Educational Measurement and Evaluation
DEDU504

LOVELY PROFESSIONAL UNIVERSITY

EDUCATIONAL MEASUREMENT AND EVALUATION

SYLLABUS

Educational Measurement and Evaluation

Objectives :

To enable the learners to

(1) familiarise with the utility of measurement evaluation

(2) familiarise with different kinds of psychological tests

(3) understand the steps for the construction of an achievement test and the characteristics of good test and type of test i.e. standardized and non-standardised tests

Sr. No.	Description
1.	Measurement and evaluation: concept, need, scope; difference and relevance. Educational testing and assessment: concept, context, issues and current trends. Scales of measurement: ordinal, nominal, interval, ratio.
2.	Characteristics of a good test. Planning for different types of test. Validity-- types and methods and usability. Reliability-- types and methods and usability.
3.	Test construction. Test standardization. Item analysis: Item difficulty, discrimination index. Effectiveness of distracters. Development of Norms of a test.
4.	Conversion of raw scores into standard scores, T-scores, C-scores, Z-scores, Stanine scores, percentiles. Interpretation of test scores: qualitative and quantitative.
5.	Criterion referenced test, Norm reference test, Factors influencing test scores: nature of test, psychological factors and environmental factors.
6.	Integrated approach of evaluation. Marking system: need, problems and components. Grading—need, problems, components and methods. Methods of feedback for students
7.	Semester system vs annual system, Continuous assessment, Portfolio assessment Question bank, Use of computer in evaluation.
8.	Achievement test: concept, types and construction. Diagnostic test: concept and construction, remedial teaching. Objective type test: advantages and limitations
9.	Short answer type test: advantages and limitations. Essay type test: advantages and limitations
10.	Formative and Summative Evaluation. Measurement of Attitude, Aptitude, personality and intelligence

CONTENTS

Unit 1: Educational Testing and Assessment : Concept, Context, Issues and Current Trends

---

CONTENTS

Objectives

Introduction

Objectives

Objectives of this unit can be summarised as below :

- To explain the Concept of Educational Testing
- To discuss about the Context of Educational Testing
- To describe the issues of Educational Testing
- To discuss about current trends in Educational Testing
- To explain about the Concept of Educational Assessment
- To discuss about the Context of Educational Assessment
- To describe about issues of Educational Assessment
- To discuss about current trends in Educational Assessment

Introduction

Almost all the Commissions on education as also the National Policy on Education (1986), Programme of Action (1992) and National Curriculum Framework (2000) of NCERT have all stressed the need for a more continuous and comprehensive evaluation of students in order to pass more sound judgments about students' learning and growth. Regular Testing assessment through pupils' lives in school is considered essential for remedial treatment of those who fall below the acceptable performance level. It is in this context that the role of Continuous and Comprehensive Evaluation (C.C.E.) is highlighted in appraising the whole child and recognising it as a complement of external examination by providing due weight age to C.C.E. in the final assessment of the total development of the child.

1.1   Concept of Educational Testing

Testing is neither assessment nor appraisal, but at the same time it may become a means to getting information, data or evidences needed for assessment and appraisal. Testing is one of the significant and most usable technique in any system of examination or evaluation. It envisages the use of instruments or tools for gathering information or data. In written examinations, question paper is one of the most potent tools employed for collecting and obtaining information about pupils' achievement.

1.2   Context of Educational Testing

A test of educational achievement is one designed to measure knowledge, understanding, or skills in a specified subject or group of subjects. The test might be restricted to a single subject, such as arithmetic, yielding a separate score for each subject and a total score for the several subjects combined.

Tests of educational achievement differ from those of intelligence in that the former are concerned with the quantity and quality of learning attained in a subject of study, or group to subjects, after a period of instruction and the latter are general in scope and are intended tor the measurement and analysis of psychological *processes*, although they must of necessity employ some acquired content that resembles the content found in achievement tests.

1.2.1 Nature and Scope

Most educational achievement tests are devoted largely to the measurement of the amount of information acquired or the skills and techniques developed. In recent years, however, an increasing number have been devised to measure such educational results as attitudes and appreciations, ability to solve problems, to draw inferences from subject matter, to apply generalizations to specific situations and problems.

Emphasis upon these more recent types of test materials is clearly illustrated in a publication in which educational objectives are analyzed in detail. There are six major classes of objectives. These are knowledge, comprehension, application, analysis, synthesis, and evaluation. Each of these is analyzed into several aspects. For example, knowledge is divided into knowledge of specifics, ways and means of dealing with specifics, universals and abstractions in a field. These, in turn, are broken down further into knowledge of :

specific information

terminology

ways and means of presenting ideas and phenomena

trends and sequences

classification and categories

criteria

methodology

major ideas

principles and generalizations

theories and structures

1.2.2 Testing Techniques

Written Examinations

In the process of evaluation of cognitive outcomes of learning, the most common and reliable technique is that of testing. For collection of data both testing and non-testing techniques (observation, inquiry, analysis) are usable. With each technique a number or tools are associated, which help

gather the data or evidences required. In testing technique we may use written examinations, oral testing and practical examination corresponding to the assessment of cognitive or psychomotor outcomes of learning. Written tests are more reliable for measuring the learning outcomes in cognitive domain, whereas practical examinations are more dependable for measuring the learning outcomes in psychomotor domain. Likewise personality tests, interest inventories and attitude scale are more relevant for measuring the learning outcomes in affective domain. Written tests are still most acceptable due to better reliability than oral and practical tests for judging the students' achievement, in spite of some limitations.

Oral Examinations

Need for development of oral skills and expressions, which are necessary in day-to-day living, was stressed as back as in 1964 in the fifth Conference of Chairman and Secretaries of the Boards of Secondary Education. But unless oral skills are tested in the external examinations or certified in school-based assessment, these are not going to attract the needed attention of the teachers in developing these skills. Subjectivity in assesment, greater number of examiners, time span, inter-examiner variance, subjective interpretation etc. are a few among many more problems and difficulties in using oral tests in external examinations. However, their use in instructional process would continue to provide the diagnosis, feedback and their use as instructional tool for readiness testing and review of lessons.

Purpose of oral examination is :

(a) to test oral skills that cannot be tested through written examinations.

(b) to confirm and probe further evidences gathered through written examination whenever desired (viva);

(c) to judge the extent to which such skills are warranted by the nature of subject; and

(d) to make quick oral review for informal assessment of what the pupils have learnt or their deficiencies.

For improving validity and reliability of oral examinations, and systematic evaluation in the field of languages it is necessary to :

(*i*) identify clearly listening, speaking, reading and written expression;

(*ii*) delineate the aspects of evaluation of language and content elements in case of oral expression;

(*iii*) define oral expression in terms of expected competencies like use of appropriate words in right context using correct sentence pattern, pronouncing words correctly, modulating voice according to the situation, speaking at a reasonable speed, arrange idea in logical sequence, using proper gestures, avoiding mannerism etc.;

(*iv*) develop criteria of assessment;

(*v*) define each criterion;

(*vi*) prepare a scale for each component;

(*vii*) develop marking scheme;

(*viii*) decide about the mode of conducting the examination;

(*ix*) devise questions and exercises;

(*x*) interpret responses; and

(*xi*) provide feedback of results for remediation.

Oral examination can profitably be used to validate content, oral skills, work experience, project work, functional understanding about practical skills etc. when used as an external moderation device. For younger children it is an indispensable device for testing to get on-the-spot feedback about their learning and provide basis for remediation.

'Tests can be and should be, among the most useful instructional tools for planning new learning activities and for monitoring student's progress in attaining the learning goals presented to them. Tests can be used to provide recognition and rewards for success in learning and teaching. They can be used to motivate and direct effort to learn. In short, they can be used to contribute substantially to effective instruction."

1.2.3 Classification of Tests

Tests have been classified in a number of ways. Some of these even overlap. Here we are stating only one classification which seems to be quite comprehensive.

Tests vary according to form, use and type etc. Yoak and Simpson give the following classification of tests :

Form

(a) Oral examinations (b) Written examination

Purposes

(a) Prognostic. (b) Diagnostic. (c) Power. (d) Speed. (e) Accuracy (f) Quality. (g) Range.

Organisation

(a) Essay. (b) Objective.

Period or Time of Administering

(a) Daily (b) Weekly (c) Monthly (d) Term. (e) Year.

Duration

(a) Short. (b) Long.

Method of Scoring and Interpreting Results

(a) Non-standardized. (b) Standardized.

Abilities Involved

(a) Speed. (b) Comprehension. (c) Organization. (d) Judgement. (e) Retention. (f) Appreciation etc.

Nature of Material Included

(a) Arithmetic. (b) Language. (c) Reading. (d) Spelling. (e) Writing etc.

Mental Functions Involved

(a) Association. (b) Memory. (c) Recall. (d) Recognition. (e) Problem-solving.

Types of Response Involved

(a)   Alternate response : (1) True - False (2) Yes - No (3) Plus - Minus.
(b)   Multiple response : (1) Best Answer. (2) Correct Answer.
(c)   Completion.
(d)   Matching.
(e)   Identification
(f)   Enumeration.
(g)   Eassy.

1.3 Issues of Educational Testing

The traditional question paper is associated with the essay type questions. Testing of this type suffers from the following limitations :

(*i*)   It consists of a very limited number of questions from 6 to 10 questions.

(*ii*) Questions do not cover the major portion of the syllabus.

(*iii*) Chance element predominates.

(*iv*) Questions mostly encourage memorisation and cramming.

(*v*) Options are very often provided. Students are asked to attempt any five or six questions out of ten questions (usually this is the number).

(*vi*) Directional words like 'State', 'Describe', 'Explain', 'What do you know' and 'Interpret' make the question paper vague. Answers vary to a considerate extent.

(*vii*) Students depend upon 'guess work'.

(*viii*) Marking becomes very subjective.

(*ix*) Standardization of marking becomes very difficult.

Limitations

- **Lowering of Moral Standard :** Examinations teach the students different ways of becoming dishonest. Books are smuggled in the examination centres and attempts are made on the life of the supervisors who try to catch the students using unfair means.

- **Ignoring Qualities of Character :** They fail to provide any measure to test the originality, initiative, truthfulness, honesty, sociability of an individual and thus they fail to test real education. "Training in originality and independence of thought, correctness of judgement or reasoning, responsiveness to noble ideas and sentiments and enjoyment of beautiful things - these cease to be the aims of teacher's effort for they are not judged by the traditional examination. Examination is the sole aim and cram the sole method.

- **Subjectivity :** Subjective attitude of examiners influence the marks of individuals and leads to a great variability in marking. Vernon points out, "The same script might receives a different mark if read after instead of before dinner."

- **Heavy Mental Strain :** Most of the students are in the habit of working strenuously just near the annual examination and this severely tells upon their health, leads to mental dyspepsia and discourages the formation of healthy mental habits.

- **Develop Frustration :** Failures in examinations lead to frustration and even to suicides in some cases.

- Heavy weightage to verbal ability in many tests.

- Difficulty in obtaining level norms.

- Relative narrowness of the traits measured by tests.

- Indifference, lack of attention, low commitment and generally low motivation of the students, tending to invalidate the test results.

- Difficulty to follow the directions of the tester by the testee.

- Undue influence of socio-cultural conditions or environmental influences on the testees or the students.

- Low validity and reliability in the case of some pupils.

As a matter of fact, the examinations have met with heavy criticism from all quarters. This criticism is due to the excessive domination of the examination system and a large number of students subjected to it. Candidates complain against the examination system because of its harmful influence on school work; the parents denounce it because of its mental strain; the teachers against the examination because of its injurious effect on the physical and mental health of the children; the practical psychologists speak ill of it because of its unreliability and invalidity and the educational theorists attack it because of its lack of definiteness in aim and purpose.

