGRESS

11. What is theory of correlation?
12. What does correlation analysis help us in determining?

4.6 PERCENTILE AND PERCENTILE RANK

NOTES

Percentile is the point below which the given percent of cases lies. We know that median is the point in the distribution below which 50% of the case lies. We also know how to calculate median and quartile deviation. It was discussed earlier that Q_1 and Q_3 mark points in the distribution below which lie 25% and 75% of the cases respectively. By using the same method we can compute points below which any percent of cases, say 10%, 45%, 70%, 80%, 85% or any percent of the scores or cases lie. These points are called as percentiles and are represented by the symbol 'Pp', the P referring to the percentage of cases below the given value or score. Thus P_{15} refers to the point below which 15% of the scores lie, P_{85} refers the point or score below which 85% of the scores lies. Suppose a student is declared as achieving 84^{th} percentile (P_{84}) in a test, it means that his position in the test is very good as 84% of the examinees have achieved marks which are less than this student. In other words 84% of the students of his class are inferior to him, or he is better than 84% of the students of his class.

The method of finding out the percentile is essentially the same as we have seen and used in computing median and quartile deviation. Following formula may be used to compute percentile:

$$Pp = l + [(pN - F)/fp] \times i$$

Where,

P = percentage of the distribution to be calculate, such as 15%, 30%, 50% (median), 70% etc.

l = Exact lower limit of the class interval upon which Pp lies.

pN = Part of N to be counted off in order to reach Pp.

f = Sum of all scores upon intervals below l.

fp = Number of scores within the interval upon which Pp falls.

i = Size or length of the class interval.

Percentile is the very useful concept to represent the exact position of an individual in the class. Merely percentage of marks such as 80% in mathematics does not give much information about the learner progress or achievement. It simply gives information that this student has got 80 % marks. May be this 80% marks is the highest in the class or may be that this marks is the lowest in the class or may be that this is the average marks obtained by students in the class. In the other way if we present marks of student in percentile then it represent exact performance of students with reference to the class. For example if we say that a student A has got P_{80} (80^{th} Percentile) in the class then it gives information that 80% of his or her classmates have scores below him or her i.e. the learner is very-very good in comparison to others in the class. He or she is one of the 20% good performers of the class.

NOTES

Use of Quartile Deviation, Standard Deviation and Percentile in Classroom Situation

Quartile deviation (QD) should be used when :

- The median is the measure of central tendency.
- There are scattered scores in the distribution or there are extreme scores in the distribution which can influence the standard deviation of the distribution disproportionately.
- The interest of the researcher is the concentration of scores around the median, in other words the middle 50% of the cases.

Standard deviation (SD) should be used when:

- The statistics having the greatest stability is required by the researcher.
- The extreme deviations exercise a proportionally greater effect upon the variability.
- Coefficient of correlation and other statistics are subsequently to be calculated.

Percentile should be used when:

- You wish to evaluate and compare the achievement of a given student in tests of different subjects.
- You are interested in locating the exact position of a student in the class in terms of his or her performance.
- You want to find out how much percent of students are below a particular score in the class.

4.6.1 Skewness and Kurtosis

Skewness

Skewness refers to lack of symmetry in a distribution. In a symmetrical distribution, the mean, median and mode coincide. Positive and negative skewness is shown in Figure 4.3.

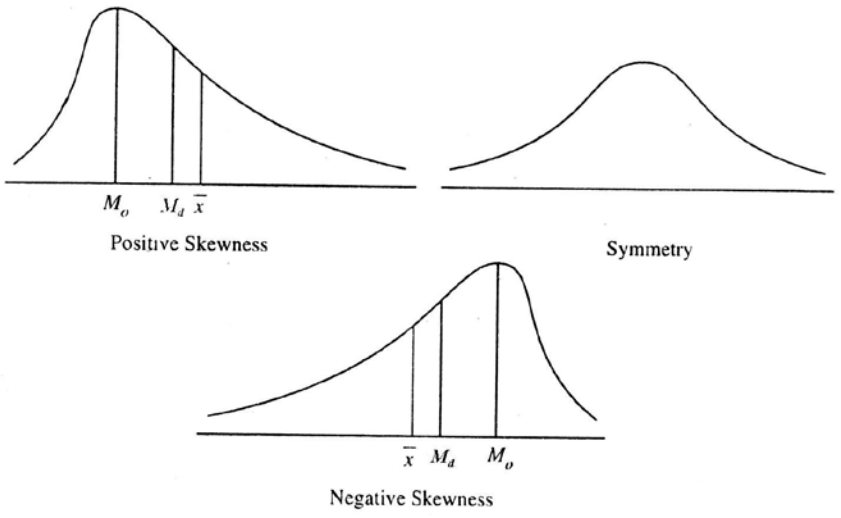


Fig. 4.3 *Skewness*

NOTES

In a positively skewed distribution, the longer tail is on the right side and the mean is on the right of the median.

In a negatively skewed distribution, the longer tail is on the left and the mean is on the left of the median.

In a skewed distribution, the distance between the mean and the median is nearly one-third of that between the mean and the mode.

How to check the presence of skewness in a distribution

In the following cases skewness is present in the data:

(a) The graph is not symmetrical.

(b) The mean, median and mode do not coincide.

(c) The quartiles are not equidistant from the mean.

(d) The sum of positive and negative deviations from the median is not zero.

(e) Frequencies are not similarly distributed on either side of the mode.

Measure of skewness

A measure of skewness gives a numerical expression and the direction of asymmetry in a distribution. It gives information about the shape of the distribution and the degree of variation on either side of the central value.

We consider some relative measures of skewness that are as follows:

(a) *Pearson's coefficient of skewness*

$$PSk = \frac{\bar{x} - M_O}{s} = \frac{3(\bar{x} - M_d)}{s}$$

It may have any value, but usually it lies between –1 and +1.

Example 4.35: If for a given data it is found that

$\bar{x} = 10,$ Mode = 8, s = 4, we have

$$PSk = \frac{\bar{x} - M_O}{s} = \frac{10 - 8}{4} = 0.5$$

(b) *Bowley's coefficient of skewness*

$$BSk = \frac{Q_3 - Q_1 - 2M_d}{Q_3 - Q_1}$$

Its value lies between –1 and +1.

Example 4.36: If for a given data $Q_1 = 2,\ Q_3 = 8,\ M_d = 5$

$$BSk = \frac{Q_3 + Q_1 - 2M_d}{Q_3 - Q_1} = \frac{8 + 2 - 5}{8 - 2} = 0.83$$

(c) *Kelley's coefficient of skewness*

$$KSk = P_{50} - \frac{1}{2}(P_{10} + P_{90})$$

where P_{10}, P_{50} and P_{90} are the 10th, 50th and 90th percentiles of the data.

(*d*) *Method of moments*

NOTES

If μ_2, μ_3 are moments about the mean we have the coefficient of skewness

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \mu_3^2 / \sigma^6$$

Sometimes, we define the coefficient of skewness as follows:

$$\gamma_1 = \sqrt{\beta_1} = \sqrt{\frac{\mu_3^2}{\mu_2^3}} = \frac{\mu_3}{\sigma^3}$$

Kurtosis

Kurtosis is a measure of peakedness of a distribution. It shows the degree of convexity of a frequency curve.

If the normal curve is taken as the standard, symmetrical and bell-shaped curve then kurtosis gives a measure of departure from the normal convexity of a distribution. The normal curve is mesokurtic. It is of intermediate peakedness. The flat-topped curve, broader than the normal, is termed platykurtic. The slender, highly peaked curve is termed leptokurtic.

Measures of kurtosis

(a) Moment Coefficient of Kurtosis : $\beta_2 = \frac{\mu_4}{\mu_2^2}$

Instead of β_2 we often use $\gamma_2 = \beta_2 - 3$ which is positive for a leptokurtic distribution, negative for a platykurtic distribution and zero for the normal distribution.

(b) Percentile Coefficient of Kurtosis $k = \frac{Q}{P_{90} - P_{10}}$, where $Q = \frac{1}{2}(Q_3 - Q_1)$ is the semi-interquartile range.

CHECK YOUR PROGRESS

13. What is skewness?
14. What occurs in a skewed distribution?

4.7 NORMAL PROBABILITY CURVE

In probability theory, the normal probability curve or normal (or Gaussian) distribution is considered as the most frequently occurring continuous probability distribution. Normal distributions are exceptionally significant in statistics and are typically used

NOTES

in the context of natural and social sciences generally for real-valued random variables whose distributions are not known. The curve of normal distribution is illustrated using the normal probability curve and is the most common type of distribution. Typically, a normal distribution is a very important statistical data distribution pattern occurring in many natural phenomena, such as height, blood pressure, lengths of objects produced by machines, etc. Certain data, when graphed as a histogram (data on the horizontal axis, amount of data on the vertical axis), creates a bell-shaped curve known as a normal probability curve or normal distribution.

The graph of the normal distribution depends on two factors - the mean (average) and standard deviation (σ, sigma). The mean of the distribution determines the location of the center of the graph and the standard deviation determines the height and width of the graph. If the standard deviation is large, the curve is short and wide while if the standard deviation is small, the curve is tall and narrow. All normal distributions look like a symmetric, bell-shaped curve. Thus, a normal probability curve shows the theoretical shape of a normally distributed histogram. The shape of the normal probability curve is also based on two parameters: mean (average) and standard deviation (σ, sigma). Thus, the normal distribution is a very important class of statistical distribution. All normal distributions are symmetric and have bell-shaped density curves with a single peak. Specifically, in any normal distribution, two quantities have to be specified, the mean μ where the peak of the density occurs and the standard deviation σ which indicates the spread or girth of the bell curve.

Among all the probability distributions, the normal probability distribution is by far the most important and frequently used continuous probability distribution. This is so because this distribution fits well in many types of problems. This distribution is of special significance in inferential statistics since it describes probabilistically the link between a statistic and a parameter (i.e., between the sample results and the population from which the sample is drawn). The name Karl Gauss, 18th century mathematician–astronomer, is associated with this distribution and in honour of his contribution, this distribution is often known as the Gaussian distribution.

Normal distribution can be theoretically derived as the limiting form of many discrete distributions. For instance, if in the binomial expansion of $(p + q)^n$, the value of 'n' is infinity and $p = q = \frac{1}{2}$, then a perfectly smooth symmetrical curve would be obtained. Even if the values of p and q are not equal but if the value of the exponent 'n' happens to be very very large, we get a curve of normal probability smooth and symmetrical. Such curves are called normal probability curves (or at times known as normal curves of error) and such curves represent the normal distributions.[1]

The probability function in case of normal probability distribution[2] is given as:

$$f(x) = \frac{1}{\sigma.\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Where, μ = The mean of the distribution.

σ^2 = Variance of the distribution.

The normal distribution is thus defined by two parameters viz., μ and σ^2. This distribution can be represented graphically as in Figure 7.1.

NOTES

Characteristics of Normal Distribution

The characteristics of the normal distribution or that of a normal curve are, as given as follow:

1. It is a symmetric distribution.
2. The mean μ defines where the peak of the curve occurs. In other words, the ordinate at the mean is the highest ordinate. The height of the ordinate at a distance of one standard deviation from the mean is 60.653% of the height of the mean ordinate and similarly the height of other ordinates at various standard deviations (σ_s) from mean happens to be a fixed relationship with the height of the mean ordinate (see Figure 4.4).
3. The curve is asymptotic to the base line which means that it continues to approach but never touches the horizontal axis.
4. The variance (σ^2) defines the spread of the curve.
5. Area enclosed between mean ordinate and an ordinate at a distance of one standard deviation from the mean is always 34.134% of the total area of the curve. It means that the area enclosed between two ordinates at one sigma (SD) distance from the mean on either side would always be 68.268% of the total area. This can be shown as follows:

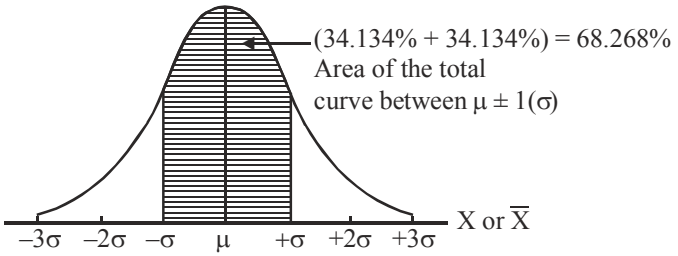


Fig. 4.4 *Area of the Total Curve between* μ ± 1 (σ)

Similarly, the other area relationships are as follows:

Between		*Area Covered to Total Area of the Normal Curve*[4]
μ ± 1	S.D.	68.27%
μ ± 2	S.D.	95.45%
μ ± 3	S.D.	99.73%
μ ± 1.96	S.D.	95%
μ ± 2.578	S.D.	99%
μ ± 0.6745	S.D.	50%

6. The normal distribution has only one mode since the curve has a single peak. In other words, it is always a unimodal distribution.
7. The maximum ordinate divides the graph of normal curve into two equal parts.
8. In addition to all the above stated characteristics the curve has the following properties:
   (a) $\mu = \bar{x}$
   (b) $\mu_2 = \sigma^2 =$ Variance

(c) $\mu_4=3\sigma^4$

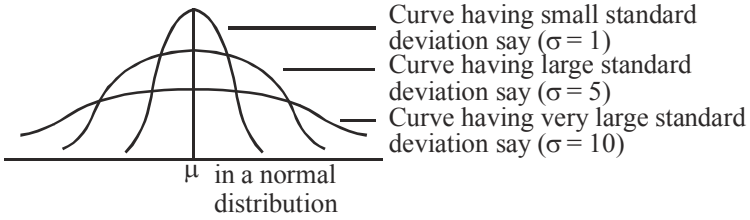
(d) Moment Coefficient of Kurtosis = 3

Family of Normal Distributions

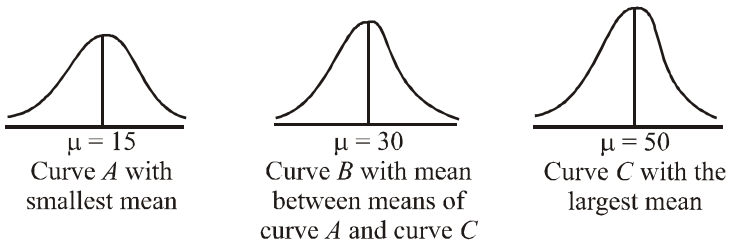
NOTES

We can have several normal probability distributions but each particular normal distribution is being defined by its two parameters viz., the mean (μ) and the standard deviation (σ). There is, thus, not a single normal curve but rather a family of normal curves (see Figure 4.5). We can exhibit some of these as under:

Normal curves with identical means but different standard deviations:



Normal curves with identical standard deviation but each with different means:



Normal curves each with different standard deviations and different means:

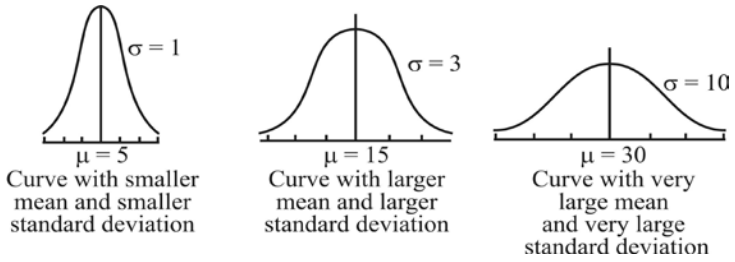


Fig. 4.5 *Family of Curves*

How to Measure the Area under the Normal Curve

We have stated above some of the area relationships involving certain intervals of standard deviations (plus and minus) from the means that are true in case of a normal curve. But what should be done in all other cases? We can make use of the statistical tables constructed by mathematicians for the purpose. Using these tables we can find the area (or probability, taking the entire area of the curve as equal to 1) that the normally distributed random variable will lie within certain distances from the mean. These distances are defined in terms of standard deviations. While using the tables showing the area under the normal curve we talk in terms of standard

NOTES

variate (symbolically Z) which really means standard deviations without units of measurement and this 'Z' is worked out as under:

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

Where, Z = The standard variate (or number of standard deviations from X to the mean of the distribution).

X = Value of the random variable under consideration.