1.4 Current Trends in Testing

There are lot of current trends are working in Testing. These are as follows:-

Introduction of New Types of Tests : Attempts should be made to minimise the subjective element. Essay type questions should be reduced and supplemented by new-type of objective.

Thought-provoking Questions : Questions should be thought-provoking and evenlydistributed over the entire course. Question should be such as to discourage cramming.

Class Work : Due consideration should be given to the regularity of the students in class work.

Appointment of Examiners : The paper-setters and examiners for external examinations should be drawn from the teachers who actually teach subjects in schools.

Viva Voce Tests : External examinations may be supplemented by viva voce tests, it possible.

Standard of Markings : Standard of marking should be prescribed so as to minimise the variability in marking.

Balanced Questions : Difficult as well as easy questions should find a place in the question papers. These should not be either too difficult or too easy.

Monthly Tests : Instead of terminal examinations, a system of monthly tests should exist. However, a recapitulatory test may be held at the end of the year.

Faith in the Teacher : The teacher should be trusted. He should be given a fair opportunity to know and study closely the students he teaches.

Cumulative Records : These should be maintained in respect of all the students.

Grades : In place of numerical marking, the system of symbolic marking be adopted. Five-point scale seems to be the best type of symbolic marking scale i.e., 'A' stands for excellent, 'B' for good, 'C' for fair and average, 'D' for poor and 'E' for very poor. **Examination as Means :** Examinations should be regarded as means and not ends. These should be conducted in such a manner that they become effective instruments of education. They should serve as stimulus to learn new facts, gain new experiences, discover weak points and estimate progress.

Oral Examinations : Oral examinations should be held in addition to the written ones. Some qualities like candidate's alertness, intelligence, special interests, mental outlook, his personal qualities of mind and character and his mastery of acquired knowledge can be better tested by viva voce tests than by written examinations.

Self Assessment

1. *Fill in the blanks :*

(*i*)   Testing is one of the significant and most usable techniques in any _____ or evaluation.

(*ii*)   _____ are more reliable for measuring the learning outcomes in Cognitive domain, while practical examinations measure learning out comes in _____.

(*iii*)   _____ of test provide indication of success or failures of educational programmes.

(*iv*)   Questions should be _____ and evenly distributed over the entire course.

(*v*)   _____ of examiners influence the marks of individuals and leads to a great variability in marking.

1.5 Concept of Educational Assessment

Assessment is an ongoing process aimed at understanding and improving student learning. It involves making expectations explicit and public; setting appropriate criteria and high standards for learning quality; systematically gathering, analyzing, and interpreting evidence to determine how well performance matches those expectations and standards, and using the resulting information to document, explain, and improve performance.

- Ongoing process

- To understand and improve student learning
- Explicit and public expectations
- Appropriate criteria and high standards
- Systematic data collection, analysis, and interpretation to :
  - determine whether learning meets expectations and standards
  - document, explain, and/or improve performance

1.6 Context of Educational Assessment

1.6.1 Objectives of Educational Assessment

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

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

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

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

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

1.6.2 Aspects of Assessment

Comprehensiveness is the significant factor in the assessment of the whole chi1d vis-a-vis his total development to form the basis of assessment. Accordingly, the following areas of assessment will be included in the CCE.

1.6.3 Assessment of Scholastic Aspects

Use of the following tools will be made in assessment of scholastic aspects.

1. **Tests :** Different types of tests required for asessing the scholastic aspects are given below :

   (a) Readiness tests

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

   (b) Criterion-referenced unit-tests

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

1.6.4 Assessment of Non-Scholastic Aspects

This would include assessment of the following aspects which may be introduced in second phase.

1. Physical health, covering basic understanding about nutrition and health, physical fitness, development of positive attitudes etc.

2.   Habits like health habits, study habits and work habits.

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

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

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

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

Did u know? Different types of co-curricular activities, projects and practical work, which form a part of school programmes like games, sports, gardening, dancing, music, album making etc., involve, the development of different types of psychomotor outcomes of learning.

1.6.5 Modes of Assessment

Three modes of assessment or evaluation are visualised.

1.   **Rational mode :** It will be used at planning stage for planning any programme or activity using participative methodology to get experts' opinion for arriving at consensus.

2.   **Formative mode :** It will be used at the implementation stage to identify inadequacies and weaknesses in the programmes, strategies and development of materials.

3.   **Summative mode :** It will be used at the review stage to judge the efficacy and effectiveness of the materials, strategies, programmes, outputs and collateral effects.

Evaluation at the first stage will be *logical*; at the second stage *diagnostic*, and at the third stage *judgmental* in nature. Nevertheless, all the three modes are not considered exclusive but are interrelated and interdependent.

1.6.6 Tools and Techniques of Assessment

All the four basic techniques of assessment, i.e. testing, observation, inquiry and analysis, will be employed in different situations to collect the relevant evidences about pupils' growth and development in both cognitive and non-cognitive areas.

Notes   It is essential to undertake functional research at various stages. Formative evaluation will be in-built in the CCE project to get regular feedback.

1.7 Issues of Present System of Assessment

(a)   Unless protagonists of reform develop a deep conviction and full faith (not 20% or 25%) in teachers' assessments and conceptualise dichotomy observable in theory and practice, establishment of a pedagogicaly sound system of SBA would be difficult.

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

(c)   Unless within a given state, board of school education or the concerned examining agency where curriculum requirements are the same for a particular stage of school education, a

common blueprint of CCE based on the same design of SBA is adopted in all schools under the same board, comparability of attainment targets and pupils' achievement across the schools would not be possible, which in turn would pass it difficult to pass valid norm-referenced judgments.

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

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

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

Task   What is Readiness test ?

1.8 Current Trends in Educational Assessment

1.   **Visualise :** the context and background of the institution

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

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

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

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

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

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

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

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

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

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

12.   **Analyse :** evidences in terms of specified intended learning outcomes.

13.   **Interpret :** the evidences in terms of assessment objectives.

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

15. **Summarise :** results and other evidences into meaningful individual and group profiles for different areas and aspects.

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

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

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

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

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

Self Assessment

2. State whether the following statements are 'true' or 'false'.

(*i*)   Educational assessment is an ongoing process aimed at understanding and improving student learning.

(*ii*)   Comprehensiveness is the significant factor in the assessment of the whole child.

(*iii*)   Unit test should be used at the beginning of each unit.

(*iv*)   The rational mode of assessment are used at the implementation stage to identify inadequacies and weaknesses in the programmes.

1.9 Summary

- Testing is one of the significant and most usable techniques in any system of examination or evaluation. It envisages the use of instruments or tools for gathering information or data.

- A test of educational achievement is one designed to measure knowledge, understanding, or skills in a specified subject or group of subjects.

- Tests of educational achievement differ from those of intelligence in that (1) the former are concerned with the quantity and quality of learning attained in a subject of study, or group to subjects, after a period of instruction and (2) the latter are general in scope and are intended for the measurement and analysis of psychological *processes.*

- Most educational achievement tests are devoted largely to the measurement of the amount of information acquired or the skills and techniques developed.

- The principles upon which tests of educational achievement are standardized are the same as those of the other types already presented; the same principles of definition of aim, sampling, validity and reliability apply here as elsewhere. A standardized test of educational achievement should be based upon a careful analysis of materials taught in a given field;

- Testing Techniques

- Written Examinations

- In the process of evaluation of cognitive outcomes of learning, the most common and reliable technique is that of testing. For collection of data both testing and non-testing techniques (observation, inquiry, analysis) are usable. With each technique a number or tools are associated, which help gather the data or evidences required.

- Oral Examinations

- Need for development of oral skills and expressions, which are necessary in day-to-day living, was stressed as back as in 1964 in the fifth Conference of Chairman and Secretaries of the Boards of Secondary Education.

- Purpose of oral examination is :

  (a) to test oral skills that cannot be tested through written examinations.

  (b) to confirm and probe further evidences gathered through written examination whenever desired (viva);

  (c) to judge the extent to which such skills are warranted by the nature of subject; and

  (d) to make quick oral review for informal assessment of what the pupils have learnt or their deficiencies.

- Classification of Tests : Tests have been classified in a number of ways. Some of these even overlap.

- Form

  (a) Oral examinations   (b) Written examination

- Purposes

  (a) Prognostic. (b) Diagnostic. (c) Power. (d) Speed. (e) Accuracy (f) Quality. (g) Range.

- Organisation

  (a) Essay. (b) Objective.

- Period or Time of Administering

  (a) Daily   (b) Weekly   (c) Monthly   (d) Term   (e) Year.

- Duration

  (a) Short.   (b) Long.

- Method of Scoring and Interpreting Results

  (a) Non-standardized. (b) Standardized.

- Abilities Involved

  (a) Speed. (b) Comprehension. (c) Organization. (d) Judgement. (e) Retention.

  (f) Appreciation etc.

- Types of Response Involved

  (a) Alternate response : (1) True - False (2) Yes - No (3) Plus - Minus.

  (b) Multiple response : (1) Best Answer. (2) Correct Answer.

  (c) Completion.

  (d) Matching.

  (e) Identification

  (f) Enumeration.

  (g) Eassy.

Notes
- System at Work
- Centralised Responsibility : There should be one person in each school who will look after the administrative details given below :

  (a) Ordering for relevant tests and test material.

  (b) Arranging schedules of time.

  (c) Arranging place for testing.

  (d) Arranging scoring and recording results.

- Testing Schedule : Tests should be given under standard conditions that permit each testee to perform at his best level.Testing schedule should never be made too tight.
- Preparation by Testers : The testers should prepare themselves in advance especially for the following aspects of the testing programme.

  (a) The tester should follow the manual verbatim.

  (b) The tester should make the procedures and the tasks clear to the testee before administering the test.

  (c) The tester should have a stopwatch and he should allow the subjects to work on a particular test for that much time only which is mentioned in its manual.

- Appropriate Environment for Testing : The desirable environment for testing should satisfy the following conditions :

  (*i*) Examinees are physically comfortable.

  (*ii*) Examinees are emotionally relaxed.

  (*iii*) Examinees are free from interruptions and distractions.

  (*iv*) Examinees are seated in such a way as to prevent copying.

- Recording of Test Results : Results should be kept systematically so that these are made use of conveniently by the appropriate agency. Of course results should be kept confidential.
- General Purposes of Testing

  (*i*) To determine the achievement level and progress of the pupils.

  (*ii*) To obtain data for diagnostic purposes.

  (*iii*) To identify the interests of the students.

  (*iv*) To identify the underachievers.

- The traditional question paper is associated with the essay type questions. Testing of this type suffers from the following limitations :

  (*i*) It consists of a very limited number of questions from 6 to 10 questions.

  (*ii*) Questions do not cover the major portion of the syllabus.

  (*iii*) Chance element predominates.

  (*iv*) Questions mostly encourage memorisation and cramming.

- Limitations

  (*i*) Heavy weightage to verbal ability in many tests.

  (*ii*) Difficulty in obtaining level norms.

  (*iii*) Relative narrowness of the traits measured by tests.

  (*iv*) Indifference, lack of attention, low commitment and generally low motivation of the students, tending to invalidate the test results.

- Lowering of Moral Standard : Examinations teach the students different ways of becoming dishonest. Books are smuggled in the examination centres and attempts are made on the life of the supervisors who try to catch the students using unfair means.

- Ignoring Qualities of Character : They fail to provide any measure to test the originality, initiative, truthfulness, honesty, sociability of an individual and thus they fail to test real education.

- Subjectivity : Subjective attitude of examiners influence the marks of individuals and leads to a great variability in marking.

- Heavy Mental Strain : Most of the students are in the habit of working strenuously just near the annual examination and this severely tells upon their health, leads to mental dyspepsia and discourages the formation of healthy mental habits.

- Develop Frustration : Failures in examinations lead to frustration and even to suicides in some cases.

- Introduction of New Types of Tests : Attempts should be made to minimise the subjective element. Essay type of questions should be reduced and supplemented by new-type of objective.

- Thought-provoking Questions

- Class Work

- Appointment of Examiners

- Viva Voce Tests

- Standard of Markings

- Balanced Questions

- Monthly Tests

- Faith in the Teacher

- Cumulative Records

- Grades : In place of numerical marking, the system of symbolic marking be adopted. Five-point scale seems to be the best type of symbolic marking scale i.e., 'A' stands for excellent, 'B' for good, 'C' for fair and average, 'D' for poor and 'E' for very poor.

- Assessment is an ongoing process aimed at understanding and improving student learning. It involves making expectations explicit and public; setting appropriate criteria and high standards for learning quality; systematically gathering, analyzing, and interpreting evidence to determine how well performance matches those expectations and standards, and using the resulting information to document, explain, and improve performance.

- Objective of SBA

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

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

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

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

- Aspects of Assessment

    - Comprehensiveness is the significant factor in the assessment of the whole child vis-a-vis his total development to form the basis of assessment. Accordingly, the following areas of assessment will be included in the CCE.

Notes
- Rational mode
- Formative mode
- Summative mode
- All the four basic techniques of assessment, i.e. testing, observation, inquiry and analysis, will be employed in different situations to collect the relevant evidences about pupils' growth and development in both cognitive and non-cognitive areas.
- Tests

    (a) Readiness tests

    (b) Criterion-referenced unit-tests

    (c) Annual or half-yearly tests

    (d) Diagnostic tests
- Depending upon the hard spots of learning as revealed in different unit-tests, specific diagnostic tests would be used on such selected areas.
- Performance on Assignments : Home and school assignments will be regularly used by teachers to reinforce what is taught and develop in students the ability to use what is learnt.
- Performance on Orals : Oral testing by means of class questioning, oral unit-tests, oral review sessions, quizzes, peer discussion etc.

    (a) Unless protagonists of reform develop a deep conviction and full faith (not 20% or 25%) in teachers' assessments and conceptualise dichotomy observable in theory and practice, establishment of a pedagogicaly sound system of SBA would be difficult.

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

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

1.10 Keywords

Testing : The activity of examining something/somebody to find something out.

Assessment : An opinion or a judgement about something that has been thought about very carefully

Context : The situation in which something happens and that helps you to understand it.

1.11 Review Questions

1. What do you understand by Concept of Educational testing ?
2. What is the nature of Educational Testing ?
3. Write the Testing Techniques.
4. What is Educational Assessment ?
5. Give current trends in Educational Assessment
6. What are the issues of Educational Testing?

Answers : Self Assessment   Notes

1. (*i*)   System of Examination   (*ii*)   Written tests, psychomotor domain
    (*iii*) Results   (*iv*)   Thought provoking
    (*v*)   Subjective attitude
2. (*i*)   T   (*ii*)   T   (*iii*) F   (*iv*) F

1.12 Further Readings

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 2: Measurement and Evaluation : Concept, Need, Scope, Difference and Relevance

CONTENTS

Objectives

Introduction

2.1 Concept of Measurement

2.2 Need of Measurement

2.3 Scope of Measurement

2.4 Steps of Measurement in Education

2.5 Concept of Evaluation

2.6 Need of Evaluation

2.7 Scope of Evaluation

2.8 Steps of Evaluation in Educational Achievements

2.9 Difference between Measurement and Evaluation

2.10 Relevance between Measurement and Evaluation

2.11 Summary

2.12 Keywords

2.13 Review Questions

2.14 Further Readings

Objectives

The objectives of this unit can be summarised as below :

- To explain the concept of Measurement
- To discuss about the need of Measurement
- To describe the scope of Measurement
- To discuss about the concept of Evaluation
- To explain about the need of Evaluation
- To describe about the scope of Evaluation
- To discuss about the steps of Evaluation in Educational achievements.
- To explain about the difference between measurement and Evaluation
- To discuss about the relevance between measurement and Evaluation

Introduction

In the present age, measurement has influenced the progress in education and psychology too. Today, the age of theoretical education is over, and effort is being made to make education and psychology more and more practical. Under education and psychology are studied different human behaviours and problems. For this, it becomes necessary to measure human behaviours.

Educational measurement is not a new concept. The teacher has been testing students since times immemorial in order to know their progress in studies, and to see what type of behavioural changes have occurred in students, if they are optimal and what direction these behavioural changes have taken. A teacher wants to know the shortcomings of the method of teaching he uses, for which these tests have been found to be very important. These tests have become all the more important in the present age, and it is expected from the future teacher or pupil-teacher that he would gain skill in constructing several types of tools of measurement and evaluation.

The introduction of evaluation in the educational world is comparatively new. In fact, it was introduced in this field in order to get rid of the serious shortcomings of measurement. From the beginning of the twentieth century, three types of important progresses were noted down in the field of educational measurement, they were testing, measurement and evaluation. This progress in education is known as progress in measurement. However, in the present times, evaluation is considered to be an important contribution of psychology and education. Recently, the attention of educationists and psychologists has been drawn towards evaluation.

2.1 Concept of Measurement

Generally, to measure and show the weight, length and volume of an object in definite units is called measurement; for example, to show the weight of a person in kilograms, length of cloth in metres and volume of milk in litres. But the field of measurement is very wide. It includes to define any characteristic of any object or person or activity in words, symbols or units.

As far as explaining the qualities of objects, persons and activities is concerned, it has been in vogue from very ancient times, of course, without any definite base of measurement. In the present times, the bases of most of the qualities of objects, persons and activities have been defined; their standards and units have been specified; measuring tools and methods have been devised and methods to demonstrate the results of measurement in brief have been decided. Now, a characteristic of an object, person or activity is described in definite words, symbols and units in brief. Many scholars have attempted to delimit the definition of this process. Most scholars are in agreement with the definition given by **James M. Bradefield.** In his words :

Measurement is the process of assigning symbols to the dimension of phenomenon in order to characterise in order to characterise the status of phenomenon as precisely as possible.

--- James M. Bradefield

In this definition of measurement only the measurement of qualities of objects and activities has been included, and not the measurement of qualities of persons. Though the persons are included in the objects of the universe; however, the objects are meant to include only concrete materials, so it is necessary to show the measurement of qualities of the persons separately. This definition of Bradefield does not point to any such basis of measurement. We are of the opinion that it should also be necessarily included in it and in that case measurement should be defined as :

Measurement is the process by which a characteristic of an object, person or activity is perceived and understood on specific standards and is described in standard words, symbols or definite units.

Factors of Measurement

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

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

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

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

(4)   The person who measures it.

2.1.1 Measuring Variables and Their Types

From the viewpoint of measurement, variables are meant by those characteristics of objects, persons and activities in which the objects, persons and activities are not uniform in a group; for example, the weight, intelligence and personality of persons. There is possibility that the weight of a few members may be equal, however all persons of a group cannot have the equal weight; and if a group is such formed that the weight of all members is equal, then the weight of such persons would become constant in place of being variable. It is evident that a characteristic can be variable for one group while it can be constant for another. We can clarify it further by another illustration. Sex in a group of boys and girls is a variable and they can be divided into two groups of boys and girls on the basis of their sex; however, in separate groups of boys or girls the sex is not a variable.

Variables are of two types—Qualitative variables and Quantitative variables.

Qualitative Variables

Some qualities of objects and persons are such which can only be perceived, but they cannot be measured in definite units; for example, the complexion, caste, religion and sex of people. These qualities or characteristics are called qualitative variables. The level or class of the students is another example of qualitative variable. On the basis of this variable, they can be classified into the students of primary, middle and higher classes or levels. The students of higher classes can also be classified on the basis of their subjects — art, commerce, engineering, medical, etc. At this level, the subjects (disciplines) also function as qualitative variable.

Quantitative Variables

Some qualities of objects and persons are such which can be measured in definite units or quantity; for example, height, weight and I.Q. of persons. Such qualities of persons are called quantitative variables. Proficiency of students in a particular subject is also included in quantitative variable because it can be measured in definite units by testing. Quantitative variables are of two types — Continuous variables and Discrete variables.

1. **Continuous Variables :** Those quantitative variables are included in continuous variables which can be of any quantitative value between any two continuous quantitative whole numbers for example, height of a person. It is not necessary for a person's height to be in the whole number, such as 171 *cm* following 170 *cm*. It can also be 170.1 *cm*, 170.2 *cm*, 170.3 *cm*, 170.4 *cm*, 170.5 *cm*, 170.6 *cm*, etc. or it can be even 170.11 *cm* and 170.12 *cm*. The units used in the continuous variables are never whole numbers by themselves, rather they are approximate numbers, they have a part in them.

2. **Discrete Variables :** Those quantitative variables are included in discrete variables which are always measured in whole numbers; for example the number of students in a class. This number can always be a whole number (40, 41, 50, etc.) and never a part of it (40.5, 41.51, 45.52, etc.). The discrete variables are always in whole numbers, so their units of measurement are always exact numbers.

2.1.2 Qualitative and Quantitative Measurement

Qualitative Measurement

Perceiving the characteristics of an object, person or activity in the form of a quality is called qualitative measurement; for example, describing a student as very intelligent, or dull is qualitative measurement.

Quantitative Measurement

Measuring the characteristics of an object, person or activity in the form of quantity is called quantitative measurement; for example, to measure the I.Q (Intelligence Quotient) of a student as 140, 120 or 110 is quantitative measurement.

Difference between Qualitative and Quantitative Measurement

Qualitative measurement is different from quantitative measurement in the following ways :

(1) The bases of qualitative measurement are often norms; and those of quantitative measurement are units.

(2) The norms of qualitative measurement are generally not universal, while the units of quantitative measurement are universal.

(3) There is never a situation of zero (0) in qualitative measurement; for example, the intelligence of a child can never be zero. On the contrary, the basis of quantitative measurement is zero under all circumstances; for example, when we say that the weight of a child is 40 *kg.*, it means that his weight is more than zero (0) by 40 *kg.* The most suitable example of this is temperature which may be both more than zero (0) and less than zero (0).

Notes   Qualitative measurement can never be done completely in itself, for example, inability to fully measure the interest of a child. On the other hand, quantitative measurement can be done fully in itself; for example, measuring the weight of a child fully.

---

(4) Qualitative measurement has no mathematical relationship; for example, the proficiency in mathematics of a child securing 60 marks is not necessarily the double than that of a child securing 30 marks. On the contrary, quantitative measurement has mathematical relationship; for example, the weight of a child with 60 *kg.* is double than that of a child with a weight of 30 *kg.*

2.1.3 Levels, Scales and Types of Measurement

S.S. Stevens has divided the process of measurement into four levels on the basis of their facts — Nominal level, Ordinal level, Interval level and Ratio level. Some scholars term them as scale,— Nominal scale, Ordinal scale, Interval scale, and Ratio scale. On the basis of these four levels or scales, measurement is of four types — Nominal measurement, Ordinal measurement, Interval measurement and Ratio measurement. We shall discuss them here in brief. They are four in numbers.