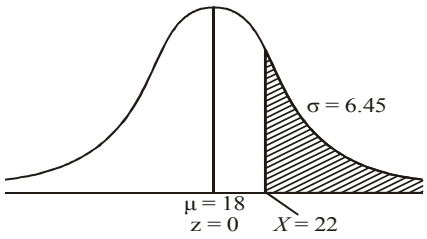
μ = Mean of the distribution of the random variable.

σ = Standard deviation of the distribution.

The table showing the area under the normal curve (often termed as the standard normal probability distribution table) is organized in terms of standard variate (or Z) values. It gives the values for only half the area under the normal curve, beginning with $Z = 0$ at the mean. Since the normal distribution is perfectly symmetrical the values true for one half of the curve are also true for the other half. We now illustrate the use of such a table for working out certain problems.

Example 4.37: A banker claims that the life of a regular savings account opened with his bank averages 18 months with a standard deviation of 6.45 months. Answer the following: (a) What is the probability that there will still be money in 22 months in a savings account opened with the said bank by a depositor? (b) What is the probability that the account will have been closed before two years?

Solution: (a) For finding the required probability we are interested in the area of the portion of the normal curve as shaded and shown in figure given below:



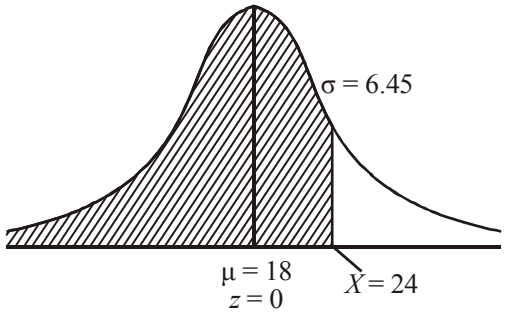
Let us calculate Z as under:

$$Z = \frac{X - \mu}{\sigma} = \frac{22 - 18}{6.45} = 0.62$$

The value from the table showing the area under the normal curve for $Z = 0.62$ is 0.2324. This means that the area of the curve between $\mu = 18$ and $X = 22$ is 0.2324. Hence, the area of the shaded portion of the curve is (0.5) – (0.2324) = 0.2676 since the area of the entire right hand portion of the curve always happens to be 0.5. Thus, the probability that there will still be money in 22 months in a savings account is 0.2676.

NOTES

(b) For finding the required probability we are interested in the area of the portion of the normal curve as shaded and shown in figure given below:



For the purpose we calculate,

$$Z = \frac{24 - 18}{6.45} = 0.93$$

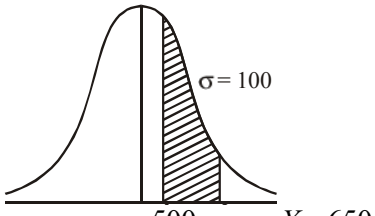
The value from the concerning table, when $Z = 0.93$, is 0.3238 which refers to the area of the curve between $\mu = 18$ and $X = 24$. The area of the entire left hand portion of the curve is 0.5 as usual.

Hence, the area of the shaded portion is (0.5) + (0.3238) = 0.8238 which is the required probability that the account will have been closed before two years, i.e., before 24 months.

Example 4.38: Regarding a certain normal distribution concerning the income of the individuals we are given that mean=500 rupees and standard deviation =100 rupees. Find the probability that an individual selected at random will belong to income group,

(a) ₹ 550 to ₹ 650 (b) ₹ 420 to ₹570

Solution: (a) To find the required probability we are interested in the area of the portion of the normal curve as shaded and shown below:



To find the area of the curve between $X = 550$ to 650, let us do the following calculations:

$$Z = \frac{550 - 500}{100} = \frac{50}{100} = 0.50$$

Corresponding to which the area between $\mu = 500$ and $X = 550$ in the curve as per table is equal to 0.1915 and,

$$Z = \frac{650 - 500}{100} = \frac{150}{100} = 1.5$$

NOTES

Corresponding to which, the area between $\mu = 500$ and $X = 650$ in the curve, as per table, is equal to 0.4332.

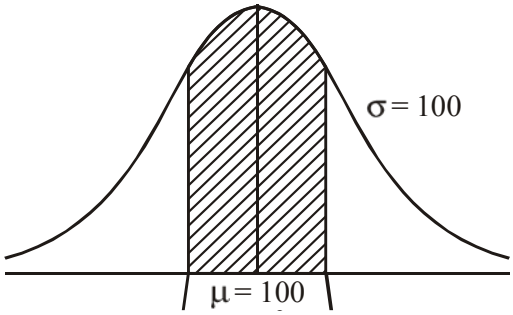
Hence, the area of the curve that lies between $X = 550$ and $X = 650$ is,

$$(0.4332) - (0.1915) = 0.2417$$

This is the required probability that an individual selected at random will belong to income group of ₹ 550 to ₹ 650.

(b) To find the required probability, we are interested in the area of the portion of the normal curve as shaded and shown below:

To find the area of the shaded portion we make the following calculations:



$$Z = \frac{570 - 500}{100} = 0.70$$

Corresponding to which the area between $\mu = 500$ and $X = 570$ in the curve as per table is equal to 0.2580.

and $$Z = \frac{420 - 500}{100} = -0.80$$

Corresponding to which the area between $\mu = 500$ and $X = 420$ in the curve as per table is equal to 0.2881.

Hence, the required area in the curve between $X = 420$ and $X = 570$ is,

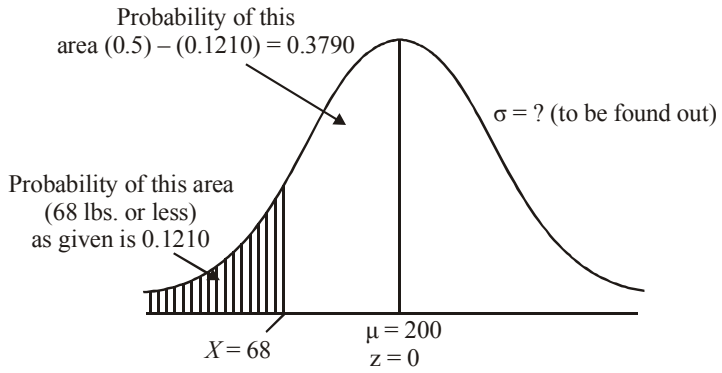
$$(0.2580) + (0.2881) = 0.5461$$

This is the required probability that an individual selected at random will belong to income group of ₹ 420 to ₹ 570.

Example 4.39: A certain company manufactures $1\frac{1''}{2}$ all-purpose rope using imported hemp. The manager of the company knows that the average load-bearing capacity of the rope is 200 lbs. Assuming that normal distribution applies, find the standard deviation of load-bearing capacity for the $1\frac{1''}{2}$ rope if it is given that the rope has a 0.1210 probability of breaking with 68 lbs. or less pull.

NOTES

Solution: Given information can be depicted in a normal curve as shown below:



If the probability of the area falling within $\mu = 200$ and $X = 68$ is 0.3790 as stated above, the corresponding value of Z as per the standard statistical tables showing the area of the normal curve is – 1.17 (minus sign indicates that we are in the left portion of the curve).

Now to find σ, we can write,

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

or $$-1.17 = \frac{68 - 200}{\sigma}$$

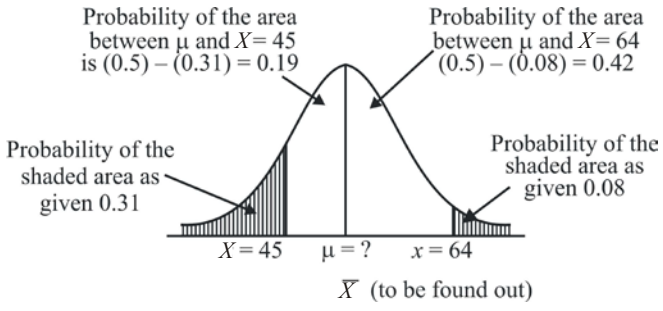
or $$-1.17\sigma = -132$$

or $$\sigma = 112.8 \text{ lbs. approx.}$$

Thus, the required standard deviation is 112.8 lbs. approximately.

Example 4.40: In a normal distribution, 31 per cent items are below 45 and 8 per cent are above 64. Find the $\bar{X}$ and σ of this distribution.

Solution: We can depict the given information in a normal curve as shown below:



If the probability of the area falling within μ and $X = 45$ is 0.19 as stated above, the corresponding value of Z from the table showing the area of the normal curve is – 0.50. Since, we are in the left portion of the curve, we can express this as under,

$$-0.50 = \frac{45 - \mu}{\sigma} \quad (1)$$

Similarly, if the probability of the area falling within μ and $X = 64$ is 0.42, as stated above, the corresponding value of Z from the area table is, +1.41. Since, we are in the right portion of the curve we can express this as under,

$$1.41 = \frac{64 - \mu}{\sigma} \quad (2)$$

If we solve Equations (1) and (2) above to obtain the value of μ or $\bar{X}$, we have,

$$-0.5\,\sigma = 45 - \mu \quad (3)$$

$$1.41\,\sigma = 64 - \mu \quad (4)$$

NOTES

By subtracting Equation (4) from Equation (3) we have,

$$-1.91\,\sigma = -19$$

$$\therefore \quad \sigma = 10$$

Putting $\sigma = 10$ in Equation (3) we have,

$$-5 = 45 - \mu$$

$$\therefore \quad \mu = 50$$

Hence, $\bar{X}$ (or μ)=50 and σ=10 for the concerning normal distribution or probability curve.

Applications of normal distribution or probability curve

The following are the applications of normal distribution or probability curve:

1. **Random Processes:** Many naturally occurring random processes tend to have a distribution that is approximately normal. Examples can be found in any field, these include: SAT test scores of college bound students and body temperature of a healthy adult.
2. **Approximation of Binomial Distribution:** When $np>5$ and $n(1-p)>5$, the normal distribution provides an good approximation of the binomial distribution. Distributions that are based on multiple observations, for example the Binomial distribution, approach the normal distribution when *n* gets large. The value $n>30$ is usually considered large.
3. **Standardization:** It is used where it is usually hypothesize that the theoretical distribution of a certain variable is normal, whereas the measurement of such variable may not give a normal distribution.

   For example, in the introductory classes of Statistics there are 200 students and it has been assumed that the performance of all the students in the examination should be normally distributed. In addition, for giving reasonable distribution of marks, the mean should be 55 and the standard deviation should be 10. After the examinations being over, the lecturer marked all the papers, and the mean and standard deviation of the raw scores given by the lecturer are 50 and 6, respectively. For converting the raw score to standardize score, the follows steps were taken:

   (a) The standard score is obtained by $Z = (X-50)/6$.

   (b) Then the converted (standardized) = $10(Z) + 55$.

   Hence, a raw score of 56 will be converted into 65.
4. **Composite Scores:** When more than one measure is used to measure a variable, the distribution of each measure usually differs from each other. In order to obtain an unbiased measure using several different measurements, each sub-measure is standardized before added together.

NOTES

For example, if the marks are awarded according to the average of the marks given by the Marker I and Marker II, then clearly the final grades are greatly affected by the Marker I than by Marker II as Marker I is awarded the marks with higher standard deviation as shown in table below:

For computing the composite score, the standardized scores of Marker I and Marker II should be averaged. If the ideal average score (μ) and standard deviation (σ) is taken to be 60 and 10, respectively, then the Z scores is converted into the standard score for each marker. The following table shows the resulted average standardized score 60 for every student.

	Marker I			Marker II		
Student	Raw score	z=(x-μ)/σ	Standard score	Raw score	z=(x-μ)/σ	Standard score
A	80	1.4	74	50	-1.4	46
B	70	0.7	67	55	-0.7	53
C	60	0	60	60	0	60
D	50	-0.7	53	65	0.7	67
E	40	-1.4	46	70	1.4	74

5. **Probability Distribution:** The probability distribution of $\bar{X}$ for large *n* is the normal distribution. The Central Limit Theorem states that if the observations are independent for one population which has a mean (μ) and standard deviation (σ) then for large *n* (*n*>30) $\bar{X}$ has a normal distribution with the same mean and a standard deviation of $\sigma / \sqrt{n}$.

Normal Probability Curve and its properties and uses

The Normal Probability Curve (NPC), simply known as normal curve, is a symmetrical bell-shaped curve. This curve is based upon the law of probability and was discovered by French mathematician Abraham Demoivre (1667–1754) in the 18th century. In this curve, the mean, median and mode lie at the middle point of the distribution. The total area of the curve represents the total number of cases and the middle point represents the mean, median and mode. The base line is divided into six sigma units (σ units). The scores more than the mean come on the + σ side and the scores less than the mean come on the – σ side. The mean point (middle point) is marked as zero (0). All the scores are expected to lie between – 3σ to + 3σ.

Properties of Normal Probability Curve

The NPC or Normal Probability Curve has several features which are essential to understand for its use. The major characteristics are limited. They are as follows:

- It is a bell-shaped curve.
- The measures of central tendency are equal, i.e, mean, mode and median concentrate on one point.
- The height of the curve is 0.3989.
- It is an asymptotic curve. The ends of the curve approach but never touch the *X*-axis at the extremes because of the possibility of locating in the population,

NOTES

in cases where scores are still higher than our highest score or lower than our lowest score. Therefore theoretically, it extends from minus infinity to plus infinity as illustrated in Figure 4.6. Here, *M* is the mean or expectation (location of the peak) and σ is the standard deviation.

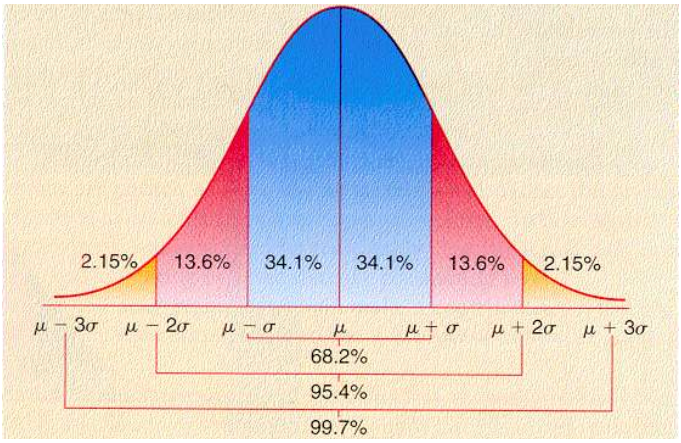


Fig. 4.6 *Normal Curve Showing Areas at Different Distances from the Mean*

- It has 50 per cent frequency above and 50 per cent below the mean. The mean is zero and it is always reference point.
- Standard deviation of a normal curve is always 1.
- The points of inflection of the curve occur at points –1 unit above and below mean.
- The distribution of frequency per cent has the definite limits.
- There is a definite relation between quartile deviation and standard deviation in a normal distribution curve.
- It is a mathematical curve and is an open-ended curve.

Some limits are as follows:

- The middle 68 per cent frequency is between –1 and +1.
- The middle 95 per cent frequency is between –1.96 and + 1.96.
- The middle 99 per cent frequency is between –2.58 and + 2.58.

The total area under the normal curve is arbitrarily taken as 10,000. Every score should be converted into standard score (Z score) by using the following formula:

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

The area in proportion should be converted into a percentage at the time of reading the table. From the table, we can see the areas from mean to σ and also we can read the value of σ scores from the mean for the corresponding fractional area.

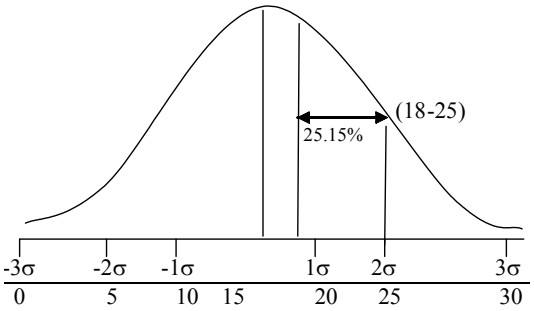
Uses of Normal Probability Curve: Determining Mean and Median

The uses of normal probability curve are discussed in this section.

NPC is used to Determine the Percentage of Cases within Given Limits

Example 4.41: Given a distribution of scores with a mean of 15 and a standard deviation of 5, what percentage of cases lie between 18 and 25. Refer to Figure given below to calculate the answer.