1. **Nominal Measurement :** In the measurement of this level or scale, a particular characteristic of an object or person is seen, and on its basis is allotted the classification or symbol; for example, to name the students as boys or girls on the basis of sex, to name the students as urban or rural on the basis of place, to name the students as belonging to primary class, middle class, higher class, medical class or engineering class, on the basis of level of education.

   It is evident that nominal measurement is a qualitative measurement and is very ordinary in its approach. It is the least appropriate type of measurement.

2. **Ordinal Measurement :** In the measurement of this level or scale, the quantity of a characteristic of an object or person is measured, and is allotted a classification, name or symbol on the basis of quantity; for example, to classify the students as belonging to high intelligence, medium intelligence or low intelligence on the basis of quantity of their intelligence; or to classify the students as belonging to first class, second class or third class on the basis of the marks obtained.

   It is evident that the measurement of this level is quantitative, and the classifications made within it have inter-relation. In the examples given above, the students of the first classification are better than those of the second classification, and the students of the second classification are better than those of the third classification; they are inter-related. The measurement of this level is better than nominal measurement.

3. **Interval Measurement :** In the measurement of this level or scale, a trait of an object or person is measured in unit numbers and there is equal difference between any two continuous unit numbers; for example, to award the students with marks 40, 38, 52, 64, etc. in an examination.

These unit numbers have a difference of one (1) between 40-41, and 38-39 to have a difference of a unit number; but in this level there is no unit like zero (0) which may express absence of trait. For example, if a student scores zero in an achievement test of a subject, it does never mean that his achievement is nil or his ability in the subject is nil. This zero only signifies that the student has not succeeded in solving any of the questions asked in the achievement test. Because there is possibility that he can solve other questions of the subject, so his knowledge in the subject cannot be zero.

It is evident that interval measurement is better than the first two types of measurements, still it is not a fully appropriate measurement. But this type of measurement is mostly used in the field of education.

4. **Ratio Measurement :** The measurement of this level possesses all the characteristics of interval measurement, besides having the concept of true zero. The true zero is the point at which the trait is completely absent. From the concept of this zero, the obtained results can be compared and they can be given ratio value. So it is called ratio measurement. For example, if the weight of Ram is 60 *kg* and that of Shyam 20 *kg*, then it can be said that the ratio in their weights is 3 : 1.

It is clear that the measurement of this level can be used for physical traits (variables), and not for abstract traits (variables). For example,

---

Notes If I.Q of Ram in an intelligence test is 150, and that of Shyam 100, then it cannot be said that the ratio in the intelligence of Ram and Shyam is 3 : 2. Ram's intelligence can be more or less than $\frac{3}{2}$ of that of Shyam.

Absolute, Normative and Ipsative Measurement

On the basis of tools or devices and the results obtained therefrom, measurements are classified into three types—Absolute measurement, Normative measurement and Ipsative measurement. We introduce them here for your convenience.

Absolute Measurement

Absolute measurement is that measurement in which the scale starts from zero. It measures + ve on being more than zero, and — ve on being less than zero. For example, take the temperature of a place. It can be zero, it can be more than zero and it can be less than zero. Such measurement is possible only in physical variables, and not in educational and psychological variables, because the educational and psychological variables have no possibility of an absolute zero. For example, if a student obtains zero in an achievement test, it does never mean that his ability is absolutely zero. It only means that the student has failed to solve any of the questions asked in the achievement test.

There exists absolute zero in absolute measurement, so the analysis of its results can be carried out on mathematical basis. For example, if the temperature of a place is 40 degrees, and that of another place 20 degrees, then it can be said that the temperature of the first place is double than that of the second place. Such kind of mathematical analysis cannot be carried out in the measurement of educational and psychological variables. For example, if Student A scores 60 marks and Student B scores 30 marks in an achievement test, then it cannot be said that the ability of Student A is double than that of Student B in the subject. It can be more or less than double.

Normative Measurement

Normative measurement is such measurement in which the marks obtained are not affected by one another. Its another characteristic is that there is no possibility of absolute zero in it. For example, if

a student scores zero in an achievement test, it does never mean that his ability in the subject is zero. It only signifies that he has failed to solve any of the questions in the achievement test. There may be possibility that he can solve other questions on the same subject.

This measurement has not absolute zero, so its results cannot be analysed mathematically. For example, if Student A scores 60 marks and Student B scores 30 marks, then it cannot be said that the ability of Student A is double than that of Student B. It can be more or less than double. The analysis of such measurements can be done with the help of statistical calculations — measures of central tendency and measures of deviation. We shall clarify the form of normative measurement by an example.

Ipsative Measurement

There are several tools and devices of measurement. One such tool or device is the one in which a person or student is forced to choose items. Measurement by this method has been termed as Ipsative measurement by **Ramond Cattle.** In this type of measurement certain questions, statements or problems are put before the persons or students and they are asked to give them preference 1, 2, 3,4, etc. It is obvious that if a person gives preference (1) to a statement, he cannot give the same preference to any other statement. Such measurement tool is called forced choice item and the measurement is called the ipsative measurement. In such measurement, all persons or students have to allot 1, 2, 3, 4, etc. marks, so the mean (M) and standard deviation (σ) of the marks of all persons or students are bound to be equal. We shall clarify it with the help of an example.

2.2 Need for Measurement

The needs of measurement are as follows:-

(1) For Measure the ability of students, finding out their interest sand aptitude at the time of admission, and admit than on its basis.

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

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

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

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

2.3 Scope of Measurement

Generally, it is understood that in the field of education only educational achievements of the students are measured, and are expressed in terms of marks. However, in the field of education, in present times, various variables, such as intelligence, interest, aptitude, attitude and personality of the students are also measured; the aims of education, usefulness of the curriculum and of the teaching methods are also measured; basis of educational policy is also measured; and the educational activities of the persons concerned such as administrators, teachers, and guardians, are also measured. There is no area in the field of education which is not subject to measurement in present times and the result of which is not expressed in more and more objective terms. Therefore, we can define educational measurement as follows :

Educational measurement is that process by which the usefulness of various factors of educational process, the usefulness of various activities of persons concerned, and the intelligence, interest, attitude, aptitude, personality and educational achievements of the students are measured on the basis of definite standards and are expressed in definite words, symbols or units.

Measurement is an extensive process which has several tools and devices. The physical aspects of a person, such as height, weight and temperature are measured with the help of scale, weights and

thermometer respectively and his interest, aptitude, intelligence, personality and educational achievement are measured with the help of Interest Tests, Aptitude Tests, Intelligence Tests, Personality Tests and Achievement Tests respectively. Evidently, measurement is a process to measure the characteristic of an object, person or activity; and testing is only a device or method of measurement. They are different like whole and its part.

2.4 Steps of Measurement in Education

The process of measurement is completed in four steps in any field, including the field of education. These steps are :

1. **Determination and Defining the Measuring Traits or Aims and Objectives :** At first the measurer determines which quality of which person (teacher, student, administrator, guardian, etc.) or which educational achievements of the students he has to measure. After having determined it, he gives it a definite form and defines it. For example, if he has to provide educational guidance, he has to determine the traits that he has to measure, such as intelligence, interest, aptitude, attitude, etc. In case he has to measure the educational achievements, then he has to determine the aims and objectives that he has kept in mind while teaching a subject or training in an activity, and he has to determine the extent to which he has to achieve it.

2. **Selection of Suitable Measurement Tools or Methods :** After having determined the traits aims and objectives to be measured, the measurer selects suitable tools or techniques for their measurement. For example, if he has to measure the intelligence of students, then he has to select the intelligence test according to the age of the students, and if he has to measure the educational achievements of the students, then he has to select a suitable performance test according to the aims and objectives. In the absence of clear knowledge of the traits or aims and objectives to be measured, he would not be able to select a suitable measurement tool or technique; and in the absence of it, he cannot measure the traits or educational achievements of the students.

3. **Use or Administration of Measurement Tool or Technique :** After having selected the suitable measurement tool or technique, the measurer uses or administers it. For example, if he has selected an intelligence test to measure the students' intelligence, then he will administer it; or if he has selected or constructed an achievement test for the measurement of educational achievement, then he will administer it. He will be able to administer it properly only when he is acquainted with the knowledge of its administration.

4. **Results and Records :** This is the last step of measurement process. At this step, the measurer obtains the results of measurement and records them. For example, if he has administered an intelligence test on students, he will calculate intelligence quotient (IQ) on the basis of the obtained results and will record it. In case he has administered a performance test, then he will award marks on it and record the scores.

Did u know? At first, the attention of educationists was drawn to the testing of intelligence and achievement of students. The tests in the nineteenth century were generally subjective. The objective tests were introduced in order to do away with the shortcomings of these tests. This effort was reactionary. After these tests were found to be very successful in Europe and America, Indian educationists used these tests in our country in order to measure educational results which were not feasible and possible to be measured by subjective tools.

Self Assessment

1. *Fill in the blanks :*

    (*i*)   _____ is the process of assigning symbols to the dimension of phenomenon in order to characterise the status of phenomenon.

    (*ii*)   Complexion, Caste, religion, of people are called _____.

    (*iii*)   _____ has divided the process of measurement into four levels.

    (*iv*)   _____ is measurement in which the marks obtained are not affected by one another.

    (*v*)   _____ is that measurement in which the scale starts from Zero.

2.5 Concept of Evaluation

We are aware that measurement is used to express a trait of an object, person or activity in standard words, symbols or units. In evaluation, these results are analysed and this analysis is done on the basis of certain social, cultural or scientific standards (Norms) and by this analysis, the relative condition of the trait of the object, person or activity is clarified. **James M. Bradefield** has defined this process of evaluation in the following words :

Evaluation is the assignment of symbols to phenomenon in order to characterise the worth or value of the phenomenon usually with reference to some social, cultural and scientific standards.

—James M. Bradefield

If we want to further clarify this definition in simple terms, we can do it in the following words :

Evaluation is the process in which the analysis of the result obtained from measurement of a trait of an object, person or activity is done on the basis of certain social, cultural or scientific standards (Norms), and the relative position of the object, person or activity is determined as relative to that trait.

Factors of Evaluation

Two processes have to be undertaken in evaluation — first, the measurement and the second, analysis of the information or data obtained from measurement. And we are aware that there are the following four factors of measurement :

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

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

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

(4)   And the person who measures it.

    There are two factors of analysis of the data or result received from the measurement, and we can assign them serial numbers 5 and 6 respectively, which are :

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

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

Subjective and Objective Evaluation

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

Formative Evaluation

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

Summative Evaluation

Summative evaluation is such evaluation which is conducted in order to test the utility of an already existing educational policy, planning or programme, curriculum, teaching method, teaching aid or evaluation method. For it, the evaluator constructs the most suitable measurement tool or method based on interview, questionnaire or rating scale, for evaluation of the educational policy, planning or programme, curriculum, teaching method, teaching aid or evaluation method. After this, he tests its utility on the basis of related standards (Norms) and statistical calculations. Finally, he decides whether such educational policy, planning or programme, curriculum, teaching method, teaching aid or evaluation method should continue or not, and if it is to continue, what should be its form.

2.6 Need of Evaluation

The need of the Evaluation are as follows:-

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

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

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

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

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

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

2.7   Scope of Evaluation

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

Educational evaluation is the process in which measurement of the decisions related to education and of the traits and activities of the persons concerned with education is carried out and the results of such measurement are analysed on the basis of predetermined standards (Norms) and on its basis the relative results are declared and suggestions for improvement in them are given.

Generally, people use the words educational evaluation and measurement in identical meanings. However, they have the difference of the whole and the part. Measurement is the first step of evaluation, evaluation is a process to analyse the results of measurement.

From the functional steps viewpoint, the process of measurement has only three steps — (i) selection or construction of suitable measurement tool or device, (ii) use of measurement tool or device, and (iii) recording of administration and result. Evaluation has three additional steps besides the above— (i) selection and use of suitable standards (Norms) for analysing the measurement results, (ii) analysis of measurement results on the basis of these standards (Norms), and (iii) prediction, suggestions or guidance on the basis of this analysis.

The analysis of results of educational measurement under educational evaluation cannot be done as of the results of physical measurement, because the physical measures are whole measures, and educational measures are relative in nature. For example, if the weight of Ram's weight is double than that of Shyam 30 *kg*, then it can be said that Ram's weight is double than that of Shyam. However, in an educational achievement of Ram is double than that of Shyam. It can be more or less than the double. Now statistical methods are used for their analysis.

Difference between Educational Evaluation and Testing

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

Difference between Educational Evaluation and Examination

Educational evaluation is a broad concept. Under it, the measurement and analysis of educational policy, educational planning, educational objectives, activities of educational administrators, teachers and guardians, and students' intelligence, interest, aptitude and personality, etc. is carried out, and in the end, predictions, suggestions and guidance are provided. On the other hand, examination is used for measurement and evaluation of only educational achievements.

> Notes
>
> Though there is a difference between measurement and evaluation, but because measurement is a part of evaluation and is its first step, so now it is used as a joint concept in the field of education as 'measurement and evaluation'. Secondly, there is no significance of measurement in the field of education until its results are not analysed. When the results of measurement are analysed, it is called evaluation. Thirdly, whether we join measurement with evaluation or not, the process of measurement is included in it.

2.8 Steps of Evaluation in Educational Achievements

The educational achievements of the students are mainly evaluated in the field of education. It is done in two forms — at the end of the session for promotion purposes, and during teaching to see the effect of teaching. The first kind of evaluation is complete with passing or failing the students, classifying the students in first, second and third divisions and to provide them results. However, the second kind of evaluation goes on until the students learn what we want to teach them, and the desired change in their behaviour is effected.

We know that the aims of teaching of different subjects in schools are different. Various activities are carried out for achievement of these aims. And finally, measurement and evaluation of the achievement of the students is conducted. In the above evaluation, the evaluator clarifies, after evaluation, how far the aims have been achieved. In case of failure the evaluator re-inspects the teaching aims, evaluates the teaching-learning activities and suggests for improvement. This process goes on until the aims are realised, the students learn what we want them to learn, and a desired

change of behaviour is effected in them. This process of evaluation may be clarified with the help of the following sketch :

Teaching Objectives

Evaluation

Teaching-Learning
activities

Evaluation
Techniques
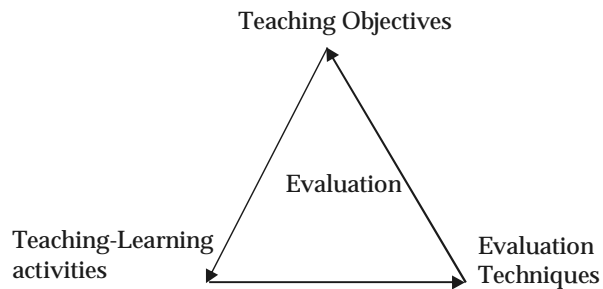
With reference to evaluation of educational achievements, the steps of evaluation may be written and understood as follows :

(1) Clarification of teaching objectives.

(2) Selection of suitable performance test or practical work.

(3) Administration of performance test or practical work.

(4) Marking and recording of scores on the answer-sheet or practical work.

(5) Analysis of measurement results.

(6) Suggestions on the basis of analysis of measurement results (evaluation).

(7) Evaluation of the effect of the suggestions.

(8) Evaluation is used to give educational and vocational guidance to students.

(9) Evaluation is used to ascertain the amount of learning of students.

(10) Evaluation helps to know about the weaknesses and strengths of students.

(11) Evaluation is used to ascertain the effectiveness of instruction and to plan activities accordingly.

(12) A purpose of evaluation is to ascertain how different individual and social needs of students can be met.

(13) The chief purpose of evaluation is to classify students.

(14) By evaluation, the skill and success of teachers are measured.

(15) A purpose of evaluation is to enable students to understand their problems and progress.

(16) Suitable amendments can be effected to the curriculum by evaluation.

(17) Evaluation tests suitability of methods of teaching.

Task What are forced choice items ?

2.9 Difference between Measurement and Evaluation

Generally the term 'measurement' is confused with 'Evaluation', while they are quite apart from each other. Measurement is a quantitative form of an object, while evaluation presents its qualitative together with its measurement. In brief, measurement is numerical and evaluation quantitative. Measurement tells us how much of an object, while evaluation tells us how good it is. Besides, evaluation pays attention to the fact of realization of objectives, while by measurement is meant only the number of specific objectives that have been realized. Measurement is incomplete without evaluation.

Distinguishing between measurement and evaluation, **Wrightstone** has written :

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

Measurement	Evaluation
1. Measurement does not express any clear assumption about a student.	1. Clear assumption about a student can be formed on the basis of evaluation.
2. Measurement does not require much energy and time.	2. Evaluation requires more energy and time.
3. The scope of measurement is limited, only some dimensions of personality can be tested under measurement.	3. The scope of evaluation is wide, in it, the entire personality of a student is tested.
4. Measurement is content-oriented.	4. Evaluation is objective-oriented.
5. Measurement is a means, and not an end in itself.	5. Evaluation is an end in itself.
6. The purpose of measurement is to gather evidences.	6. Evaluation is to deduce inferences from evidences, that is, its work is appraisement of evidences.
7. Measurement may not be an essential part of education.	7. Evaluation is the integrated or necessary part of education.
8. Measurement answers the question 'how much', such that Sunanda has scored 56 marks in mathematics. This is measurement.	8. Evaluation answers the question 'what value'. Sunanda has scored 50% marks in mathematics and passed second division in the class, this is evaluation.
9. Prediction cannot be made meaningfully on the basis of measurement.	9. Evaluation can predict meaningfully.
10. Measurement acquaints with a situation. This is isolated from the entire environment.	10. Evaluation acquaints about the entire situation.
11. Measurement indicates those observations which are displayed numerically.	11. Evaluation comprises of both quantitative and qualitative observations.
12. Measurement can be conducted at any time.	12. Evaluation is a continuous process.

2.10 Relevance between Measurement and Evaluation

Measurement and Evaluation both are used for assess the internal qualities. I.Q. aptitude, attitude and intelligence of students.

The main purposes of measurement and evaluation are:-

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

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

- Certification: This helps to certify that a student has achieved a particular level of performance.

- Stimulating learning: this can be motivation of the student or teacher, providing feedback, suggesting suitable practice etc.
- Improving teaching: by helping to review the effectiveness of teaching arrangements.

Self Assessment

2. State Whether the following statements are 'true' or false'.
  (*i*) Summative Evaluation is such evaluation which is conducted before giving final shape to an educational policy or program, curriculum teaching method.
  (*ii*) The educational achievements of the students is done in three forms.
  (*iii*) The measurement is the first step of evaluation
  (*iv*) According to James M. Bradefield, The evalution is the assignment of symbols to phenomenon in order to characterise the worth or value of the phenomenon with reference to some social, cultural and scientific standards.

2.11 Summary

- Measurement is the process by which a characteristic of an object, person or activity is perceived and understood on specific standards and is described in standard words, symbols or definite units.
- Variables are meant by those characteristics of objects, persons and activities in which the objects, persons and activities are not uniform in a group.
- Variables are of two types—Qualitative variables and Quantitative variables.
- Qualitative Variables
  - Some qualities of objects and persons are such which can only be perceived, but they cannot be measured in definite units; for example, the complexion, caste, religion and sex of people.
- Quantitative Variables
  - Some qualities of objects and persons are such which can be measured in definite units or quantity; for example, height, weight and I.Q. of persons.
- Quantitative variables are of two types — Continuous variables and Discrete variables.
  (*i*) Continuous Variables : Those quantitative variables are included in continuous variables which can be of any quantitative value between any two continuous quantitative whole numbers.
  (*ii*) Discrete Variables : Those quantitative variables are included in discrete variables which are always measured in whole numbers; for example the number of students in a class.
- Qualitative Measurement
  - Perceiving the characteristics of an object, person or activity in the form of a quality is called qualitative measurement;
- Quantitative Measurement
  - Measuring the characteristics of an object, person or activity in the form of quantity is called quantitative measurement;
- Qualitative measurement is different from quantitative measurement in the following ways:
  (*i*) The bases of qualitative measurement are often norms; and those of quantitative measurement are units.
  (*ii*) The norms of qualitative measurement are generally not universal,
- We shall discuss them here in brief. They are four in numbers.
  (*i*) Nominal Measurement : In the measurement of this level or scale, a particular characteristic of an object or person is seen, and on its basis is allotted the classification or symbol.
  (*ii*) Ordinal Measurement : In the measurement of this level or scale, the quantity of a characteristic of an object or person is measured, and is allotted a classification, name or symbol on the basis of quantity.

(*iii*) Interval Measurement: In the measurement of this level or scale, a trait of an object or person is measured in unit numbers and there is equal difference between any two continuous unit numbers.

(*iv*) Ratio Measurement : The measurement of this level possesses all the characteristics of interval measurement, besides having the concept of true zero. The true zero is the point at which the trait is completely absent.

- On the basis of tools or devices and the results obtained therefrom, measurements are classified into three types—Absolute measurement, Normative measurement and Ipsative measurement.

- Absolute measurement is that measurement in which the scale starts from zero. It measures +ve on being more than zero, and — ve on being less than zero.

- Normative measurement is such measurement in which the marks obtained are not affected by one another. Its another characteristic is that there is no possibility of absolute zero in it.

- There are several tools and devices of measurement. One such tool or device is the one in which a person or student is forced to choose items. Measurement by this method has been termed as Ipsative measurement by Ramond Cattle. In this type of measurement certain questions, statements or problems are put before the persons or students and they are asked to give them preference 1, 2, 3,4, etc.

(*i*) For Measure the ability of students, finding out their interest sand aptitude at the time of admission, and admit than on its basis.

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

- Educational measurement is that process by which the usefulness of various factors of educational process, the usefulness of various activities of persons concerned, and the intelligence, interest, attitude, aptitude, personality and educational achievements of the students are measured on the basis of definite standards and are expressed in definite words, symbols or units.

- The process of measurement is completed in four steps in any field, including the field of education. These steps are :

(*i*) Determination and Defining the Measuring Traits or Aims and Objectives :

(*ii*) Selection of Suitable Measurement Tools or Methods :

(*iii*) Use or Administration of Measurement Tool or Technique :

(*iv*) Results and Records :

- Evaluation is the process in which the analysis of the result obtained from measurement of a trait of an object, person or activity is done on the basis of certain social, cultural or scientific standards (Norms), and the relative position of the object, person or activity is determined as relative to that trait.

- Formative evaluation is such evaluation which is conducted before giving final Shape to an educational policy or programme, curriculum, teaching method, teaching aid or evaluation method.

- Summative evaluation is such evaluation which is conducted in order to test the utility of an already existing educational policy, planning or programme, curriculum, teaching method, teaching aid or evaluation method.

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

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

Notes
- In the field of education, generally, the measurement of educational achievements of the students is called evaluation. The first thing in this context is that the measurement is the first step of evaluation, it is not evaluation in itself.