NOTES



Solution: Both the raw scores (18 and 25) are to be converted into *Z* scores.

$$\text{Z score of } 18 = \frac{X - M}{\sigma} = \frac{18 - 15}{5}$$

$$= \frac{3}{5}$$

$$= 0.6\sigma$$

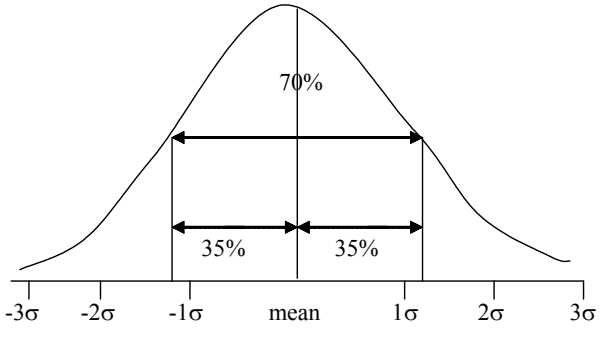
$$\text{Z score of } 25 = \frac{X - M}{\sigma} = \frac{25 - 15}{5}$$

$$= \frac{10}{5}$$

According to the table of area of a normal probability curve, the total percentage of cases lie between the mean and 0.6σ is 22.57. The percentage of cases lie between the mean and 2σ is 47.72. So, the total percentage of cases that fall between the scores 18 and 25 is 47.72 – 22.57 = 25.15.

NPC is used to determine the limit which includes a given percentage of cases

Example 4.42: Given a distribution of scores with a mean of 12 and an σ of 6, what limits will include the middle 70 per cent of the cases? Refer to Figure given below to calculate the answer.



Mean = 12, σ = 6

NOTES

Solution: The middle 70 per cent of the cases in a normal distribution signifies that 35 per cent cases above the mean and also 35 per cent cases below the mean. According to the table of area under NPC, 35 per cent of cases fall between the mean and 1.04 σ. So the middle 70 per cent of the cases will lie between –1.04σ to + 1.04σ.

The value of 1 σ = 6

So 1.04 σ = 6 × 1.04 = 6.24

The value of mean = 12

So the lowest limit for the middle 70 per cent cases of the distribution is:

12 – 6.24 = 5.76.

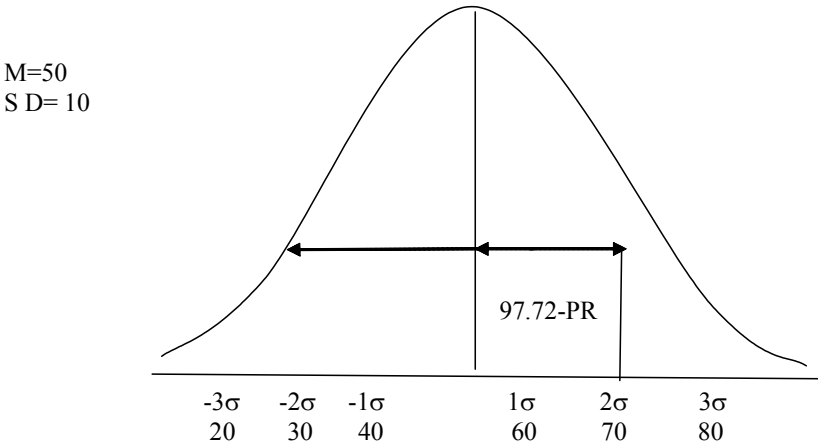
The highest limit for the middle 70 per cent cases of the distribution is:

12 + 6.24 = 18.24.

Thus, the middle 70 per cent cases lie in between 5.76 and 18.24.

NPC is used to Determine the Percentile Rank of a Student in his Class

Example 4.43: The score of a student in a class test is 70. The mean for the whole class is 50 and the σ is 10. Find the percentile rank of the student in the class. Refer the Figure given below to find the answer.



Solution: The Z score for the score 70 is $\frac{70-50}{10} = 2\sigma$

As per the table of area under the NPC, the area of the curve that lies between mean and 2σ is 47.72 per cent. The total percentage of cases below 70 is:

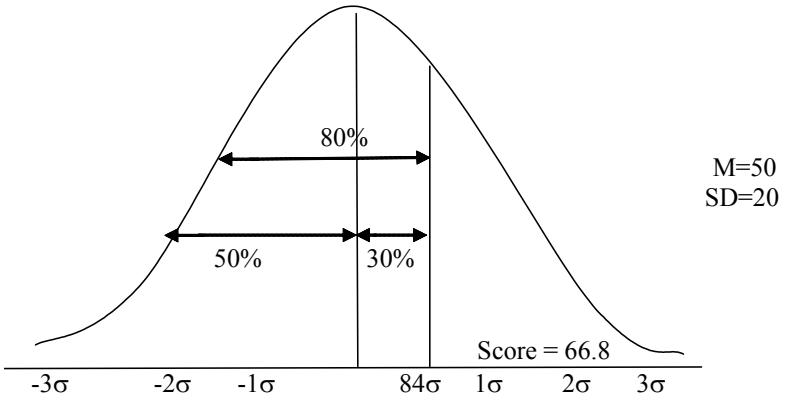
50 + 47.72 = 97.72 per cent or 98 per cent.

Thus, the percentile rank of the student is 98.

NPC is used to Find out the Percentile Value of a Student whose Percentile Rank is Known

Example 4.44: The percentile rank of a student in a class test is 80. The mean of the class in the test is 50 and the σ is 20. Calculate the student's score in the class test. Figure given below illustrates the case.

NOTES



Solution: The student has scored 30 per cent scores above the mean. According to the table of area under NPC, 30 per cent cases from the mean is 0.84σ.

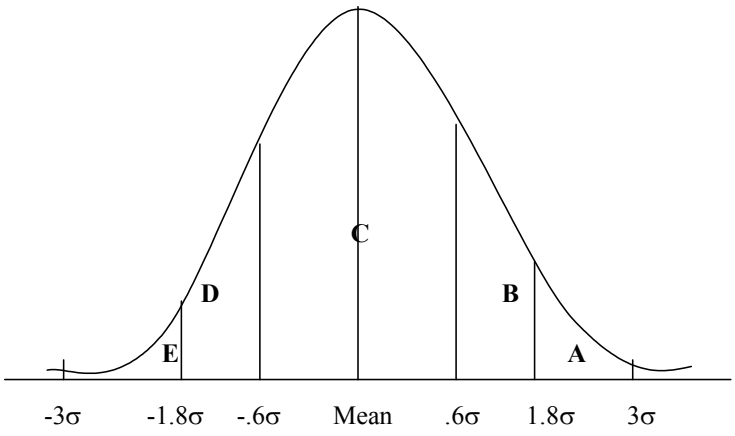
$1\sigma = 20$.

$0.84\sigma = 20 \times .84 = 16.8$

Thus, the percentile value of the student is $50 + 16.8 = 66.8$.

NPC is used to Divide a Group into Sub-Groups According to their Capacity

Example 4.45: Suppose there is a group of 100 students in a Commerce class. We want to divide them into five small groups A, B, C, D and E according to their ability, the range of ability being equal in each sub-group. Find out how many students should be placed in each category.



Solution: The total area under NPC is –3σ to + 3σ, that is 6σ. This 6σ should be divided into five parts, so $6\sigma \div 5 = 1.2\sigma$.

According to the table of area under NPC:

3.5 per cent of the cases lie between 1.8 σ to 3σ (Group A, the high scorers). 23.8 per cent of the cases lie between .6σ to 1.8σ (23.8 per cent of the cases for B and also 23.8 per cent of the cases for D), the middle 45 per cent of the cases lie – 0.6σ to + 0.6σ (Group C), and the lowest 3.5 per cent of the cases lie between –3σ to – 1.8σ (Group E)

NOTES

In category 'A' the number of students = 3.5 per cent = 3 or 4 students.

In category 'B' the number of students = 23.8 per cent = 24 students.

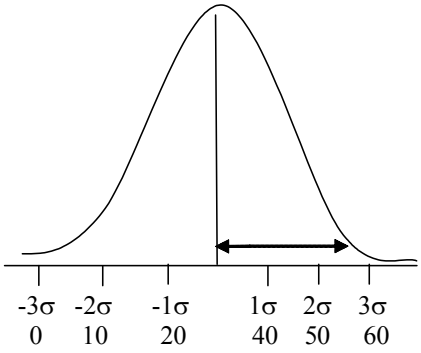
In category 'C' the number of students = 45 per cent = 45 students.

In category 'D' the number of students = 23.8 per cent = 24 students.

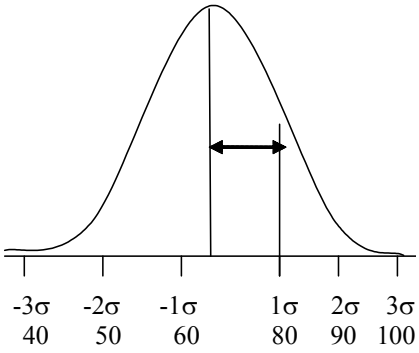
In category 'E' the number of students = 3.5 per cent = 3 or 4 students.

NPC is used to Compare the Scores of Students in Two Different Tests

Example 4.46: Suppose, a student scored 60 marks in English test and 80 marks in statistics test. The mean and SD for the English test is 30 and 10 respectively, whereas for the statistics test the mean is 70 and SD is 10. Find out, in which subject the student performed better. Refer to Figure given below.



M = 30
SD = 10
English Test



Mean = 70
SD = 10
Statistics Test

Solution: In case of the English test:

Raw score = 60

Mean = 30

SD =10

So Z score for the English test = $\frac{X - M}{\sigma} = \frac{60 - 30}{10} = \frac{30}{10} = 3\sigma$

In case of statistics test raw score = 80

Mean = 70

SD = 10

So Z Score for the statistics test = $\frac{X - M}{\sigma} = \frac{80 - 70}{10} = \frac{10}{10} = 1\sigma$

So, the student has done better in the English than the statistics on.

NPC is Used to Determine the Relative Difficulty Level of Test Items

Example 4.47: In a standardized test of psychology, question numbers A, B, C and D were solved by the students, 45 per cent, 38 per cent, 30 per cent and 15 per cent respectively. Assuming the normality, find out the relative difficulty level of the

questions. Also explain the difficulty levels of questions. Table 4.7 displays the information in tabular form.

Solution:

NOTES

Table 4.7 *Determining the Difficulty Level of Test Items*

Question Number	Percentage of Successful Students	Percentage of Unsuccessful Students	Percentage distance of Mean of Unsuccessful Students	Difficulty Level
A	45	55	55-50=5	0.13σ
B	38	62	62-50=12	0.31σ
C	30	70	70-50=20	0.52σ
D	15	85	85-50=35	1.04σ

As we know that in an NPC, 50 – 50 cases lie both the sides of mean. The mean of NPC is that point which is shown as 0. In an NPC, the explanation of difficulty level is done on the basis of σ — distance. Therefore, if a question is at the positive side of the NPC and σ has more distance from the mean, the question of a test will be much difficult. The relative difficulty value of the test items has been shown:

The question

A to B is 0.18σ is more difficult (0.31σ –0.13σ = 0.18σ)

A to C is 0.39σ is more difficult (0.52σ –0.13σ = 0.39σ)

A to D is 0.91σ is more difficult (1.04σ –0.13σ = 0.91σ)

B to C is 0.21σ is more difficult (0.52σ –0.31σ = 0.21σ)

B to D is 0.73σ is more difficult (1.04σ –0.31σ = 0.73σ)

C to D is 0.52σ is more difficult (1.04σ –0.152σ = 0.52σ)

Statistical Significance

Statistical significance is the result that is not likely to occur randomly, but rather it is likely to be attributable to a specific cause. Statistical significance can be strong or weak and is important feature of research in many mathematics and science related fields. Statistical significance does not always indicate practical significance. In addition, it can be misinterpreted when researchers do not use language carefully in reporting their results.

The calculation of statistical significance (significance testing) is subject to a certain degree of error. The researcher must define in advance the probability of a sampling error which exists in any test that does not include the entire population. Sample size is considered as an important component of statistical significance because larger samples are less prone to accidents. Only random, representative samples should be used in significance testing.

The level at which one can accept whether an event is statistically significant is known as the significance level or P value. Hence, statistical significance is the

NOTES

number called a P value and defines the probability of the result being observed given that the null hypothesis is true. If this P value is sufficiently small, the experimenter can safely assume that the null hypothesis is false.

In experiments of statistics one must define a level of significance at which a correlation will be estimated to have been verified, though the option is often actually made after the event. It is significant to know that however small the value of P is there is always a finite chance that the result is a pure accident. A typical level at which the threshold of P is set would be 0.01, which means there is a one per cent chance that the result was accidental. The significance of such a result would then be indicated by the statement $P < 0.01$. Further, in some cases the researcher can use the much lower levels of significance. A level frequently referenced is $P < 0.05$. This means that there is a one in twenty chance that the whole object was accidental.

It is difficult to generalize, but on the whole $P < 0.01$ would normally be considered significant and $P < 0.001$ highly significant. The origin of the $P < 0.05$ criterion goes back to the great pioneer of significance testing, R A Fisher, who did not in fact proved this. Many leading scientists and mathematicians today believe that the emphasis on significance testing is grossly overdone. $P < 0.05$ had become an end in itself and the determinant of a successful outcome to an experiment.

Tests for Statistical Significance

Tests for statistical significance are specifically used to answer the question, such as what is the probability that what we think is a relationship between two variables is really just a chance occurrence? If number of samples is selected from the same population then can we still find the same relationship between these two variables in every sample? If we could do a census of the population, would we also find that this relationship exists in the population from which the sample was drawn? Or is our finding due only to random chance?

Tests for statistical significance tell us what the probability is and also the relationship that occurs only due to random chance. This illustrates what the probability is and what would be the error if it is assumed that the relationship exists. It is always not 100% certain that a relationship exists between two variables. There are too many sources of error to be controlled, for example, sampling error, researcher bias, problems with reliability and validity, simple mistakes, etc. But using probability theory and the normal probability curve, the probability of being wrong can be estimated if it is assumed that the finding a relationship is true. If the probability of being wrong is small, then it is assumed that the observation of the relationship is a statistically significant finding.

Statistical significance means that there is a good chance that one may accurately find that a relationship exists between two variables. But statistical significance is not the same as practical significance. We can have a statistically significant finding, but the implications of that finding may have no practical application. The researcher must always examine both the statistical and the practical significance of any research finding. For example, consider that there is a statistically significant relationship between a citizen's age and the satisfaction level with city recreation services. It may be that older citizens are 5% less satisfied than younger citizens with city recreation services. But is 5% large enough difference to be concerned.

At times, when differences are small but statistically significant which is due to a very large sample size then in a sample of a smaller size, the differences would not be enough to be statistically significant. The following are some significant tests for testing statistical significance.

NOTES

Steps in testing for statistical significance

1. State the research hypothesis
2. State the null hypothesis
3. Select a probability of error level (α or Alpha Level)
4. Select and compute the test for statistical significance
5. Interpret the results

There is always a possibility that the researcher will make a mistake regarding the relationship between the two variables. There are two possible mistakes or errors. The first is called a Type I error. This occurs when the researcher assumes that a relationship exists when in fact the evidence is that it does not. In a Type I error, the researcher should accept the null hypothesis and reject the research hypothesis, but the opposite occurs. The probability of committing a Type I error is called alpha (α). The second is called a Type II error. This occurs when the researcher assumes that a relationship does not exist when in fact the evidence is that it does. In a Type II error, the researcher should reject the null hypothesis and accept the research hypothesis, but the opposite occurs. The probability of committing a Type II error is called beta (β).

Researchers generally specify the probability of committing a Type I error, i.e., the value of alpha. Most researchers select an alpha as 0.05. This means that there is a probability of 5% of making a Type I error assuming that a relationship between two variables exists when it is really does not. However, an alpha of 0.01 is also used when researchers do not want to have a probability of being wrong more than 0.1% of the time or one time in a thousand.