- Generally the term 'measurement' is confused with 'Evaluation', while they are quite apart from each other. Measurement is a quantitative form of an object, while evaluation presents its qualitative together with its measurement. In brief, measurement is numerical and evaluation quantitative. Measurement tells us how much of an object, while evaluation tells us how good it is. Besides, evaluation pays attention to the fact of realization of objectives, while by measurement is meant only the number of specific objectives that have been realized. Measurement is incomplete without evaluation.

2.12 Keywords

Measurement : The act or the process of findings the size, quantity or degree of something.

Evaluation : The process of forming an opinion of the amount, value or quantity of something.

Determination : The quantity that makes you continue trying to do something even when this is difficult.

2.13 Review Questions

1. What do you understand by measurement ? Clarify the difference between absolute normative and Iptative measurement ?

2. What are qualitative and quantitive measurement ?

3. What do understand by evaluation ? What is the difference between formative and Summative evaluation ?

4. What is the difference between measurement and evalution ?

5. What is the scope of evaluation in education.

Answers : Self Assessment

1. (*i*) Measurement (*ii*) qualitative variables
   (*iii*) S.S. Stevens (*iv*) Normative measurement
   (*v*) Absolute measurement
2. (*i*) F (*ii*) F (*iii*) T (*iv*) T

2.14 Further Readings

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*
2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*
3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*
4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 3: Scales of Measurement : Ordinal, Nominal, Interval, Ratio

---

CONTENTS

Objectives

Introduction

3.1 Concept of Scales of Measurement

3.2 Properties of Measurement Scales

3.3 Nominal Scale

3.4 Ordinal Scale

3.5 Interval Scale

3.6 Ratio Scale

3.7 Summary

3.8 Keywords

3.9 Review Questions

3.10 Further Readings

Objectives

The objectives of this unit can be summarised as below :

- To explain about the Concept of Scales of Measurement

- To discuss about the Properties of Measurement Scales

- To describe about the Nominal Scale

- To explain about the Ordinal Scale

- To discuss about the Interval Scale

- To describe about the Eatio Scale

Introduction

Measurement scales are used to categorize and/or quantify variables. This lesson describes the four scales of measurement that are commonly used in statistical analysis : nominal, ordinal, interval, and ratio scales. The interval scale of measurement has the properties of identity, magnitude, and equal intervals. A perfect example of an interval scale is the Fahrenheit scale to measure temperature. The scale is made up of equal temperature units, so that the difference between 40 and 50 degrees Fahrenheit is equal to the difference between 50 and 60 degrees Fahrenheit.

The ratio scale of measurement satisfies all four of the properties of measurement : identity, magnitude, equal intervals, and an absolute zero. Absolute zero is a property of the weight scale because objects at rest can be weightless, but they cannot have negative weight. We will study about all the Scales of measurement.

3.1 Concept of Scales of Measurement

Any test can be useful only when it is reliable, and it should be able to measure only that attribute or characteristic for which it has been constructed. The tools for measurement have been needed from ancient times, and we need them in our daily life.

The basis of educational measurement are data. Whatever the type of measurement-physical, social, economic or psychological, it is necessary to gather data. From the viewpoint of convenience, we place the available data into four levels. These four levels are arranged in a definite order. The lower level can be easily measured, but the measurement done by it will be under some doubt. On the contrary, measurement in the higher level is more complex, but the inferences drawn from it will be more accurate. Thus, accuracy of measurement depends on its level. Generally, individuals, objects, events, observations and characteristics are given a quantitative form under the measurement process. Measurement Scales are used to categorize or quantity variables. In essence, each scale has a single objective, and its rules, theory, characteristics, limitations and statistical techniques are different from other levels fully.

Measurement has the following four chief levels :

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

3.2 Properties of Measurement Scales

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

- **Identity :** Each value on the measurement scale has a unique meaning.
- **Magnitude :** Values on the measurement scale have an ordered relationship to one another. That is, some values are larger and some are smaller.
- **Equal intervals :** Scale units along the scale are equal to one another This means, for example, that the difference between 1 and 2 would be equal to the difference between 19 and 20.
- **Absolute zero :** The scale has a true zero point, below which no values exist.

Values assigned to variables represent a descriptive category, but have no inherent numerical value with respect to magnitude.

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

> *Notes* The ordinal scale has the property of both identity and magnitude. Each value on the ordinal scale has a unique meaning, and it has an ordered relationship to every other value on the scale.

3.3 Nominal Scale

This is the lowest level of measurement. Some people call it by the name of classification level too. Under this scale, the measured objects or events are classified into separate groups on the

basis of their certain attributes, and this group is given a separate name, number or code for its easy identification. The chief feature of this group is that all elements or individuals will be similar to each other within the group but they will be entirely different when compared to those of another group. This feature of the group is called internal homogeneity. For example, the cricket teams of Sri Lanka and Australia will be given different colour dresses in order easy identification and their dresses will be marked with the letter S and A respectively. In the same way, women and men, fair and dark, rural and urban people will be kept under separate groups in order to distinguish the difference. In the same manner, allotting PIN codes for post distribution, dividing a metropolitan city into zones, such as New Delhi 110001, 110009, 110065 or 10081 etc., the railway division being allotted a different logo, the banks being given separate signs, as Bank of lndia-3, the New Bank of lndia-7, Canara Bank-7, Syndicate Bank-5 etc. Different types of fruits can be given codes, or the football players are allotted numbers etc. All these types of classifications are used at this level of measurement. This level is not important from the viewpoint of research, because the only statistical operation or technique involved is counting or calculation.

3.4 Ordinal Scale

In the arrangement of scales, the ordinal scale is put at the second place from down below. In this scale, objects, individuals, events, characteristics or responses are arranged in hierarchical order in ascending or descending order depending on the basis of certain attributes. After that, they are given ranks. Giving first, second or third position or rank to students on the basis of their scores, giving preference in employment to candidates on the basis of eligibility and experience, awarding trophy to players on the basis of their performance, selecting Miss World or Miss Universe on the basis of beauty, selecting the best industrialist, selecting professors for the college proctorial board and arranging them in hierarchical order in view of their administrative accomplishment, arranging fruits on the basis of their taste and flavour, etc. are some of the illustrations of this scale. In this scale, we generally make use of two methods for arranging objects : linear method and pair comparison method.

The first method, the linear method, is quite simple. In this, objects are arranged as per their rank, such as the sequence of Indian cricket players can be arranged in a sequence on the basis of their performance in a series : Sachin Tendulkar, Gautam Gambhir, Rahul Drawid, Virendra Sahwag, Yuvraj Singh, M.S. Dhoni, Virat Kohli, Praveen Kumar, Harbhajan Singh, Srisant, Zaheer Khan. In the second method, the pair comparison method, all members of the group are compared in pairs. For determining the number of pairs in the group, the following formula is used :

$$^n C_2 = \frac{n(n-1)}{2}$$

Where, $^n C_2$ = The number of total pairs
and n = Number of total players in the group

Though this scale is used more than the nominal scale, yet from the standpoint of research, this is not accepted as very valid and reliable. Under this scale, though the median, percentiles, correlation multiple (r) etc. can be used to distinguish the difference between two individuals, yet it does not clarify the actual difference between the two. This is the chief limitation of this scale.

Did u know? If a student scores zero in an achievement test of a subject, it does never mean that his achievement is nil or his ability in the subject is nill. The zero only signifies that the student has not succeeded in solving any of the questions asked is the achievement test.

3.5 Interval Scale

This is the third level of measurement. This scale endeavours to do away with the limitations of the above two scales. Under this scale, we display the difference between any two classes, individuals or objects by the medium of scores. The distance between two differences is equal.

Lack of exact zero point is a shortcoming of this scale, due to which the measurement done by this scale is relative measurement, and not absolute; that is, if a student obtains zero marks in this scale, then it should not be concluded that the student is fully ignorant of the given subject. Some examples of this scale are thermometer, hour, minute, week month, year etc. In a thermometer, the normal temperature of an individual is considered at 98.4°F, but if due to certain reasons, this temperature is read at 97°F, then on the one hand, this shows that that person has no fever, but it should not be concluded that the individual's body has no heat or temperature at all. A thermometer is the most appropriate example of this scale. A thermometer indicates from 98°F to 108°F. It has the same distance between 98°F and 99°F, so it is between 107°F and 108°F. Under this scale, several statistical calculations can be used, such as mean, percentiles, standard deviation etc.

Notes   Though there is no exact or absolute zero point, in internal scale, still the scores arranged at equal distances are considered the constant unit of this scale.

3.6 Ratio Scale

This is the highest level of measurement. This scale comprises of all features of all other scales. The presence of exact or true zero point is the chief feature of this scale. This zero point is not arbitrary point, rather it is related with the zero amount of certain attribute or feature. In physical measurement, there is always an absolute zero point, such as meter, *km*, gram, liter, millimeter etc. Measurement of height, length, weight or distance is started from zero point. In ratio scale, the true zero point is considered the initial point of the scale. So, we can find out the ratio between the distance of any two places, and on its basis, we can say with certainty how distant is one place from another. Thus, if Rekha, Pinki or Puja are awarded 10, 20 and 40 marks on the basis of certain attribute, then we will say as per this scale in what measure this attribute exists in Rekha, and it exists in Pinki in the double measure and in Puja it is four times. It means to say that each unit in the scale explains different amounts of the attribute, and there is possibility of applying all basic operations in this scale.

Task   If I.Q of Ram is an intelligence test is 150, and that of Shyam 100, then what is that the ratio of the intelligence ?

Self Assessment

1. *Fill in the blanks :*

    (*i*)   Nominal scale is also called _____ too.

    (*ii*)   Each value on the measurement scale has a _____ meaning.

    (*iii*)   The presence of exact or true _____ is the chief feature of the ratio scale.

(*iv*) A thermometer is the most appropriate example of _____.

(*v*) The _____ is the chief feature of nominal scale.

3.7 Summary

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

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

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

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

  - Absolute zero : The scale has a true zero point, below which no values exist.

- The basis of educational measurement are data. Whatever the type of measurement-physical, social, economic or psychological, it is necessary to gather data. From the viewpoint of convenience, we place the available data

- This is the lowest level of measurement. Some people call it by the name of classification level too. Under this scale, the measured objects or events are classified into separate groups on the basis of their certain attributes, and this group is given a separate name, number or code for its easy identification.

- In the arrangement of scales, the ordinal scale is put at the second place from down below. In this scale, objects, individuals, events, characteristics or responses are arranged in hierarchical order in ascending or descending order depending on the basis of certain attributes.

- This is the third level of measurement. This scale endeavours to do away with the limitations of the above two scales. Under this scale, we display the difference between any two classes, individuals or objects by the medium of scores.

- This is the highest level of measurement. This scale comprises of all features of all other scales. The presence of exact or true zero point is the chief feature of this scale. This zero point is not arbitrary point, rather it is related with the zero amount of certain attribute or feature.

3.8 Keywords

Scales : A range of levels or numbers used for measuring something.

Absolute : Total and complete

3.9 Review Questions

1. What is scale of measurement ? Explain

2. What are the properties of scales of measurement

3. Explain the different scale of measurement
    (*i*) Nominal (*ii*) Ordinal (*iii*) Interval (*iv*) Ratio

Answers : Self Assessment

(*i*) Classification level (*ii*) Unique

(*iii*) Zero point (*iv*) Interval scale

(*v*) Internal homogeneity

3.10 Further Readings

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 4: Characteristics of a Good Test

CONTENTS

Objectives

Introduction

Objectives

The objectives of this unit can be summarised as below :

- To explain the characteristics of good test
- To discuss about question
- To describe the characteristics of a good question
- To discuss about the genotypic varieties of a question

Introduction

There are different varieties of tests : achievement tests, diagnostic tests, mental-ability tests, selection tests and psychological tests, besides the class-room unit tests. It is always assumed in testing that the measuring instrument or tool (test, question etc.) is of good quality and whatever evidences are gathered are dependable. This assumption is, however, questionable unless we know or determine the quality of the tool we use. In this unit we will study about characteristics of a good test.

4.1 Characteristics of a Good Test

A good test should possess the following qualities:-

Validity

    A test is considered as valid when it measures what it is supposed to measure.

Reliability

    A test is considered reliable if it is taken again by the same students under the same circumstances and the score average is almost the constant, taking into consideration that the time between the test and the retest is of reasonable length.

Objectivity

- A test is said to be objective if it is free from personal biases in interpreting its scope as well as in scoring the responses.

- Objectivity of a test can be increased by using more objective type test items and the answers are scored according to model answers provided.

Objective Basedness

- The test should be based on pre-determined objectives.
- The test setter should have definite idea about the objective behind each item.

Comprehensiveness

- The test should cover the whole syllabus.
- Due importance should be given all the relevant learning materials.
- Test should be cover all the anticipated objectives.

Validity

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

There are different types of validity :

- Operational validity
- Predictive validity
- Content validity
- Construct validity

Operational Validity

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

Predictive Validity

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

Content Validity

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

Construct Validity

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

Reliability

- Reliability of a test refers to the degree of consistency with which it measures what it indented to measure.
- A test may be reliable but need not be valid. This is because it may yield consistent scores, but these scores need not be representing what exactly we want measurer.
- A test with high validity has to be reliable also. (the scores will be consistent in both cases)
- Valid test is also a reliable test, but a reliable test may not be a valid one

Different method for determining Reliability

- **Test-retest method**
  - A test is administrated to the same group with short interval. The scores are tabulated and correlation is calculated. The higher the correlation, the more the reliability.
- **Split-half method**
  - The scores of the odd and even items are taken and the correlation between the two sets of scores determined.

- **Parallel form method**
  - Reliability is determined using two equivalent forms of the same test content.
  - These prepared tests are administrated to the same group one after the other.
  - The test forms should be identical with respect to the number of items, content, difficult level etc.
  - Determining the correlation between the two sets of scores obtained by the group in the two tests.
  - If higher the correlation, the more the reliability.

Discriminating Power

- Discriminating power of the test is its power to discriminate between the upper and lower groups who took the test.
- The test should contain different difficulty level of questions.

Practicability

- Practicability of the test depends up on...
- Administrative ease
- Scoring ease
- Interpretative ease
- Economy

Comparability

- A test possesses comparability when scores resulting from its use can be interpreted in terms of a common base that has a natural or accepted meanings
- There are two method for establishing comparability
  - Availability of equivalent (parallel) form of test
  - Availability of adequate norms

Utility

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

1. **Simplicity**

   Simplicity means that the test should be written in a clear, correct and simple language, it is important to keep the method of testing as simple as possible while still testing the skill you intend to test. (Avoid ambiguous questions and ambiguous instructions).

2. **Scorability**

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

4.2 Question

It is necessary for every teacher and evaluator to know : what is the concept of a question; what are the types of questions; what is a good question; how to frame a good question and how does a question bring dynamism in the teaching-learning process. Therefore, whatever the type of test we use in educational testing, the quality of the measuring instrument depends on the quality of the questions it is composed of and their efficient use. Therefore a question is the most significant and indispensable tool in the hands of a teacher, who must use it efficiently both in instruction and for evaluating the students. To visualise its nature, scope and purpose, we may define a *question* as:

(a) verbal or non-verbal tool of 'inquiry framed in a particular context or setting';

(b) based on certain assumptions about learners, learning and requirement of the subject area;

(c) reflecting the intended content elements and the assessment objectives;

(d) worded properly with in-built direction, indicating the nature and scope of expected response;

(e) ensuring maximum scoring objectivity to form appropriate judgments about pupils' learning; and

(f) aiming at getting needed feedback for taking right decisions about examiner and instruction.

4.3 Characteristics of a Good Question

Basic source of all questions is the syllabus prescribed (curriculum, content). However, a question can also be set independent of a syllabus. But every question as its roots in some content area of achievement. One has to think whether that question has to be used for instructional purpose or for evaluation purpose. Thus a question can be developmental when used as a teaching device, or it can be judgmental when used as a testing device. Focus of both is on measurement of instructional objective. Not going into further details about the sources, we turn now to the qualities of a good question that a teacher uses as a measuring device in achievement testing.

(*i*) Question Should be Objective Centred

Instructional objectives are prescribed in the curriculum of every subject. It is these objectives that become the focus of instruction. It is assumed that these objectives are formulated, specified and stated properly in terms of learning competencies. Therefore every question must be objective based or competency based.

(a) A question can cover one (K) or two (K + U or U + A) objectives at a time depending on the purpose.

Example

(*i*) State Archimedes principle. (K)

(*ii*) State Archimedes principle and illustrate with the help of examples. (K + U)

(b) A question may be based on a single or more than one competencies.

Example

(*i*) List the conditions necessary for germination of a seed. (Recall)

(*ii*) Observe the given population graph. What does this graph indicate about population ? (Observe + interpret)

Whether one or more than one objectives are to be covered (Observe + interpret); whether one or more competencies within the same objective or different objectives are to be used in a question depends on the purpose of the test and the use to which the evidence is to be put. However, care may be taken to ensure adequate coverage of competencies in a test (not possible in the same question)

(*ii*) It Must be Context and Purpose Specific

Each question is framed in a particular context and setting. It may be for use in a unit test, an annual test, a review exercise of a text-book chapter, a selection test, or for developing institutional question bank, as home assignment, a diagnostic test etc. Each situation has different purpose to serve; it may be to assess the pupils' learning, identifying their weaknesses, reinforcement of what is learnt, ranking of students, standard setting, comparing students' performance etc. Accordingly, the question is to be framed as individual question or a part of a set of questions used for a test. Cognizance of context and purpose is therefore the first requisite for framing a question.

(*iii*) It Must Ensure Adequate Coverage of Specified Content

When a question is framed in terms of objectives, content elements form the basis or medium.

Therefore a question must reflect the coverage of the content elements like facts (F), concepts (C) and principles (P) that constitute the content or the subject matter. Idea is that the question must be based on pre-decided content elements covering one or more content elements. Of course, for this the content analysis is necessary. Since the content elements like terms, facts, concepts, principles, processes and generalisations are hierarchical in nature, care has to be taken that the higher-level content elements are more amenable to higher-order objectives. The more the content coverage of a unit of testing, the more valid the question is.

Example

Q1. List one chemical property of carbon dioxide. (Limited coverage)

Q 2. List one chemical property of carbon dioxide that differs from that of oxygen. (More coverage)

Notes   Thus a question is indeed the heart and soul of teaching learning and testing on which the teachers and evaluators should pay more attention for effective instruction and evaluation of students. Attempt is made here to highlight these different aspects related to attributes of a question and its conceptual framework

---

(*iv*) It Should Evoke Intended Response

This refers to the hidden face of the question, which indicates the objective, content and form dimension. Question should correspond exactly to these three attributes represented in the hidden or constitutional face. When the question is framed to reflect these three dimensions appropriately, it is sure to evoke the intended responses in terms of relevant competencies, content elements and the form of questions. It is likely to yield quite valid evidence.

(*v*) It Should Generate Responses at Acceptable Level of Performance

Every question must yield responses that can be accepled in terms of its criterion. If one is interested in judging the students' ability to differentiate, interpret, analyse or predict, the question should be based on that particular ability so that the responses given by the students relate to that very ability. Suppose we are interested in judging the ability to interpret which of the two questions given below reflect the acceptable level :

 (*i*) Convert the given population data (of 10 years) in the form of a graph. (No)

 (*ii*) Observe the given graph showing population growth over the years. What does this graph indicate ? (Yes)

The first is at lower level of performance-testing ability to translate, whereas the second is set at the desired level to assess the students' ability to interpret. Likewise it also relates to the hierarchical level of content elements, which are analysed from simple (terms and facts) to more and more complex (concepts, principles, themes etc.) forms; for example, if one is testing the concept of osmosis and puts the following questions :

 (*i*) Define osmosis. (Fact) — lowest level

 (*ii*) Give one example of osmosis from daily life. (Concept) — higher level than (i)

 (*iii*) Why do potato chips wilt when kept in salt water ? (Principle) —still higher level

Idea is that what is the acceptable level in a given context. In mastery learning approach, question like the following indicates more clearly the acceptable level.

Q. Pupil should be able to solve 8 out of 10 additional sums correctly in 30 minutes.

(*vi*) A Question Should Ensure Congruence between Process and Product Responses

When a question is put to students it is expected that the intended response is generated. This is called the product response. In objective questions the key is the product response. In the

questions that demand free or open-ended responses, as in essay-type or short-answer questions, the product responses are their answers— rather expected answers or acceptable answers. However, to arrive at a correct, good, desired or acceptable answer it is essential for the examinee to use the same mental processes or exercises. The same abilities are required to write the correct response, i.e. the product response. If the product response (as intended) is produced correctly without the use of this relevant mental process, it is not a good question. Look at the following two questions tried by the writer himself, used for testing students' ability to apply. (AP objective)

   (a)   Why does an iron needle sink in water and a ship float ?

   (b)   An empty football bladder sinks in water but when filled with air (which has weight) it floats, why ?

      (*i*)   Both these questions when administered to the same group of pupils, it was found that the first question was attempted by more than 80% students correctly but the second hardly by about 20%. Why ? The first question, though of application, was given in the prescribed text-book exercise taught in the class, but the students could attempt correctly (product response) by just recalling or relating, without exercising the relevant mental process involved (analysing, reasoning etc.). In the second question, which involved unfamiliar situation, only the students who used the related mental process could elicit the intended product response. Therefore a good question must ensure this congruence of product and process response.

Did u know? When a teacher or an evaluator constructs a question, he puts it on the paper in different forms : (a) an interrogative sentence; (b) a statement followed by a directional word like why justify etc.; (c) only selection of correct response out of 3 or 4 given alternatives as in objective tests; and (d) just to draw a diagram, read a thermometer or solve a sum.

4.4 Genotypic Varieties of a Question

The hidden or constitutional make-up of every question has three major dimensions, viz (a) the assessment objectives, (b) the content elements, and (c) the form of questions that forms the basis of written questions. For illustration of common usable type the nature and scope of each of these dimensions, is specified first.

1.   Commonly used assessment objectives are : (i) Knowledge (K), (ii) Understanding or comprehension (U) and (iii) Application (A).

2.   Commonly used content elements are (i) Facts (F), (ii) Concepts (C) and Principle

3.   Commonly used forms of questions are : (i) Objective type (O), Short-answer type (S) and Essay type (E).

From this we can develop configuration of $(3 \times 3 \times 3)$, i.e. 27 types of questions with different bases or genotypes. This would lead to 9 varieties of knowledge questions, 9 of understanding and 9 of application objective. Each variety reflects different (1 out of 3) content elements and different (1 out of 3) forms of questions, as depicted in tabular form :

Knowledge based			Understanding based			Application based		
1.	K F O		4.	U F O		7.	A F O	
	K F S			U F S			A F S	
	K F E			U F E			A F E	

2.	K	C	O	5.	U	C	O		A	C	O
	K	C	S		U	C	S		A	C	S
	K	C	E		U	C	E		A	C	E
3.	K	P	O	6.	U	P	O		A	P	O
	K	P	S		U	P	S		A	P	S
	K	P	E		U	P	E		A	P	E

All the 27 types can be classified into three groups, ranging from easy (K) to difficult (A). However, it may be admitted that certain types of questions are more difficult to construct than others because some forms of questions are more amenable to testing of a particular objective in relation to a given content element. Similarly, it is more difficult, sometimes impossible to develop, higher-level objective (A) using lower-level content element (F). Still, in other cases some forms cannot be used to test particular content element efficiently. Thus out of 9 sets of questions, as we proceed from 1st to ninth set it requires more expertise to develop than others. Questions in 9th set are the most difficult to develop and those of 1st set are the most easy to construct. The first question is based on *fact testing knowledge* objective using *objective* form, whereas the last question in set 9 is based on content element *principle* and of *essay-type* testing *application* objective. If one tries to frame question on each of the 27 questions, it would be possible to judge which of them are quite difficult to handle.