The level of alpha can vary, but the smaller the value, the more stringent the requirement for reaching statistical significance becomes. Alpha levels are often written as the 'P value' or 'P = 0.05'. Usual levels are P = 0.05 or the chance of one in 20 of making an error; P = 0.01 or the chance of one in 100 of making an error; P = 0.001 or the chance of one in 1,000 of making an error. When accounting the level of alpha, it is usually accounted as being 'less than' some level, using the 'less than' sign or '<'. Thus, it is accounted as $P < 0.05$ or $P < 0.01$, etc.

For nominal and ordinal data, Chi-square test is used as a test for statistical significance. To calculate Chi-square, we compare the original, observed frequencies with the new, expected frequencies. *t*-test is considered as the important test for statistical significance and is used with interval and ratio level data. *t*-tests can be used in several different types of statistical tests.

Tests for statistical significance are used to estimate the probability that a relationship observed in the data occurred only by chance and that the probability variables are really unrelated in the population.

4.8 DERIVED SCORES (Z-SCORE, STANDARD SCORE AND T-SCORE)

NOTES

In the words of American psychologists Thorndike and Hagen, 'Norms are defined as the average performance on a particular test made by a standardization sample'. Frank S. Freeman defines a norm in measurement as 'the average or standard score on a particular test made by a specified population'.

Norms are representations of average performance based upon the results of testing a specified group of students. It is a device of transforming raw scores into standard scores in a group. Educational measurement is not a case of physical measurement, but a case of mental measure. Therefore, educational measurement is relative. In case of this, we have to go for the evaluation of an individual in relation to other in a class or a group. Let us take an example. Suppose Prakash is a student of Class X. In the classroom examination, he scored 60 in English and 75 in Mathematics. If a layman analyses the result of Prakash, he will say that Prakash is better in Mathematics than in English. Prakash scored the highest marks in English test but the lowest marks in Mathematics in the class. So the interpretation made by the layman was wrong because the two scores are raw scores.

In order to get a valid and reliable result, we should go for a norm, through which we can get the perfect place of an individual in a group. For this purpose of interpretations, the raw scores are to be converted into derived scores. Norms tell us where a student or a class stands in relation to a reference group. Norms are representation of average performance based upon the results of testing a specified group of students. It is a statistical procedure to minimize the interpretive error of a test score. The norms of any educational test represent the average test performance of the standardized group or sample selected from a specified population. In this, an individual score is compared with the standardized sample as a reference group. The importance of norms in the field of measurement and evaluation are explained as follows:

- Norms are the basis of interpreting raw scores into derived scores.
- Norms place an individual in the exact place within a group.
- Norms are helpful for selection and classification of students.
- Norms are helpful in provision of guidance and counselling the students.
- Norms speak about the attainment of instructional objectives by the students.
- Norms help in minimizing the interpretive error of a measuring instrument.

Types of Norms

In the field of educational and mental measurement, we use four types of norms, which are as follows:

(i) Age norm

The concept of age norm was developed by French psychologist Alfred Binet in 1908. It basically deals with mental age. This age norm is also known as 'mental age

norms' or 'age equivalent norms'. The 'age norm' is defined as the average performance of a representative sample of certain age group on the measure of intelligence or ability. Let us consider a suitable example to have clarity about age norms.

Suppose the average score of students of age 15 years 2 months on an achievement test is 80. So the age norm for the score of 80 will be 15 years 2 months. Suppose Mohan is 12 years old and he scores 80 in the achievement test. Here, though his chronological age is 12, Mohan's mental age is 15 years 2 months.

So, when that age norm is fixed, standardized test is given to a representative sample of students of a particular age level and the average score is calculated, and this score is considered as the norm for the group. The students who achieve that score are considered within that age norm.

Limitations of age norm

The limitations of age norm are as follows:

- It is very difficult to get a true representative sample of individuals of a selected age group.
- In case of very high and very low scores, it is difficult to interpret it with age norms.
- Mental age units are not fixed in case of different tests; it may vary.
- It has the limited scope to be used in some psychological and educational tests.
- Age norms lack a standard and uniform unit throughout the period of growth of physical and psychological traits.
- It is the difficult and a time consuming task to develop the age norms and mental age.
- The mental age of a particular age group may differ from locality to locality and test to test.

(ii) Grade norm

Grade norms are also like age norm. However, here, measurement is based upon class or grade level, not on age level. Grade norms have been widely used with standardized achievement tests, especially at the elementary school level. The grade equivalent that corresponds to a particular raw score is identified as the grade level at which the typical student obtains that raw score. A grade norm corresponding to a raw score is the grade level of those pupils whose average raw score is the raw score in question.

Suppose we conducted a test on the VIIIth grade students. After getting the result, we get 60 to be the average of that test. Therefore, 60 will be the grade norm for the students of VIIIth grade. Grade norms are based on the average performance of students at various grade levels. Grade norms are most useful for reporting growth in the basic skills during the elementary school period. They are least useful for comparing a student's performances on different tests.

NOTES

Limitations of grade norms

The limitations of grade norms are as follows:

- The rate of growth from grade to grade is not uniform throughout.
- Grade norms lack a comparability of scores on different tests.
- When a student of VIIth grade gets a credit of IXth grade, it does not mean that the student has the ability to be a student of IXth grade.
- Fractional grades do not have any meaning.
- Grade norms are affected by quality of schools, quality of teachers and quality of students.
- The interpretation of grade norm is very confusing because it provides only level of performance with respect to a subject rather than the educational level of the students.

(iii) Percentile norm

Percentile norms are about the position of an individual in relation to the norming group. 'Percentile norm' is a point on the scale of measurement determined by the percentage of individuals in given populations that lie below this point. It describes a student's performance in terms of the percentage of other students in some clearly defined group that earn a lower score. This might be a grade or age group, or any other group that provides a meaning comparison. If the percentile norm of a score 60 is 65, we mean that 65 per cent of the students of the normative group lie below a score of 60.

Percentiles should not be confused with the common 'percentage score'. The percentage scores are raw scores, whereas percentiles are transformed scores. It provides a basis for interpreting an individual's score on a test in terms of his own standing in a particular standardization sample. It should be based upon a sample which has been made homogeneous with respect to age group, sex group grade level, socio-economic status, etc. This is applicable for all types of tests: intelligence, attitude, aptitude and achievement.

Limitations of percentile norm

- The percentile units are not equal on all parts of the scale. The percentile difference of 5 near the middle of the scale (e.g., 45 to 50) represents a much smaller difference in test performance than the same percentile difference at the ends (e.g., 85 to 90), because a large number of students receive scores near the middle, whereas relatively few students have extremely high or low scores.
- Percentile norms are generally confused with percentage scores which affects the interpretation.
- Percentile norm indicates only the relative position of an examinee in the standardization sample. It conveys nothing regarding the amount of the actual difference between the scores.
- Percentile rank of one group cannot be compared with percentile rank of another group.

- Conversion of raw scores to percentile scores give more differences in the middle than at the extremes.

We can use the following formula to compute percentile rank of raw scores:

NOTES

$$PP = L + \frac{P_n - f_b}{f_a} \times i$$

where PP = Percentile point of a raw score

L = Lower limit of the raw score falls in the particular class interval

P_n = Percentage of the frequency

f_b = Frequency below the class interval

f_a = Actual frequency of the class interval

i = Size of the class interval

(iv) Standard score norms

The most important method used to indicate an individual's relative position in a group by showing how far his achieved score is above or below average. This is the approach of standard score and standard scores express performance of the individuals in terms of standard deviation units from the mean. There are numerous types of standard scores used in testing. They are as follows:

(a) ***Z-Scores:*** Z-score represents test performance directly as the number of standard deviation units a raw score is above or below the mean. The Z-scores are the units of normal probability curve, ranges from -3σ to $+3\sigma$, with mean value zero and standard deviation one.

The formula for computing Z-score is $= \frac{X - M}{SD}$

Where,

X = Raw score

M = Arithmetic mean of raw scores

SD = Standard deviation of raw score

A Z-score is always negative when the raw score is smaller than the mean.

(b) ***T-Scores:*** T-scores are also the standard scores but the mean value is 50 and standard deviation is 10. T-scores can be obtained by multiplying the Z-score by 10 and adding the product to 50. Thus,

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

One reason that T-scores are preferred to Z-scores for reporting test results is that only positive integers are produced in T-scores.

(c) ***Stanines*****:** The stanine norm is developed by the technique of normalized standard scores. It was developed by the US Air Force during the World War II. Stanines are single digit scores ranging from 1 to 9. This system of scores is so-called because the distribution of raw scores is divided into nine equal parts. Stanine 5 is in the centre of the distribution and includes all cases within

NOTES

one-fourth of a standard deviation of either side of the mean. Here, the mean score is 5 and standard deviation is 1.96 or 2. When raw scores are transformed into stanine scores, the distributions of scores take the shape of normal curve.

In the Stanine system, a 9-point scale is used, in which 9 is high, 1 is low and 5 is average. Stanines are normalized standard scores that make it possible to compare a student's performance on different tests.

The stanines are distributed on normal curve. The area of normal probability curve has been divided into nine standards with a fixed percentage. The first stanine includes 4 per cent, second stanine includes 7 per cent, third stanine includes 12 per cent, fourth stanine includes 17 per cent, fifth stanine includes 20 per cent, sixth stanine includes 17 per cent, seventh stanine includes 12 per cent, eighth stanine includes 7 per cent and the ninth stanine includes 4 per cent of the total cases.

Stanine	*Description*	*Percentage*	*Stanine's position*
1, 9	Bottom and top	4	(lst) (9th)
2, 8	Above bottom and below top	7	(2nd) (8th)
3, 7	Near to second or eighth	12	(3rd) (7th)
4, 6	Above or below mean	17	(4th) (6th)
5	Middle or mean	20	(5th)

Judging the adequacy of norms

- Test norms should be relevant to the reference group.
- Test norms should be comparable with other norms.
- A clear description is necessary for every test norm.
- Test norms should be changed from time to time.
- Test norm should represent the whole group.

Limitations of standard score norms

The limitations of standard score norms are as follows:

- The interpretation of the standard scores becomes problematic when the distribution is not normal.
- It needs expertise knowledge to deal with the standard scores.
- The minus values are very confusing in case of educational and mental measurement.
- Sometimes the raw scores are not normally distributed; they are positively skewed or negatively skewed. This linear transformation is based on the assumption of normal distribution.

4.9 SUMMARY

NOTES

- Processing of data involves analysis and manipulation of the collected data by performing various functions.
- The functions that can be performed on data are as follows:
  - o Editing
  - o Coding
  - o Tabulation
  - o Classification
- Editing of data involves the testing of data collection instruments in order to ensure maximum accuracy. This includes checking the legibility, consistency and completeness of the data.
- Data is classified on the basis of similar features as follows:
  - o Descriptive classification
  - o Simple classification
  - o Manifold classification
- There are certain basic statistical methods that can be classified into the following three groups:
  - o Descriptive statistics
  - o Inferential statistics
  - o Measures of central tendency and dispersion
- There are also two chief statistical methods based on the tendency of data to cluster or scatter. These methods are known as measures of central tendency and measures of dispersion.
- Analysis of data is the process of transformation of data for the purpose of extracting some useful information, which in turn facilitates the discovery of some useful conclusions.
- Data interpretation refers to the identification of trends in different variables. The researcher uses statistics for this purpose. The researcher is required to be familiar with the knowledge of the scales of measurement.
- Measurement is the process of assigning numbers or symbols or grades to objects or events in a well-designed manner. There are four scales of measurements, listed as follows:
  - o Nominal scale
  - o Ordinal scale
  - o Interval scale
  - o Ratio scale
- After the collection of data, it requires to be presented in a systematic form so that it can be managed properly. Raw data gives no full meaning until it is arranged properly; hence, it is very important to present data in arranged form. There are two methods of presenting data. These are:

NOTES

(i) Grouping or tabular presentation

(ii) Graphical and diagrammatic presentation

- There are several commonly used measures of central tendency, such as arithmetic mean, mode and median. These values are very useful not only in presenting the overall picture of the entire data but also for the purpose of making comparisons among two or more sets of data.
- Correlation analysis is the statistical tool generally used to describe the degree to which one variable is related to another. The relationship, if any, is usually assumed to be a linear one. This analysis is used quite frequently in conjunction with regression analysis to measure how well the regression line explains the variations of the dependent variable.
- Percentile is the point below which the given percent of cases lies. The method of finding out the percentile is essentially the same as we have seen and used in computing median and quartile deviation.
- A measure of skewness gives a numerical expression and the direction of asymmetry in a distribution. It gives information about the shape of the distribution and the degree of variation on either side of the central value.
- Kurtosis is a measure of peakedness of a distribution. It shows the degree of convexity of a frequency curve.
- If the normal curve is taken as the standard, symmetrical and bell-shaped curve then kurtosis gives a measure of departure from the normal convexity of a distribution. The normal curve is mesokurtic. It is of intermediate peakedness. The flat-topped curve, broader than the normal, is termed platykurtic. The slender, highly peaked curve is termed leptokurtic.
- In probability theory, the normal probability curve or normal (or Gaussian) distribution is considered as the most frequently occurring continuous probability distribution.
- Normal distributions are exceptionally significant in statistics and are typically used in the context of natural and social sciences generally for real-valued random variables whose distributions are not known.
- The curve of normal distribution is illustrated using the normal probability curve and is the most common type of distribution.
- Statistical significance is the result that is not likely to occur randomly, but rather it is likely to be attributable to a specific cause. Statistical significance can be strong or weak and is important feature of research in many mathematics and science related fields.
- Statistical significance does not always indicate practical significance. In addition, it can be misinterpreted when researchers do not use language carefully in reporting their results.
- In words of American psychologists Thorndike and Hagen, 'Norms are defined as the average performance on a particular test made by a standardization sample'. Frank S. Freeman defines a norm in measurement as 'the average or standard score on a particular test made by a specified population'.

4.10 KEY TERMS

NOTES

- **Data:** It refers to facts and statistics collected together for reference or analysis.
- **Median:** Something that denotes or relates to a value or quantity lying at the midpoint of a frequency distribution of observed values or quantities, such that there is an equal probability of falling above or below it.
- **Mode:** It means the type of average that refers to the most-common or most-frequently occurring value in a series of data.
- **Mean:** It refers to the mean or average that is used to derive the central tendency of the data in question.
- **Correlation analysis:** It is the degree and type of relationship between any two or more quantities (variables) in which they vary together over a period.
- **Normal probability curve:** It refers to a family of distributions that are bell shaped.
- **Kurtosis:** It is a measure of peakedness of a distribution. It shows the degree of convexity of a frequency curve.

4.11 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The functions that can be performed on data are as follows:
   - Editing
   - Coding
   - Tabulation
   - Classification
2. Coding of data can be defined as representing the data symbolically using some predefined rules.
3. According to Smith, descriptive statistics is the formulation of rules and procedures where data can be placed in a useful and significant order. The foundation of applicability of descriptive statistics is the need for complete data presentation.
4. Univariate analysis is performed with the purpose of describing each variable in terms of mean, median or mode, and variability.
5. Measurement is the process of assigning numbers/symbols/grades to objects or events in a well-designed manner.
6. When the mass of collected raw data is arranged in an order of magnitude, in increasing order or decreasing order, it is called an array.
7. Numerical data may be classified as cardinal and ordinal data whereas in cardinal data, absolute difference between any two given numbers is meaningful.

NOTES

8. Tate defined the measures of central tendency as, 'A sort of average or typical value of the items in the series and its function is to summarize the series in terms of this average value'.
9. Median is the value of the item which is located at the centre of the array.
10. Mode can be defined as the most frequently occurring value in the data. It is not affected by extreme values in the data and can be easily obtained from an ordered set of data.
11. The theory by means of which quantitative connections between two sets of phenomena are determined is called the 'Theory of Correlation'.
12. Correlation analysis helps us in determining the degree to which two or more variables are related to each other.
13. Skewness refers to a lack of symmetry in a distribution.
14. In a skewed distribution, the distance between the mean and the median is nearly one-third of that between the mean and the mode.
15. Normal distributions are exceptionally significant in statistics and are typically used in the context of natural and social sciences generally for real-valued random variables whose distributions are not known.
16. The Normal Probability Curve (NPC), simply known as normal curve, is a symmetrical bell-shaped curve. It is based upon the law of probability.
17. The steps in testing for statistical significance are as follows:
    (i) State the research hypothesis
    (ii) State the null hypothesis
    (iii) Select a probability for error level
    (iv) Select and compute the test for statistical significance
    (v) Interpret the results
18. In the words of American psychologists Thorndike and Hagen, 'Norms are defined as the average performance on a particular test made by a standardization sample'. Frank S. Freeman defines a norm in measurement as 'the average or standard score on a particular test made by a specified population'.

4.12 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What does the editing of data involve?
2. What is inferential statistics?
3. List the characteristics of mean.
4. How is median for ungrouped data calculated?
5. List the assumptions of Karl Pearson's method.

6. When is quartile deviation and standard deviation used?
7. What is normal probability curve? What are its characteristics and application?
8. What are norms? Discuss its different types.

Long-Answer Questions

1. Discuss how data is represented on a graph.
2. Illustrate how arithmetic mean is calculated with the help of an example.
3. What are the advantages and disadvantages of mode? Describe how mode is calculated.
4. What is coefficient of determination? Illustrate with the help of an example how it can be calculated?
5. Discuss with the help of an example how relative measures of skewness can be calculated.
6. What is statistical significance? Discuss the different tests for statistical significance.

NOTES

4.13 FURTHER READING

Bloom, B. S. *Taxonomy of Objectives of Cognitive Domains*. New York: Longmans Green &Co.

Ebel, R.L. *Essentials of Educational Measurement.* New Jersey: Englewood Cliff.

Grunlund, N. E. *Measurement and Evaluation in Teaching* (3rd Ed). New York: Mcmillan Publishing Co.

Julian, C., Stanley and Hopkins. *Educational and Psychological Measurement and Evaluation*. New Delhi: Prentice Hall of India Pvt. Ltd.

Lindermon, Richard, H. *Educational Measurement*. Mumbai: B. S. Taraporevala Sons & Co.

NOTES

UNIT 5 STATISTICS IN MEASUREMENT AND EVALUATION-II

Structure

5.0 INTRODUCTION

In the previous unit, you were introduced to some statistical concepts in measurement and evaluation. In it, the measures of central tendency and variability were discussed. Measures of central tendency are of various types, such as arithmetic mean, mode and median. Also discussed was the normal probability curve and the coefficient of correlation. In this unit, the discussion on statistical concepts in measurement and evaluation will continue. In it, you will learn about the methods of determining different reliability coefficients, as well as validity coefficients. The unit will also discuss the emerging trends in evaluation.

5.1 UNIT OBJECTIVES

After going through this unit, you will be able to:

- Discuss the different methods of determining reliability coefficient
- Describe the different methods of determining validity
- Examine the relation between reliability and validity
- Discuss the concept of grading
- Describe the use of computers in evaluation

5.2 RELIABILITY: CONCEPT AND DETERMINING FACTORS

NOTES

Reliability refers to the consistency of measurement, which is how stable test scores or other assessment results are from one measurement to another. It means the extent to which a measuring device yields consistent results upon testing and retesting. If a measuring device measures consistently, it is reliable. The reliability of a test refers to the degree to which the test results obtained are free from error of measurement or chance errors.

Characteristics

The characteristics of reliability are as follows:

- It refers to the degree to which a measuring tool yields consistent results upon testing and retesting.
- It indicates the level to which a test is internally consistent, i.e., how accurately the test is measuring.
- It refers to the results obtained with measuring instrument and not to the instrument itself.
- An estimate of reliability refers to a particular type of stability with the test result.
- Reliability is necessary but not a sufficient condition for validity.
- Reliability is a statistical concept.
- It refers to the preciseness of a measuring instrument.
- It is the coefficient of internal consistency and stability.
- It is the function of the length of a test.

Factors affecting reliability

The reliability of a test is affected by a couple of factors which are explained in the following manner:

(i) Length of the test: There is positive correlation between the number of items in a test and the reliability of a test. The more the number of items the test contains, the greater is its reliability. In several tests, the scores of sub-tests and whole tests are calculated separately and their reliability is also calculated separately. The reliability of the whole test is always more than the sub-test, because whole test means more items, which is better representation of the content.

(ii) Construction of the test: The nature of items, their difficulty level, objectivity of scoring, item interdependence and alternative responses are factors which affect the reliability. More alternative responses will increase the reliability of the test.

NOTES

(iii) **Nature of the group:** Reliability of a test will be more if the test is administered to a heterogeneous group. The more the variability, the higher the reliability coefficient.

(iv) **Testing conditions:** If the testing conditions are not similar at all the places, then differences in scores are obtained. The physical conditions of the tests and the environmental factors around the test-taker affect the reliability of a test.

(v) **Guessing and chance errors:** Guessing paves the way to increase error variances and it reduces reliability. If there are more opportunities for guessing in the test, the test will yield less reliable results.

(vi) **Test instructions:** If instructions in the test are complicated or difficult to understand, there will be less consistency in the scores. If the test-taker will not understand the instruction properly, his way of response will be wrong and this will hamper the reliability of test.

(vii) **Too easy or too difficult items:** Too easy or too difficult items fail to distinguish between good and bad students which otherwise affects the reliability of a test.

The other factors which affect the reliability of tests are: subjectivity of the examiner, clerical error, interval between testing, effect of practice, etc.

5.2.1 Methods of Determining Different Reliability Coefficient

When examining the reliability coefficient of standardized tests, it is important to consider the methods used to obtain the reliability estimates. American Psychological Association (APA) introduced several methods of estimating reliability. The methods are similar in that all of them involve correlating two sets of scores, obtained either from the same assessment procedure or from equivalent forms of the same procedure. The chief methods of estimating reliability are shown here. The reliability coefficient resulting from each method must be interpreted according to the type of consistency being investigated.

We will consider each of these methods of estimating reliability in detail in Table 5.1.

(i) Test-retest method

To estimate reliability by means of the test-retest method, the same assessment is administered twice on the same group of pupils with a given time interval between the two administrations. The resulting assessment scores are correlated and this correlation coefficient provides a measure of stability, i.e., it indicates how stable the assessment results are over a given period of time. If the result shows that the students who were at good position at the first administration, and also they are at good position in the second administration, the test has stability. The closer the agreement between the two administrations of the test, the greater is the reliability co-efficient or co-efficient of stability.

NOTES

Table 5.1 *Methods of Estimating Reliability*

Method	*Types of reliability measure*	*Procedure of administration*
Test-Retest method	Measures of stability and precision	Use the same test twice to the same group with anytime interval between tests, from several minutes to several years.
Equivalent forms method	Measure of equivalence	Apply two equal forms of the test to the same group in close time gap.
Split-Half method	Measure of internal consistency	Apply the test once. Score two equivalent halves of test (e.g. , odd items and even items), correct correlation between halves to fit whole test by Spearman–Brown formula to measure reliability of the test.
Kuder–Richardson method	Measure of Internal consistency	Give test once, score total and apply Kuder–Richardson formula to know the degree of reliability.
Inter-rater method	Measure of consistency	Use a set of student response requiring judgmental scoring to two or more raters and have them independently score the responses.

A high test–retest reliability or coefficient of stability shows that there is low variable error in the sets of obtained scores. A low coefficient of stability indicates that there is high variable error in the obtained scores. The error variance contributes inversely to the coefficient of stability. Such stability is indicated by a large correlation coefficient, that a perfect positive relationship is indicated by 1.00, perfect negative correlation by –1 and no correlationship by 0.00. Measures of stability in the 0.80 range are generally reported for standardized tests. The Pearson's method of correlation is mostly used for obtaining the coefficient of reliability.

When we are dealing with the measures of stability, it is more important to keep in mind about the time gap between assessments. If the time gap is short, say a day or two, the consistency of the results will be inflated because a student remembers the tasks and responses accordingly. If the time interval is a long one, the results will be influenced not only by the instability of the assessment procedure but also by actual changes in the student over that period of time. The best time interval between assessment administrations depends largely on the use to be made of the results.

Limitations of test–retest method

The limitations of test-retest method are as follows:

NOTES

- Testing conditions during test and retest may vary which result in instability in the scores.
- The individual's health, emotional conditions, mental health, motivational conditions and mental set do not remain the same in both the administrations at two occasions.
- If the test is repeated immediately, the test-taker may recall the first answer. This may increase the scores. Besides the memory effects, practice and confidence induced by familiarity with the test will almost affect the scores the second time. This is known as 'the carry-over effect' or 'transfer effect' or 'memory effect' or 'practice effect'.
- If the time gap between two administrations is a lengthy one, additional learning or changes in the characteristics of the individual will affect the scores at the later administration, and it may decline the reliability of a test.

(ii) Equivalent forms method

This method of reliability is also known as parallel forms reliability/equivalent forms reliability/comparable forms reliability. Two parallel forms of a test can be constructed by selecting the samples in each form from the same universe of contents. By parallel forms, we mean that the forms of tests are equivalent as far as the content, objectives, format, difficulty level and discriminating value of items, length of the test, etc., are concerned. Parallel tests have equal mean scores, variance and equal inter-correlation among items. So the two parallel forms of a test must be similar in all respects, but there should not be a duplication of test items. The equivalent forms method of establishing reliability is widely used in standardized testing because for most standardized tests, two or more forms are available when the two forms are virtually alike, reliability is too high; when they are not sufficiently alike, reliability will be too low. Pearson's method of correlation is used for calculating the coefficient of correlation between two sets of scores obtained by administering the two forms of the test. This coefficient of correlation is termed as the 'coefficient of equivalence'.

Limitations of equivalent forms method

The limitations of equivalent forms method are as follows:

- Construction of equivalent test forms is too difficult in comparison to others.
- There is a chance of memory effect or practice effect to operate at the administration of the second form.
- The testing conditions while administering the two forms may be different and the test-takers may not be in a same physical, mental or emotional state at both the time of administration.

(iii) Split-half method

Reliability can also be calculated from a single administration of a single test. The test is administered to a group of pupils and then is divided into half for scoring. To divide the test into halves that are equivalent, the procedure is to score the

NOTES

even-numbered and the odd-numbered items separately. This results two scores for every student which when correlated, gives a measure of internal consistency. This coefficient indicates the degree to which consistent results are obtained from the two halves of this test and may be thought of as the half-length test reliability estimate.

Among the different methods of dividing the test, the odd and even split is used mostly. Here, items no.1, 3, 5, 7, 9, 11, etc., (the odd-numbered items) constitute one half, and items no. 2, 4, 6, 8, 10, etc., (the even-numbered items) constitute the other half of the test. Scores of individuals on these two halves are correlated to know the degree of reliability between the two. The coefficient of correlation is usually denoted by the symbol (r_{hh}). The reliability coefficient of the whole test (r_{xx}) is estimated from the value of r_{hh} using the following:

Spearman-Brown formula:

$$r_{xx} = \frac{2r_{hh}}{1 + r_{hh}}$$

$$= \frac{2 \times \text{reliability of half test}}{1 + \text{reliability of half test}}$$

Where,

r_{xx} = Reliability of whole test

r_{hh} = Reliability of half test

Limitations of split-half method

The limitations of split-half method are as follows:

- The whole test can be divided into two parts in a number of ways. As such, reliability coefficient obtained through this method may vary from method to method.
- As the test is once administered, the factor chance error may affect the scores on the two halves, and thus, tend to make the reliability coefficient too high.
- It is very difficult to divide the test into two halves in such, way that both halves are of equivalent difficulty levels and discriminating power.
- This method cannot be used in power test.

(iv) Kuder–Richardson method

This method of reliability was developed by G. F. Kuder and M. N. Richardson. This is also known as rational equivalence method of reliability. It is most useful for homogenous test reliability. Like the split-half method, the Kuder–Richardson method provides an internal consistency but it does not require splitting the assessment in half for scoring purposes. This method enables us to calculate the inter-correlation of the items of the test and the correlations of the items with the test as a whole. All

NOTES

the items in the test measure the same ability—the correlation between the items are equal, that all the items are of same difficulty and that all the items are highly homogeneous. When we use Kuder–Richardson formula, it is required that all the items of the test should be psychologically homogeneous and that every item in the test has a high correlation with every other item. This is referred to as inter-item consistency. This coefficient is called the 'coefficient of rational equivalence'. This coefficient provides some signals of how internally consistent or homogeneous the items of the test are.

Kuder–Richardson formula used to compute the internal consistency coefficient of test items is:

$$r_{tt} = \left[\frac{n}{n-1}\right]\left[\frac{\sigma x^2 - \Sigma pq}{\sigma x^2}\right]$$

Where,

r_{tt} = Reliability index

n = Number of items in the test

p = Proportion of right responses

q = Proportion of wrong responses

Example 5.1: A test consisting of 50 items and standard deviation of test score is 7.5 and the sum of the product of proportion of right and wrong responses on the item is 10.43. Calculate the reliability.

Solution:

$$r_{tt} = \left[\frac{n}{n-1}\right]\left[\frac{\sigma x^2 - \Sigma pq}{\sigma x^2}\right]$$

$$= \left[\frac{50}{50-1}\right]\left[\frac{7.5^2 - 10.43}{7.5^2}\right]$$

The reliability coefficient of the test is 0.84.

The KR-21 method is also a simple one for calculating reliability of a test. The test is administered once on a group to determine quickly the reliability. The mean and variance are calculated and then the following formula KR-21 is used.

$$r_{tt} = \frac{n\,\sigma\,x^2 - M\,(n - M)}{\sigma x^2\,(n-1)}$$

Where,

r_{tt} = Reliability of the whole test

n = Number of items in the test

M = Mean of the test scores

σx = Standard deviation

NOTES

Example 5.2: An objective test of 100 multiple items have been administered to a small group of students. The mean of test score is 50 and standard deviation is 10. Calculate the reliability coefficient of the test.

Solution:

$$r_{tt} = \frac{n\sigma x^2 - M\,(n - M)}{\sigma x^2\,(n-1)}$$

$$= \frac{100 \times 10^2 - 50\,(100 - 50)}{10^2\,(100 - 1)}$$

$$= \frac{100 \times 100 - 50\,(50)}{100 \times 99}$$

$$= \frac{10000 - 2500}{9900}$$

$$= 0.76$$

The reliability coefficient of the test is 0.76.

Limitations of Kuder–Richardson method

The limitations of Kuder–Richardson method are as follows:

- Kuder–Richardson formulae are not suitable for speed assessments—assessments with time limits that prevent students from attempting all the items.
- The formulae indicate the consistency of student response from one day to the other.
- It cannot be used for power test and heterogeneous tests.
- The different Kuder–Richardson formula results differ in reliability coefficient.
- In case all the items of the tests are not highly homogeneous, this method will produce lower reliability coefficient.

(v) Inter-rater method

This method assesses reliability through scoring/evaluating, done by two or more independent judges for every test.The various scores given by the judges are then compared to determine the consistency of estimations.The way the comparison is carried out is: each rater assigns each test item a score, which would be on a scale from 1 to 10.Then the correlations between any two ratings are calculated. There is another method for testing inter-rater reliability. In this method, raters identify a category for each observation and then compute the percentage of agreement among the raters. For instance, if the raters are in agreement 7 times out of 10, the test will be said to have a 70 per cent inter-rater reliability rate.

In case the raters seem to be in disagreement, it would imply that either the raters need to be trained again or the scale is defective. Sometimes, it so happens

that various raters would have different opinions about measurement results emerging from the same object, such as a scientific experiment or a test, wherein first the test is carried out, then its results are interpreted, recorded and presented. At any of these stages, the rater may become affected by rater's bias (the tendency to rate in the direction of what the rater expects). There may also be discrepancy during interpretation and presentation of results, for instance, the round-off may be different in terms of higher or lower digit next to the decimal.

Limitations of inter-ratermethod

The limitations of inter-rater method are as follows:

- The method can be tedious because inter-rater reliability statistics need to be calculated separately for every item and every pair of raters.
- It is a lengthy and difficult task to train the raters such that they are able to reachan exact agreement.
- Even when they are trained, the forced consensus might render the ratings inaccurate and this would be a threat to the validity of the student's scores.
- The resulting estimates might turn out to be too conservative if two raters show differences in the method used on the rating scale.

CHECK YOUR PROGRESS

1. Define reliability.
2. On what does the best time interval between assessment administrations depend?

5.3 VALIDITY: CONCEPT AND USES

The validity of a test is determined by measuring the extent to which it matches with a given criterion. It refers to the very important purpose of a test, and it is the most important characteristic of a good test. A test may have other merits, but if it lacks validity, it is valueless.

Characteristics of Validity

The characteristics of validity are as follows:

- Validity is a unitary concept.
- It refers to the truthfulness of the test result.
- In the field of education and psychology, no test is perfectly valid because mental measurement is not absolute but relative.
- If a test is valid, it is reliable; but if a test is reliable, it may or may not be valid.
- It is an evaluative judgment on a test. It measures the degree to which a test measures what it intends to measure.

NOTES

- It refers to the appropriateness of the interpretation of the result, and not to the procedure itself.
- It refers to degree means high validity, moderate validity and low validity.
- No assessment is valid for all the purpose. A test is valid for a particular purpose only.

Types

Validity is a specific characteristic of a test. There are different ways of calculating the validity of a test based on the objectives with which we validate the test. The six types of validity are discussed below.

(i) Face validity: When a test appears to measure what the test user wishes or intends to measure, it is said to possess face validity. Thus, face validity refers not to what the test measures, but what the test appears to measure, i.e., whether it seems to be relevant to its various objectives. Test contents should not appear to be inappropriate or irrelevant. If a test measures what the test author desires to measure, we say that the test has face validity. Face validity does not require any statistical technique. It is based on subjective judgment. Whenever a test is prepared, it is submitted to experts to assess if it measures what it intends to measure. It is the first step in validating the test. Once the test is validated at face, we may proceed further to compute validity coefficient. For example, suppose we prepare a test to measure 'skill in division'. If all the questions on the test are related to division, we can say that the test has face validity.

(ii) Content validity: An analysis of the content of an assessment evaluates the appropriateness of the content and determines the extent to which the assessment tasks provide a relevant and representative sample of the content under consideration. Content considerations are especially very important when validating achievement testing or constructing classroom tests. Content validity refers to the extent to which a test contains items representing the behaviour that we are going to assess. It is generally assessed by a critical scrutiny of the test contents to determine whether they cover a representative sample of behaviour or not. As it is measured from the content, such type of validity is referred to as content validity. Content validity of a test is estimated by assessing the presence of the content to be assessed in the test paper. Suppose we want to construct an achievement test on English. So all the questions of the test should be related to prose, poetry, novel, grammar, etc., and all the items must measure the different behavioural objectives like knowledge, understanding, application, analysis, synthesis and evaluation.

Content validity rests upon an expert analysis on the items included in the test. Content validity is also known as 'curricular validity' or 'rational validity' or 'logical validity'.

(iii) Concurrent validity: The term 'concurrent validity' is used to refer the process of validating a new test by correlating it with some available source of information which might have been obtained shortly before or shortly after the new test is given. Concurrent validity indicates to which extent the test

NOTES

scores correspond to already accepted measures of performance (or status made at the same time). Suppose we want to administer an intelligence test upon an individual. Now the test is administered upon the individual and the intelligence scale is administered upon the same individual. If the coefficient of correlation is high, the intelligence test is said to have high concurrent validity. Concurrent validity is relevant to the tests employed for diagnosis. When new tests are validated against previous tests, these previous or established tests are known as criteria for the new tests.

(iv) Construct validity: A 'construct' is an individual characteristic that we assume exists in order to explain some aspect of behaviour. Whenever we wish to interpret assessment results in terms of some individual characteristics (e.g., reasoning, problem-solving activity), we are concerned with a construct. The construct validity of a test is the extent to which the test may be said to measure a theoretical construct or trait. Examples of such construct are scholastic aptitude, mechanical comprehension, anxiety, neuroticism, etc. Construct validation requires the gradual accumulation of information from a variety of sources. When we interpret assessment results as a measure of a particular construct, we are implying that there is such a construct that differs from other constructs, and that the results provide a measure of the construct, i.e., little influenced by extraneous factors. Verifying such implications is the task of construct validation. Although construct validation has been commonly associated with theory building and theory testing, it also has usefulness for the practical use of assessment results. It takes place primarily during the development and try-out of a test or an assessment, and is based on an accumulation of evidence from many different sources. When selecting a published test that presumably measures a particular construct, such as logical reasoning or writing comprehension, the test manual should be tested to determine what evidence is represented to support the validity of the proposed interpretations.

(v) Predictive validity: Predictive validity of a test refers to the predictive capacity of a test. It refers to the effectiveness of a test in predicting future outcomes in a particular area.The word 'prediction' may be used in more ways than one. In the broader sense, it would mean prediction from the test to a criterion situation. In the limited sense, however, it would imply prediction over a time period. When used in the limited context, it may be referred to in the sense of 'predictive validity'. Predictive validation provides the most relevant information for tests used in the selection and classification of personnel. Some other uses include hiring job applicants, selecting students for admission to college or professional schools, and assigning military personnel to occupational training programmes. Take an example where we have prepared an entrance test for admission into B.Ed. course, and based on the scores, we have admitted the candidates. These candidates completed the B.Ed. course and appeared for the final B.Ed. examination. The scores of

NOTES

the entrance test and the scores of the final B.Ed. examination are correlated. If the coefficient of correlation is high, we say that the entrance test has high predictive validity.

(vi) Criterion validity: Criterion validity evidence tells us just how well a test corresponds with particular criteria. Criterion validity includes both predictive validity and concurrent validity. A test may be validated against as many criteria as there are specific uses for it. Whenever test scores are to be used to predict future performance or to estimate current performance on some valued measure other than the test itself (called a criterion), we are especially concerned with evaluating the relationship between the test and the criterion. For example, speaking-readiness test scores might be used to predict a student's future a achievement in speaking, or a test of dictionary skills might be used to estimate the student's current skills in the actual use of the dictionary. The first example is of predictive validity and the second example is of concurrent validity.

5.3.1 Determining Validity Co-efficient

Different types of validity needs different methods for assessment. The methods used for assessing the validity of a test are discussed below.

(i) Correlation method: Correlation method is used in most of the cases for calculating the validity of a test. Multiple correlation is used where more than two measures are involved in which English mathematician and biometrician Karl Pearson's 'r' is used. Several methods are used for this, but the following method is a popular one.

$$r = \frac{N\sum XY \times \sum X \sum Y}{\sqrt{[N\sum X^2 - (\sum X)^2]\,[N\sum Y^2 - (\sum Y)^2]}}$$

Where,

r = Validity index

N = Size of sample

X = Raw scores in the test X

Y = Raw scores in the test Y

XY = Sum of the products of each X score multiplied with its corresponding Y score

In order to make the calculation an easy one, we can use the above formula in this way:

$$r = \frac{N\sum dxdy - \sum dx \sum dy}{\sqrt{[N\sum dx^2 - (\sum dx)^2]\,[N\sum dx^2 - (\sum dx)^2]}}$$

Where,

r = Validity index

N = Size of the sample

NOTES

dx = X – M

dy = Y – M

X = Raw score of X group

Y = Raw score of Y group

M = Mean

Sometimes, we need to predict the future result of somebody with reference to the present result. The following regression equation is used for this purpose:

$$y = r\frac{\sigma_y}{\sigma_x}(X - Mx) + My$$

Where, Y = Predicted value

My = Mean of predicted score (Y)

Mx = Mean of test score (X)

σ_y = Standard deviation of predicted score (Y)

σ_x = Standard deviation of test score (X)

X = Test scores (basis of prediction)

Y = Predicted value

(ii) Cross validation: Cross validation indicates a process of validating a test by using a population sample that is different from the sample on which it was originally standardized. It is necessary because the validity data may be high or low due to chance factors peculiar to the standardization sample. When test is administered to various samples in a variety of situations, it is being cross-validated. The different types of cross validation are: validity extension, validity generalization and psychometric signs.

5.3.2 Relation between Validity and Reliability

Numerous factors tend to make assessment results invalid for their intended use.

Factors affecting validity

Let us have a discussion regarding the factors which affect the validity of a test.

(i) Lack of clarity in directions: Instructions that do not clearly indicate to the student how to respond to the tasks and how to record the responses decreases the validity of a test. If the directions are not clear, the students will misunderstand the purpose of the test, and this in turn will hamper the validity of the test.

(ii) Ambiguity: Ambiguous statements lead to confusion and misinterpretation. Ambiguity sometimes confuses the good students more than it does the poor students. So no question of the test should be ambiguous.

NOTES

(iii) **Length of the test:** Lengthening of a test not only increases its reliability but also its validity. If a test is too short to provide a representative sample of the performance we are interested in, its validity will suffer accordingly.

(iv) **Nature of the group:** A test may be valid for one group, but may not be valid for another. In heterogeneous groups, the scores are widely distributed and the validity coefficients are likely to be higher.

(v) **Difficult reading vocabulary and sentence structure:** Vocabulary and sentence structure that are too complex for the students taking the assessment, result in the measuring of reading comprehension and aspects of intelligence, which will lessen the validity of a test.

(vi) **Inadequate time:** Sometimes, the importance is given upon speed test instead of power test. In the field of achievement test, if sufficient time will not be given to the students, it reduces the validity of interpretations of results. However, assessments of achievement should minimize the effects of speed on student performance.

(vii) **Poorly constructed test items:** Sometimes, the test items are very subjective, vague, unclear, not objective, etc., and this affects the validity of a test.

(viii) **Improper arrangement of items:** The items in a test should be arranged according to the difficulty order. It means the items should be arranged from easy to difficult. If difficult items are placed first, it will take time and makes the student confused. So the items of a test should be arranged properly in order to develop the validity of a test.

(ix) **Identifiable pattern of answers:** Correct answers in some systematic order enables a student to guess the right answers more easily, and it affects the validity of a test.

(x) **Factors in administration and scoring:** In the case of teacher-made tests, the factors like insufficient time, unfair aid to individual students who ask for help, cheating and unreliable scoring by students tend to lower validity. In case of a standardized test, failure to follow the standard directions and time limits, or giving students unauthorized assistance or errors in scoring similarly contribute to lower validity. So, all these factors should be checked for ensuring validity of a test.

(xi) **Cultural influences:** Cultural influence, socio-economic status, social class structure, etc., affect the test scores as well as validity of a test.

(xii) **Criterion correlated to the test:** The criterion, for which the validity is assessed, should be a reliable one and free from bias, or else it will affect the validity.

Relationship between Validity and Reliability

Reliability and validity are closely related, even though they cannot be interchanged. An assessment that has very low reliability will also have low validity; quite obviously, a measurement that has low levels of accuracy or consistency is not likely to suitably

NOTES

fulfil its objective. However, at the same time, the factors necessary for achieving a considerably high degree of reliability can affect validity negatively. For instance, consistent assessment settings lead to better degree of reliability because they minimize the 'noise' (variations) in the results. Alternatively, something that can enhance validity is flexibility in assessment exercises and settings. Such flexibility enables assessment to be customized to the learner's context and to be made specific and suitable for certain groups of students. Insistence on entirely consistent assessment settings for achieving high reliability will lead to minimum flexibility, and may, in turn, bring down validity.

For every dimension of interest and specific query or set of queries, there are multiple ways to devise questions. Even though the focal point should always remain the predetermined purposes of the research, there are good or poor questions for any particular operationalization. How can you evaluate the measures?

Two chief criteria of evaluation in any measurement or observation are as follows:

1. If we are measuring what we set out to measure
2. If the same measurement process always shows the same results

These two criteria are validity and reliability.

The key concerns of 'reliability' are stability and consistency. This means that if you are using a certain tool to make measurements, is that tool giving you the same result each time you use it? To understand the concept more clearly, consider measurement processes in different contexts, such as for woodwork or construction work, a tape measure is an extremely reliable measuring tool.

Suppose you have a piece of wood that is 2½ feet long. You measure it once with the tape measure and find that it is of 2½ feet. Suppose you measure it again and you get the same result—2½ feet. If you measure it over and over again and you get the measurement of 2½ feet each time, this means that the tape measure is capable of giving you reliable results.

Validity is concerned with the extent of measurement of what we hope to measure (and what we think we are measuring). To continue with our previous example of measuring the wood block, a tape measure that shows accurately spaced inches, feet, etc., should show up valid results as well. Measuring the wooden block with a 'reliable' tape measure should give an accurate measurement of the block's length.

When applying these concepts to educational research, it is better to use measurement tools that are both reliable and valid. So, this means that a test should have questions that receive consistent answers even when asked a number of times– this is reliability. Similarly, a test should contain questions that get correct responses from respondents – this is validity.

NOTES

CHECK YOUR PROGRESS

3. How is validity of a test determined?
4. What does concurrent validity indicate?
5. When is correlation method used?

5.4 TRENDS IN EVALUATION: GRADING, CREDIT SYSTEM, CUMULATIVE RECORD CARD

Grades in the realm of education are standardized measurements of varying levels of comprehension within a subject area. Grades can be assigned in letters (for example, A, B, C, D, or E, or F), as a range (for example 4.0–1.0), as a number out of a possible total (for example out of 20 or 100), as descriptors (excellent, great, satisfactory, needs improvement), in percentages, or, as is common in some post-secondary institutions in some countries, as a Grade Point Average (GPA). Most nations have individual grading systems unique to their own schools. However, several international standards for grading have arisen recently.

Concept of Grading

The old Indian education system was changed as per the need of time and a new system was introduced in 2009–10 after studying the consequences of the existing system. This new system is called the grading system which is not only concerned with academics but other skills of the students are also considered in this system. Its main purpose is to reduce the stress of the child and undertake continuous evaluation in various areas.

The grading system focuses only on academics; it also provides opportunity to a child to explore his other abilities. For this, it gives stress on practical aspects of learning. This system provides a varied range of opportunities, so that the students can prove their talent and pursue their interests traditionally.

The implementation of the system encourages students to cope with the stress. Students are evaluated on a 9-point grading system, which will diminish the difference between a student scoring 99 per cent and the one scoring 91 per cent. Both students will get the A+ grade.

Merits of grading system

The merits of the grading system may be summarized as follows:

- It is useful in comparing the result of various examination institutions, because score got by the students are not considered to be very relevant in comparison to grades.
- It is more reliable technique of evaluation.
- It works as a common scale for comparing educational achievement of the students of various subjects of faculties.

NOTES

- It helps students in the selection of future course as per their aptitude and ability.
- It minimizes discrimination of students on the basis of marks.
- It helps in eliminating unhealthy competition among high achievers.
- It provides more flexibility for the learner by reducing pressure of competition.
- It creates a better learning environment with a feeling of positive competition and create stress free learning environment.

A grading system must be comprehensive and take into account all aspects of a student's development, besides the marks scored in tests and examinations. To achieve an effective, all-round grading system:

1. Teachers must collaborate on it and come up with the grading system jointly so that everyone's style of teaching is included.
2. All the class projects, unit tests, homework assignments must be factored into the grading system so that everything the student contributes is taken into account and the progress monitored. Each of these should be assigned a percentage as per its learning value and the total percentage must add up to 100%.
3. There should be a section, which lists the penalty deduction of marks for factors such as lateness, sloppy work, lack of concentration in class etc.
4. The basic points of the grading system should be shared with the students so that they know how to improve their grades and the areas on which they need to focus or they may lose points.
5. Exceptional performance must be rewarded.

Types and application of grading

The Ministry of Human Resources Development (HRD) and the CBSE have made the board examination optional from the year 2010-11. They have started a new system of evaluation called 'Continuous and Comprehensive Evaluation (CCE)' that is totally based on grades. It is based on the twice-a-year evaluation system, i.e, two terms in one academic session. Continuous and Comprehensive evaluation refers to a system of school based assessment that covers all aspects of student's development. It emphasizes two fold objectives: continuity in evaluation and assessment of broad- based learning, and behavioral out come.

The grading assessment system includes a variety of tools and techniques for assessment of the learners.

1. Scholastic evaluation: It includes:

- Academic evaluation
- Work experience
- Physical and health education
- Art education

2. Co- scholastic evaluation: It includes:

- Life skills
- Attitudes and values
- Outdoor activities

The total academic evaluation system is divided into two forms:

1. Formative evaluation (done in the mid of the session)
2. Summative evaluation (done at the end of the session)

1. Formative evaluation

The characteristics of formative evaluation are as follows:

- It is to **evaluate and grade class work, homework, assignment and project work.**
- There is one term end examination for each term.
- The first term examination carries 20 marks.
- The second term examination carries 40 marks.

2. Summative evaluation

The characteristics of summative evaluation are as follows:

- It is based on term end examination.
- There will be two academic evaluations in the session.
- Final result will be the sum of both.

Application of scholastic and co-scholastic grading system

Grades in scholastic areas are decided on the basis of both formative as well as summative evaluation. It follows a grade point system prepared by experts as indicated in Table 5.2

Table 5.2 *Grading System for Scholastic Assessment*

Marks	Grade	Grade point
91—100	A1	10
81—90	A2	9
71—80	B1	8
61—70	B2	7
51—60	C1	6
41—50	C2	5
33—40	D	4
21—32	E1	—
20 & below	E2	—

In co-scholastic areas, A+, A, B+, B and C grades are given.

NOTES

Indicators for Grading: CBSE and State- Evolved Indicators

CBSE initiated this system of awarding grades in place of marks. No students will get compartment or will fail in the examination. Students will get result in two forms, namely 'Eligible for qualifying certificate (QUAL)' and 'Eligible for improvement of performance (EIOP)'. Even if a candidate fails in all subjects, he will now have five chances to improve their performance without having to repeat a year.

Criteria for scholastic evaluation

The criteria for scholastic evaluation may be discussed as under:

- External assessment of both theory and practical examinations is in numerical form. In case of internal assessment subjects, only grades are shown.
- In an internal examination, the assessment is done on a five-point scale, i.e., A, B, C, D and E.
- The grades are based on the scores got in the external examination. In case of subjects of internal assessment, the grades are awarded by the schools.
- In an external examination, the qualifying percentage is 33, separately in theory and practical.

Till high school, average percentage is provided. A percentage over 80 is considered excellent; candidates securing marks between 60–80 are awarded 'first division'; and candidates securing marks between 40–60 are awarded 'second division'.

Indicators of assessment of life skills

The life skill assessment includes thinking skills, social skills and emotional skills. The indicators for these three categories of skills are discussed as follows:

1. **Thinking skills:** Students demonstrate the ability to:
   - Be original, flexible and imaginative.
   - Raise questions, identify and analyse problems.
   - Implement a well thought out decision and take responsibility
   - Generate new ideas with fluency
   - Elaborate/build on new ideas
2. **Social skills:** Students demonstrate the ability to:
   - Identify, verbalize and respond effectively to others' emotions in an empathetic manner.
   - Get along well with others.
   - Take criticism positively
   - Listen actively
   - Communicate using appropriate words, intonation and body language

NOTES

3. **Emotional skills:** Students demonstrate the ability to:
   - Identify their own strengths and weaknesses
   - Be comfortable with their own self and overcome weaknesses for positive self-concept
   - Identify causes and effects of stress on themselves
   - Develop and use multi-faceted strategies to deal with stress
   - Express and respond to emotions with an awareness of the consequences

Note: *Refer to Appendix 1 for CBSE's Continuous and Comprehensive Evaluation (CCE) Manual for Teachers (Classes VI–VIII).*

5.4.1 Issues and Problems

There are many debatable issues involved in the assessment and evaluation of learners, such as marking vs grading issue, non-detention policy, and objectivity vs subjectivity issue.

Marking vs Grading System

Before we discuss the merits and demerits of marking and grading systems of assessment, we need to know what these two systems are. You have already learned about the characteristic features of the grading system. Here an overview of marking system is given. A marking scheme is a set of criteria used in assessing students' learning. Evaluating the students is not an easy task. It determines students' understanding of the subject. It is a powerful means to judge the quality of teaching–learning process. It provides continuous feedback of content and effectiveness of methodology used. All these give an idea of individual's growth.

Marking scheme depends on the type of tests. If the test is of objective type, the answer key (scoring key) is prepared well before time to avoid wastage of time. Teacher cannot be biased with any student. It does not need unnecessary hard work. The second type of test is the subjective type test. In this, questions may carry different values, such as 10- mark questions, 5-mark questions and 2- mark questions. The marks are given on some pre- decided criteria. After checking of all types of questions is done, then grand total of marks is determined, which determines the academic ability of the student.

Arguments for the grading system

The grading system is a good step to change the very old mark-based education system of our country. Our country has progressed and developed in many fields. Thus, the colonial educational system is no longer relevant in the changed circumstances. The grading system of evaluation is going to relax most of the students and parents. Under the marking system of evaluation, the students and guardians were used to go under unnecessary pressure and stress twice— in class X and class XII. No student would fail under the grading system, which was a characteristic

feature of the marking system. Those who fail under the marking system are depressed and the victims of inferiority complex. The marking system can also lead to the perception that high marks, rather than real knowledge, are important.

If knowledge enhancement and overall development of students is the primary goal, the grading system is preferable. Under this system, students of similar caliber are placed on an equal footing. This removes complexes and negativity among them. Education becomes a pleasant pastime and not just a chore. Eventually, both the students and the teachers emerge wiser.

Arguments for the marking system

Marks help students identify their exact scores. Under the marking system, even if two students are of similar calibre, one can ace out the other by a fraction of a point. This level of assessment encourages competitiveness. To score that one extra mark, students devote more time to study and pay attention to detail. What will be motivational factor if there will be no competition in education? Grading system is like killing the sense of competition and removing all motivation from the education system. The marking system tells the students where they actually stand.

Non-detention policy

The role of Universal Elementary Education (UEE) for strengthening the social fabric of democracy through provision of equal opportunities to all has been accepted since the inception of our Republic. The original Article 45 in the Directive Principles of State Policy in the Constitution mandated the State to endeavour to provide free and compulsory education to all children up to age fourteen in a period of ten years. The National Policy on Education (NPE), 1986 states: 'In our national perception, education is essentially for all… Education has an acculturating role. It refines sensitivities and perceptions that contribute to national cohesion, a scientific temper and independence of mind and spirit - thus furthering the goals of socialism, secularism and democracy enshrined in our Constitution'.

Over the years there has been significant spatial and numerical expansion of elementary schools in the country. Access and enrollment at the primary stage of education have reached near universal levels. The number of out-of-school children has reduced significantly. The gender gap in elementary education has narrowed and the percentage of children belonging to scheduled castes and tribes enrolled is proportionate to their population. Yet, the goal of universal elementary education continues to elude us. There remains an unfinished agenda of universal education at the upper primary stage. The number of children, particularly children from disadvantaged groups and weaker sections, who drop out of school before completing upper primary education, remains high. The quality of learning achievement is not always entirely satisfactory even in the case of children who complete elementary education. The gross enrolment ratio (GER) and Retention rates at elementary level remain low in the country.

NOTES

It is against this backdrop that the Government of India has been emphasizing on the policy of non-detention at the primary and elementary levels of education in the country. Currently, Sarva Shiksha Abhiyan (SSA) is implemented as India's main programme for universalizing elementary education. Its overall goals include universal access and retention, bridging of gender and social category gaps in education and enhancement of learning levels of children. Section 13 of the Right to Education (RTE) Act has banned any kind of screening procedure for children and parents at the time of admission, has barred detention or expulsion of a child, and even disallowed the conduct of Board examinations till a child completes elementary schooling (class VIII). This is to give the child adequate time to develop her learning and understanding fully through an enabling educational environment, and through a system of continuous and comprehensive assessment which enhances learning.

The government has taken several initiatives in pursuance of its no detention policy, which may be discussed under the following heads:

For children

- Stress on the holistic development of the child that is physical, social, mental, language, emotional etc.
- To provide support service at the centers wherever it is possible to encourage working girls of weaker sections.
- These programmes are prepared child-oriented, focused around play and the individuality of the child. Formal methods and introduction of the 3 R's will be discouraged at this stage. The local community will be fully involved in these programmes.
- Through ECCE child care and pre-primary education is catered for the purpose to strengthen human resources development in future. Health education is also promoted.
- Three main aspect of elementary education are:
  (i) Universal access and enrolment
  (ii) Universal retention of children up to 14 years of age
  (iii) A substantial improvement in the quality of education to enable all children to achieve essential levels of learning

Child- centered education

To motivate the child to attend the school a child- centred and activity-based process of learning is adopted in the schools at primary stage. Component of cognitive learning will be increased with various skill development as child grows.

The policy of non-detention at the primary stage will be retained, making evaluation negligible. Corporal punishment will be firmly excluded from the educational system and school timings as well as vacations also adjusted to the convenience of children.

NOTES

Facilities available in the school

- It is made sure that some basic facilities are essentially available in the school.
- Separate toilets for boys and girls.
- At least three large rooms
- Blackboard, maps, charts, toys, other necessary learning aids
- School library
- At least three teachers should work in every school, the number of teachers should increase with the increase in students.
- At least 50% teachers should be women.

Non- formal education

- This programme is specially meant for school dropouts, children who can not avail facility in their areas, working children and girls who cannot attend whole-day schools.
- In these centers provision for using modern technology and other aids are used to improve the learning environment. Talented and dedicated young men and women from the local community will be chosen to serve as instructors, training is also provided to them if necessary. All necessary action will be taken to ensure the quality of centres.
- Major steps will be taken to prepare curriculum that should be based on the needs of the learner and on the lines of the national core curriculum. It should also cater to the local environmental need.
- Learning material of high quality will be developed and provided free of charge to all pupils. A provision to provide participatory learning environment, and activities such as games and sports, cultural program, excursions, etc. is also made.
- Government, few voluntary agencies and Panchayati Raj will take care of these institutions.

Secondary education

- At this level of education students are exposed to various streams like science, the humanities and social sciences, history, civics to understand their constitutional duties .
- It is equally important to emphasize on enrolment of girls, SCs and STs, particularly in science, commerce and vocational streams. Concerned boards will be responsible to maintain the quality.
- Effort will be made to provide computer education also to prepare the student for a technological world.
- Provision for conducting aptitude test to find out children with special talent is also made. The centres should come forward in providing good quality education, irrespective of their capacity to pay for it.

NOTES

- Residential schools and Navodaya Vidyalayas are committed to fulfil this purpose and have been opened all over the country. These institutions are also committed to provide equal opportunity to children of backward area and SC, ST candidates.

Implementation of the non-detention policy

Table 5.3 summarizes the state-wise implementation of non-detention policy in primary classes in India.

Table 5.3 *State-Wise Implementation of Non-Detention Policy in Primary Classes*

State	Method of evaluation
Bihar	State follows non-detention policy. However, annual examinations are being conducted by every school for promoting students to next class
Goa	Follows continuous comprehensive evaluation and useful assessment procedure. No detention up to class III
Gujarat	Examination is taken in all classes of primary schools. In lower standard, promotion is on the basis of attendance/examination
Karnataka	Continuing with the system of annual examination and 80% attendance as criteria for promotion to next class
Mahasashtra	Annual examination system is in existence for class V onwards. Competency based tests are given for continuous evaluation of standards in class I-IV
Tamil Nadu	Promotion is based on annual examination performance
Uttar Praesh	There is a system of examination in all classes. Children are promoted based on attendance and examination
West Bengal	No formal examination system at primary level. From class V to VIII, there prevails an annual exam system
A & N Island	Students are promoted to next class on the result of annual exam and 75% attendance
Chandigarh	Promotion to next class is based on annual examination. Continuous evaluation is also been introduced in all classes
Dadra & Nagar Haveli	There is no examination system up to class IV. Students are promoted on the basis of attendance.
Pondicherry	Students in class I and II are promoted invariably to next class. But class III annual exam is taken for promotion.

Source: *Ministry of Human Resources Development*

NOTES

Table 5.4 summarizes the state-wise implementation of free distribution of textbooks in schools— an important component of the no-detention policy of the government.

Table 5.4 *Incentive Schemes: Free Distribution of Textbooks in Schools*

State	Status
Arunachal Pradesh	Free textbooks are given to APST girls up to class XII and to APST boys up to class VIII
Bihar	All girls and SC & ST boys are provided with free textbooks at primary level
Gujarat	Free textbooks to all students of primary schools run by local bodies
Karnataka	Free textbooks are distributed to all children in class I-IV. V-VII to SC/ST. Now been extended to all children class V-VII
Mahasashtra	Free textbooks are provided in backward blocks to children of backward communities
Manipur	No scheme, but girl student of class I-VIII are given some selected books in low female literacy districts
Tamil Nadu	Being provided
Tripura	Free textbooks are given to all ST students from class I-XII
Uttar Pradesh	Free textbooks are distributed to all SC/ST/OBC children and girls
West Bengal	Nationalized textbooks are issued free of cost to all children of primary schools
A & N Island	Free textbooks are provided to all students up to class V
Dadra & Nagar Haveli	Free textbooks are distributed to SC/ST and low income group students
Daman & Diu	Free textbooks are given to girl students and SC/ST students
Pondicherry	Free textbooks are supplied at the beginning of the academic year

Source: *Ministry of Human Resources Development*

Table 5.5 summarizes the status of the government's incentives for promoting education among SC, ST, OBC and minority children.

NOTES

Table 5.5 *Status of Incentives to SC/ST/OBC and Minority Communities*

State	Status
Arunachal Pradesh	Stipend in lieu of ration is given to the APST hostel boarders. Free textbooks are supplied to APST girls up to class XII and boys up to VIII. Free uniform are supplied to students up to VIII depending on availability of funds. Education is free up to XII
Bihar	SC/ST students enrolled at primary level are provided free text books
Goa	Free text books, uniform, rain coats, scholarship, opportunity cost for SC/ST/Cash Incentives
Karnataka	Free textbooks to all children I-IV. Free uniforms, supply of school bags etc. are given to SC/ST girls in class V-VIII in Government school. Started pre-matric hostels and pre-matric scholarships for students of classes V-X
Tamil Nadu	Cooked meal, free text books, dress, free transport, ₹ 500/- as incentive scholarship to SC/ST (III to VI), ₹ 1000/- VI onwards being given
Tripura	SC/ST girl students are provided dress grant and attendance scholarships. Pre-matric scholarships has been introduced for all SC/ST students from class VI-X
Uttar Pradesh	Students belonging to SC/ST/OBC are provided with free textbooks and all SC/ST/OBC/Handicaped/Miniority children are given scholarships at the beginning the session
West Bengal	Free uniform is provided to all students belonging to SC/ST categories
A & N Island	Incentives such as free books, note books, mid day meals, full attendance scholarships etc. are being provided to all Tribal students
Daman & Diu	Stipend, free uniform, free shoes, socks and free stationery are given to SC/ST students. Hostel and Ashram Shalas are run for SC/ST students
Pondicherry	Free hostel facility, food, clothing and accommodation to Sc students is provided. Training of cutting and tailoring to SC school drop outs with stipend of ₹ 100/- per month is given. Pre-matric scholarships is given to such students studying in VI to VIII, ₹ 150/- per annum and IX and C standard ₹ 200/- per annum. Free supply of text books, stationery, uniform cloth and reimbursement of the cost of medium textbooks to all SC students. Provision of tutorial facilities to Sc school students. Special assistance of ₹ 250/- to Sc girl students to undertake primary education from I to V. Grant of special incentive of ₹ 300/- is given to each Sc student who score 65% of marks and above in class X public examination.

Source: *Ministry of Human Resources Development*

Reporting Student Performance

A report card is that card which contains the results of different assessments, evaluation, and judgments held from time to time during the course study of a student.

A cumulative record is an account of the child's history in the school. It begins as soon as the child enters the school and continues till he leaves the school for good. It contains information regarding all aspects of the life of the child—his physical, mental, social, moral and psychological. It seeks to give as comprehensive a picture as possible of the personality of a child. 'Periodically the significant information gathered on students through the use of various techniques—tests, inventories, questionnaires, observation, interview, case study, case conferences and the like—should be assembled in summary form on a cumulative record,' writes Jane Warters.

Contents of students' report card

Report card usually contains the following sections:

1. Cumulative records of personal data
2. Home circumstances
3. Profile of health information
4. Psychological report
5. Scholastic progress report
6. Performance in co-curricular activities/Open house
7. Personality Characteristics
8. General Overall Remarks

Format of students' report card

The format of the report card is as follows:

Name of the Pupil.....

Name of the Father.....

Occupation of Parent/Guardian.....

Address of Parent/Guardian.....

Date of Birth.....

SCHOOLS ATTENDED

Name of School	Month and year of enrolment	Class in which enrolled	Month and year of leaving	Reasons for leaving

HOME CIRCUMSTANCES

1. Economic status of the family .....
   .....
2. Child's position in the family (whether only child, eldest or youngest child) .....
   .....

NOTES

3. Type of family (whether Joint or unitary) .....
.....
4. Education of the Parent/Guardian .....
.....
5. Special circumstances .....
.....

ATTENDANCE

Date	Number of school days	Number of days attended	Nature of long absence, if any

PHYSICAL DATA

Year	Height	Weight	Any serious illness or physical disability

REPORT OF THE MEDICAL OFFICER

1	
2	
3	

PSYCHOLOGICAL REPORT (IF ANY)

1	
2	
3	

NOTES

SCHOLASTIC PROGRESS REPORT

(Average percentage of marks or corresponding letter grade based on the result of periodical and annual tests).

Group	Subject	Class ___
		Marks/Grade
Language and Literature		
General Science and Science		
Social Studies and Social Sciences		
Any other Group		
Craft/Work Experience		

PERFORMANCE IN CO-CURRICULAR ACTIVITIES/OPEN HOUSE

(Average Grade of the Child)

Activities	Grade
1. Sports and Athletics	
2. Personal Hygiene	
3. Literary (debating etc.)	
4. Dramatics	
5. Community service (sanitation drive, literacy campaigns etc.)	
6. Any other activity	

Notable Proficiency and Skill in different fields or special merit obtained or position of responsibility held, if any

NOTES

Years and class	In Scholastic Achievement	In Bal-Sabha	In Scouts Troop	In Sports	In Social Service	Any other

PERSONALITY CHARACTERISTICS

Nature	Remarks/Grade
1. Honesty 2. Punctuality 3. Courtesy 4. Habits of work 5. Co-operativeness 6. Sociability 7. Self-confidence 8. Emotional stability 9. Leadership 10. Initiative	

GENERAL REMARKS

Class	Signature of the Class Teacher	Sign or Stamp of the Head

Main characteristics of students' report card

1. **Validity:** The facts recorded must be authentic. It is better not to write any information than to give untrue information.
2. **Comprehensiveness:** A report card should contain information as complete and comprehensive as possible.
3. **Accuracy:** The information in the card should be accurately recorded from the original source. If instead of 105 one writes 150 as score of an individual in English, one can imagine the difference in the educational guidance of that student.
4. **Objectivity:** Personal opinion and judgements should be reduced to the minimum.

5. **Economy of space:** The criterion of a good report card is that the maximum possible information should be included in the space available.
6. **Recency:** Any piece of information should be recorded as soon as it is obtained.
7. **Outcome of a co-operative enterprise:** It should be the result of group thinking.
8. **Evaluation:** Students should be evaluated often so that they can be guided properly.

Open house: An open house is a periodic event conducted by the school where the parents of students are invited. This event is mainly to facilitate free communication between teachers and parents. Teachers can give a holistic report of the progress of the child and a detailed feedback on the child's progress in the specific testing areas to the parents. Besides this, the parents can voice concerns, offer ideas, and provide feedback to the teachers regarding their child's weaknesses and areas that require special attention from the teacher.

5.5 COMPUTER IN EVALUATION

Over the last so many years, computers have entered into each and every place in our society. We see computer-dependence all around – in our homes, private workplaces, government departments, schools, colleges, hospitals, railway/air ticket booking outlets, banks, etc. So, it has become essential to equip a child with computer skills to enable him/her to use computer with ease in education and numerous day-to-day tasks. The computer enhances cognition, alertness, technical skills, imagination power, creativity, etc. of the student. It enhances learning environment and creates a positive competition among the students. The computer is so easy and interesting to learn that anyone can learn it at any age.

Advantages of computer in education

Some advantages of the computer are as follows:

- It adds a lot of ease to the teaching-learning process thus enhancing the achievements of the teacher and the student.
- It enables the students to think logically.
- Computer training helps students compete in the high-tech world confidently.
- As today no aspect of life is without interference of the computer, it has opened a lot many job avenues for the students.
- The use of the Internet through the computer connects the students with the students in their own country or other countries. This helps them understand the culture of each other and also discuss their academic problems.
- The computer-enabled Internet also helps the students to stay in touch with their teachers even after leaving the school, thus form the alumni community.

NOTES

- Teachers can use computer to send regular reports of the students to their parents, or even parents can check the status of their wards at the school website through the computer-enabled Internet.
- Educated ladies can start online tuitions with the help of the computer and the Internet.
- Learning to do so many things in the right way using the computer enhances the confidence level of the students.

Disadvantages of computers in education

The following issues arise from a heavy dependence of the students on the computer. The teachers and the students have to be on guard so as not to allow these issues pose insurmountable problems.

- What happens to a student who gets used to learning with the computer?
- Is he/she going to like or understand a normal class without the computer?
- How the student's interest in reading and studying through books can be kept alive that is necessary for conceptual and other development?
- How can the concentration necessary for various extra-curricular activities be increased?
- Will the computer-obsessive student not take any interest in extra-curricular activities?
- Does the computer damage traditional learning methods?
- How will cognitive skills be developed?
- How will it impact the eyes, finger muscles and shoulders of the students?
- What will happen to the social/physical development of the student if he/she is engaged in computer games instead of going out to play?
- Will not the computer affect the development of mathematical, analytical and critical thinking?

Some precautions for computer use

The working space presented by computers is absolutely well-defined. However, it is believed that real creativity does not happen in well-defined spaces but in ill-defined ones, such as those involving handicrafts, arts, humanities and social interaction. Some say that the development of logic-symbolic thinking forced by the computer in fact hinders real creativity in non-logical, formal symbolic areas.

The computer looks to be an excellent tool for learning, and it is a big attraction for the children of all age groups. However, the important thing is that we need to identify whether this attraction is for some academic reason or for some cosmetic reason. Children are fascinated by video games. Fascinating pictures, music, multimedia, video games, etc are lots of computer-related activities that can divert the attention of the students. Therefore, the teachers and the parents need control the computer use by the children because the computer is a machine for the use of humans, it should not rule humans.

NOTES

The teacher should present his/her own content and use the computer in a right mix. He/she must know how much use of the computer should be appropriate for his/her students as per their age. The teacher must use computer less than his/her own content.

Educational software does not have the ability of 'knowing' what a student has learned the week or even the year before, what has been happening around or in the world, etc. Whatever knowledge it has, is all fed by humans. So, it is the teacher who can assess the impact of the computer-aided learning on the student.

No, doubt, the computer accelerates the children's development. By forcing a virtual setting, a formal language and a logic-symbolic thinking, the computer forces the children, particularly teenagers, to physically and mentally behave like adults. It is not good for a child to sit on a chair for long hours while using the computer. Besides, the computer also hinders a child's imagination. If the child does not use his/her imagination, or fantasize about some higher goal or happy moments (this would happen while hearing a fairy tale, for instance), it adversely affects his/her thinking ability. As with TV, educational software full of images leaves no space for the child's imagination.

Excessive use of computer may distract the child not only from the studies but also from his own people. But, social and family interaction is equally important. Social sites may help the child to associate with friends and other people. But is this social interaction equally good as compared to the social interaction that the child has while playing with the children in the neighbourhood? Do we want our children to be initiated into social interactions by a machine or through their real-life interactions with people around them?

5.5.1 Multimedia in Education

Multimedia means organization of educational or other content using a combination of different content forms such as text, audio, still images, animation, video, or interactivity content forms. The term multimedia can be used as a noun (a medium with multiple content forms) or as an adjective describing a medium which has multiple content forms. The term is used in contrast to such media that use only rudimentary computer display such as text-only or traditional forms of printed or hand-produced material.

In education, multimedia is used to produce computer-based training courses (popularly called CBTs) and reference books like encyclopaedia and almanacs. A CBT lets the user go through a series of presentations of text about a particular topic and associated illustrations in various information formats. Edutainment is an informal term used to describe combining education with entertainment, especially multimedia entertainment.

In the past decade, teaching and learning practices have expanded dramatically because of the introduction of multimedia. Several lines of research have evolved (e.g. cognitive load and multimedia learning).

NOTES

Categories of multimedia

The two main types of multimedia in use are as follows:

1. **Linear Multimedia:** Linear active content progresses without any navigation control by the viewer. Linear multimedia is similar to cinema presentation.
2. **Non-linear Multimedia:** Non-linear content offers the user interactivity to control its progress, Examples of non-linear multimedia are computer games or self-paced computer based training. Non-linear content is also known as hypermedia content.

Multimedia presentations

Multimedia presentations can be recorded or live. A recorded presentation may allow interactivity via a navigation system. A live multimedia presentation may allow interactivity via interaction with the presenter.

Benefits of Multimedia

Some important benefits of multimedia are as follows:

- Multimedia applications provide students opportunities to represent and express their prior knowledge.
- They allow students to function as designers, using tools for accessing and interpreting information, organizing their personal knowledge, and representing what they know to others.
- Multimedia applications engage students and provide them valuable learning opportunities.
- They empower students to create and design rather than 'absorbing representations created by others.'
- They encourage deep reflective thinking.
- Multimedia can be used in a wide variety of areas such as entertainment, training in various areas such as military or industrial training.
- Multimedia teaching brings students into a class where they can interact with the teacher and the subject.
- Multimedia teaching is more intuitive than old ways; the teachers can simulate situations in real life.
- In many circumstances, the teachers don't have to be there. The students will learn by themselves in the class.
- Multimedia provides the teachers more approaches to stimulate the students' passion for learning.

Demerits of multimedia

Some demerits of multimedia are as follows:

- Multimedia requires high-end computer systems. Sound, images, animation and video constitute a large amount of data in multimedia, which slows down or even may not fit into a low-end computer.
- It can be used only on good quality computers.

- It may not be accessible to a large section of intended users if they do not have access to multimedia-capable machines.
- It is highly costly and time consuming.

NOTES

Internet in education

The Internet is a network millions of private, public, academic, business, and government networks of computers, of local to global scope, that are linked by a broad array of electronic, wireless and optical networking technologies. The computer networks on the Internet use a standard internet protocol suit (IP address) (often called TCP/IP, although not all applications use TCP) to serve billions of users worldwide. Just as the postal system enables sending of mails to various addresses, a computer's IP address gives the Internet routing protocols the unique information they need to route packets of information to anybody's desktop from anywhere across the internet.

If a machine needs to contact another by a domain name, it first looks up the corresponding IP address with the domain name service. The IP address is the geographical descriptor of the virtual world, and the addresses of both the source and destination systems are stored in the header of every packet that flows across the Internet.

The IP address of a Windows computer can be found by opening an MSDOS or command window and typing one of these commands — 'winipcfg' or 'ipconfig'.

Benefits of Internet in education

Some benefits of the Internet in the field of education are as follows:

- Anytime anywhere availability of the Internet; access to people from all over the world.
- Access to interactive and dynamic material available on the Internet.
- A million new web pages/day can be visited as per need.
- More than 200 million people are connected with e-mail.
- Publication work can be done with the help of the Internet.
- Information can be sent in seconds to as many people as we wish.
- It is full of freedom and gives rise to independent learning.
- It provides a solid platform for group and/or real-world projects.
- Students can communicate with the students in their own and other countries. They can gain from others' knowledge and experiences, participate in chat rooms, share ideas and solutions and learn about the many diverse cultures out there.
- The Internet not only helps the students but also gives benefits to the parents and teachers. The interactive learning that the Internet provides can help the students and their parents with little or no English skills to learn English. Parents can become more involved in their children's education by connecting the school with homes, libraries or other access ports.

NOTES

- The teachers can adjust to the different learning styles in the classroom. They can also set their own pace of teaching.
- Individual teaching techniques can become easily available which has proved to be a factor in student achievement.
- Teachers can teach at more than one place simultaneously. They may be in a small town but through the Internet, they can be linked to the students in cities or metros.
- Also, the Internet enables administrators and teachers to spend less time on administration and recordkeeping. This would also give them more time to spend with their students.

CHECK YOUR PROGRESS

6. In the grading system, how are grades in scholastic areas decided?
7. What is a report card?
8. What is multimedia?

5.6 SUMMARY

- The reliability of a test refers to the degree to which the test results obtained are free from error of measurement or chance errors.
- Reliability is necessary but not a sufficient condition for validity.
- When examining the reliability coefficient of standardized tests, it is important to consider the methods used to obtain the reliability estimates.
- American Psychological Association (APA) introduced several methods of estimating reliability. These are:
  (i) Test-retest method
  (ii) Equivalent forms method
  (iii) Split-half method
  (iv) Kuder-Richardson method
  (v) Inter-rater method
- The validity of a test is determined by measuring the extent to which it matches with a given criterion. It refers to the very important purpose of a test, and it is the most important characteristic of a good test.
- A test may have other merits, but if it lacks validity, it is valueless.
- The six types of validity are discussed below.
  (i) Face validity
  (ii) Content validity
  (iii) Concurrent validity
  (iv) Construct validity

(v) Predictive validity
(vi) Criterion validity

- The methods used for assessing the validity of a test are:
  - o Correlation method
  - o Cross validation
- Reliability and validity are closely related, even though they cannot be interchanged. An assessment that has very low reliability will also have low validity; quite obviously, a measurement that has low levels of accuracy or consistency is not likely to suitably fulfil its objective.
- At the same time, the factors necessary for achieving a considerably high degree of reliability can affect validity negatively.
- Grades in the realm of education are standardized measurements of varying levels of comprehension within a subject area.
- It has become essential to equip a child with computer skills to enable him/her to use computer with ease in education and numerous day-to-day tasks.
- The computer enhances cognition, alertness, technical skills, imagination power, creativity, etc. of the student. It enhances learning environment and creates a positive competition among the students.

NOTES

5.7 KEY TERMS

- **Validity:** It is the extent to which a concept, conclusion or measurement is well-founded and corresponds accurately to the real world.
- **Ambiguity:** It refers to the quality or state of being ambiguous especially in meaning.
- **Grade Point Average:** It is a number representing the average value of the accumulated final grades earned in courses over time.
- **Open House:** An open house is a periodic event conducted by the school where the parents of students are invited. The event is mainly to facilitate free communication between teachers and parents.

5.8 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Reliability means the extent to which a measuring device yields consistent results upon testing and retesting.
2. The best time interval between assessment administrations depends largely on the use to be made of the results.
3. The validity of a test is determined by measuring the extent to which it matches with a given criterion.
4. Concurrent validity indicates to which extent the test scores correspond to already accepted measures of performance (or status made at the same time).

NOTES

5. Correlation method is used in most of the cases for calculating the validity of a test.
6. Grades in scholastic areas are decided on the basis of both formative as well as summative evaluation.
7. A report card is that card which contains the results of different assessments, evaluation, and judgements held from time to time during the course study of a student.
8. Multimedia means organization of education or other content using a combination of different content forms such as text, audio, still images, animation, video or interactivity content forms.

5.9 QUESTIONS AND EXERCISES

Short-Answer Questions

1. What are the characteristics of reliability?
2. What are the factors upon which the reliability of a test is dependent?
3. What is the test-retest method? What are its limitations?
4. What is the concept of grading? What are the merits of the grading system?
5. List the criteria for scholastic evaluation.
6. How do states implement the non-detention policy in primary classes in India?
7. What are the two main types of multimedia in use?

Long-Answer Questions

1. Explain the various types of validity.
2. Discuss the different methods of determining reliability coefficient.
3. Describe the different methods of determining validity.
4. Discuss the relation between reliability and validity.

5.10 FURTHER READING

Bloom, B. S. *Taxonomy of Objectives of Cognitive Domains*. New York: Longmans Green &Co.

Ebel, R.L. *Essentials of Educational Measurement*. New Jersey: Englewood Cliff.

Grunlund, N. E. *Measurement and Evaluation in Teaching* (3rd Ed). New York: Mcmillan Publishing Co.

Julian, C., Stanley and Hopkins. *Educational and Psychological Measurement and Evaluation*. New Delhi: Prentice Hall of India Pvt. Ltd.

Lindermon, Richard, H. *Educational Measurement*. Mumbai: B. S. Taraporevala Sons & Co.