Task   What do you understand by usability of test ?

Self Assessment

1. *Fill in the blanks :*

   (*i*) A test is said to be _____ if it is free from personal biases in interpreting its scope as well as in scoring the responses.

   (*ii*) If the item is the test constitute a representative sample of the total _____ to be tested, the test have _____ .

   (*iii*) _____ of a test refers to the degree of consistency with which it measures what it indented to measure.

   (*iv*) _____ of the question is very important in paper pencil tests.

   (*v*) In essay type questions there is a greater problem of _____ .

   (*vi*) Every question has three major dimensions viz the assessment objectives, content elements and _____ .

4.5 Summary

- A good test should possess the following qualities.

- Objectivity

- A test is said to be objective if it is free from personal biases in interpreting its scope as well as in scoring the responses.

- Objective Basedness

- The test should be based on pre-determined objectives.

- The test setter should have definite idea about the objective behind each item.

- Comprehensiveness

Notes
- The test should cover the whole syllabus.
- Validity
- A said to be valid if it measures what it intends to measure.
- *There are different types of validity* :
  - Operational validity
  - Predictive validity
  - Content validity
  - Construct validity
- Reliability
- Reliability of a test refers to the degree of consistency with which it measures what it indented to measure.
- A test may be reliable but need not be valid. This is because it may yield consistent scores, but these scores need not be representing what exactly we warn xo measurer.
- Different method for determining Reliability
- Test-retest method
- Split-half method
- Parallel form method
- Discriminating Power
- Practicability
- Comparability
- Utility
- Basic source of all questions is the syllabus prescribed (curriculum, content). However, a question can also be set independent of a syllabus. But every question as its roots in some content area of achievement. One has to think whether that question has to be used for instructional purpose or for evaluation purpose.
  - (*i*) Question Should be Objective Centred
  - (*ii*) It Must be Context and Purpose Specific
  - (*iii*) It Must Ensure Adequate Coverage of Specified Content
  - (*iv*) It Should Evoke Intended Response
  - (*v*) It Should Generate Responses at Acceptable Level of Performance
  - (*vi*) A Question Should Ensure Congruence between Process and Product Responses
  - (*vii*) It Must be Worded in Proper Language
  - (*viii*) It Must Yield Appropriate Indices of Difficulty (Facility Level) and of Discrimination
  - (*ix*) Every Question Should be Time and Cost Effective
  - (*x*) It Should Ensure Maximum Scoring Objectivity
- Genotypic Varieties of a question
- The hidden or constitutional make-up of every question has three major dimensions, viz (a) the assessment objectives, (b) the content elements, and (c) the form of questions that forms the basis of written questions. For illustration of common usable type the nature and scope of each of these dimensions, is specified first.
1. Commonly used assessment objectives are : (i) Knowledge (K), (ii) Understanding or comprehension (U) and (iii) Application (A).

2. Commonly used content elements are (i) Facts (F), (ii) Concepts (C) and Principle

3. Commonly used forms of questions are : (i) Objective type (O), Short-answer type (S) and Essay type (E).

4.6 Keywords

Comprehensiveness   :   including all items, details fact information etc.

Scorability   :   The ability to achieve more and more scores

Utility   :   The quality of being useful

4.7 Review Questions

1. What are the characteristics of a good test ?

2. What is the meaning of question, explain its characteristics ?

3. What are the genotypic varieties of a question ?

Answers : Self Assessment

1. (*i*) Objective   (*ii*) Course Content, Content Validity

   (*iii*) Reliability   (*iv*) Proper wording

   (*v*) Subject marking   (*vi*) Form of question

4.8 Further Readings

Books

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 5 : Planning for Different Types of Test

CONTENTS

Objectives

Introduction

Objectives

The objectives of this unit can be summarised as below :

- To discuss about the types of standardized test for school age
- To explain about the planning procedure of a test
- Different guidelines for planning a test
- To describe about the guidelines for item writing
- To discuss about the planning for multiple choice test
- To describe about the planning for essay questions
- To explain about the planning for diagnostic test

Introduction

We called an instruction success when the objectives of this instruction have been caught. After an instruction has been done, teachers should know weather its process success or not by measuring knowledge, capability and achievement of their students. Usually, teachers give some achievement test. A test ables to measure student's achievement based on the learning objectives and learning competencies. Something good must be prepared well, so a good achievement test must be prepared well by the teachers. We will discuss about the planning for different types of test.

5.1 Types of Standardized Tests for School Age

Parents always wonder if their child might have gifted and talented mind. They also wonder how their child stacks up to the competition. From early education through high school, we constantly measure academic progress. Standardized testing, in many forms, is the most common way of measuring progress and intelligence. These tests are as follows:-

(*i*) **Intelligence test**

- Intelligence tests are standardized tests that aim to determine how a person can handle problem solving using higher level cognitive thinking. Often just called an IQ test for common use, a typical IQ test asks problems involving pattern recognition and logical reasoning. It then takes into account the time needed and how many questions the person completes correctly, with penalties for guessing. Specific tests and how the results are used change from district to district but intelligence testing is common during the early years of schooling.

(*ii*) **Academic Progress**

- Standardized testing in schools for academic progress and intelligence are not the same, although they use similar questions and methodologies. Academic progress tests such as the Iowa Basic Skills Test give schools an idea of how their students perform on a national level in core areas and how well the school has taught certain subjects. While intelligence tests are often used for gifted and talented programs, academic progress tests usually identify poor performance among students and the effectiveness of teaching.

(*iii*) **College Entrace Exams**

- Colleges often require results from a standardized test, such as the SAT or ACT, to measure college readiness. College entrance exams are similar to other academic progress exams but require a higher level of reading and mathematics. The SAT and ACT allow colleges to measure the aptitude of different applicants, instead of having to compare the scores of many tests, classes and grades from different schools.

5.2 Planning Procedure of a Test

Once the teacher or the test constructor is aware of the characteristics that a good test must possess, s/he can proceed to construct a test, which may be either a unit test or a full-fledged question paper covering all the aspects of the syllabus. Planning for every type of test is almost same, Whether the test is a unit test for use in classroom testing or a question paper for use in final examinations, the steps of test construction are the same, which are as follows :

5.2.1 Prepare a Design

The first step in preparing a test is to construct a design. A test is not merely a collection of assorted questions. To he of any effective use, it has to be planned in advance keeping in view objectives and the content of the course and the forms of questions to be used for testing these. For this weightage to different objectives, different areas or content, and different forms of questions are to be decided, along with the **scheme of options** and sections, and these are the dimensions which are known as a **design of a test**.

5.2.2 Weightage to Objectives

To make a test valid, it is necessary to analyze the objectives of the course and decide which objectives are to be tested and in what properties. For this marks are allotted to each objective to be tested according to its importance. In English language testing the three major objectives are knowledge of the elements of language, comprehension and expression. The **weightages** to all these three objectives may be decided in percentages, For example for a test of 50 marks the following **weightages** may be decided.

Objective	Percentage of Marks	Marks allotted
Knowledge;.	10%	5
Comprehension	40%	20
Expression	50%	25
Total	**100%**	**50**

5.2.3 Weightage to different areas of Content

It is necessary to analyze the syllabus and allot weightages to different areas of content. This is again done to endure the validity of the test. A hypothetical example is given below for an English language test showing **weightages** to content units for a class XI test,

Content Area	Percentage of Marks	Marks allotted
Reading skills	30%	15
Writing skills	30%	15
Textual content	40%	20
Total	**100%**	**50**

5.2.4 Weightage to different forms of Questions

After analyzing the objectives and the content, it is to be seen how they are to be tested. A particular objective and content can be tested more appropriately by a particular form of questions. So, different forms of questions are to be included in the test for testing different objectives and contents. For this a number of different types of questions to be included in the test and the marks carried by each of them are decided. This takes care of the reliability of test.

As an illustration, hypothetical weightage to different forms of questions in our 50 marks question paper for class XI is given below :

Forms of Questions	No. of Questions	Marks allotted	% of marks
Essay type	3	19	38%
Short answer type	9	23	56%
Very short answer type	8	8	16%
	20	50	100%

5.2.5 Scheme of Sections

The design of a question paper may also indicate the scheme of sections for the paper. For example, a question paper may consist of both multiple choice questions and supply type questions. Such a test may have two sections, one consisting of multiple choice questions and the other consisting of supply type questions like essay type, short answer and very short answer type questions. In case the examiner wants, the question paper can also be divided into sections areawise like one section for reading comprehension, another for writing tasks, third for grammar and so on. If the multiple choice questions are not substantial in number, there is no need to keep a separate section.

5.2.6 Scheme of Options

The design may indicate the pattern of options i.e. the complete elimination of overall options or retention of internal options within limits. No options are to be provided in case of multiple choice, short answer and very short answer questions; for essay type questions the teacher may like to provide internal options. While providing options, it may be kept in mind that the options are comparable in terms of objectives to be tested, the form of questions and the difficulty level of the questions. As far as possible, the major area of content should also be the same in the options.

While planning the paper. It should be so planned that the difficulty level of the questions varies so as to cater to all the students of the class and also to discriminate between high achievers and low achievers. The suggested percentage for easy and difficult questions is 20% whereas average questions can be 60%. The difficulty level of the test paper can be varied according to the level of the students. If the class has a large number of good students, then 25% to 30% difficult questions can be given.

5.2.7 Preparing a Blue Print

After deciding on the design of the test, the blue print is prepared. The blueprint is a three dimensional chart which shows the placement of each question in respect of the objective and the content area that it tests. It also indicates the marks carried by each question. It is useful to prepare a blue print so that the test maker knows which question will test which objective and which content unit and how many marks it would carry. Without a blue print only the weightage are decided for objectives, content areas and types of questions. The blue print concretizes the design in operational terms and all the dimensions of a question (i.e. its objective, its form, the content area it would cover and the marks allotted to it) become clear to the test maker.

There is no set procedure for preparing a blue print. However, the following sequential steps would help in preparing a good blue print.

1. Transfer the decisions regarding weightages to objectives - Knowledge, Comprehension and Expression on the given proforma.

2. Transfer the weightages already decided for different content units. For this, list the content units under the content areas in the column given at the left hand and the marks under the column of total given at the right hand side.

3. Place the essay type questions first in the blue print. Place them under the objectives which you want to test through these questions. The marks of the questions may be shown in the column under the objectives and the number of questions may be given in brackets.

4. If in a question, marks are to be split between two objectives indicate it with asterisks and a dotted line as shown in the example.

5. After placing the essay type questions, place the short answer type questions under the objectives and beside the content unit that you want to test through them.

6. Place the very short answer type questions in a similar way.

7. Place the multiple choice questions in the same way - marks outside the bracket, number of questions inside the bracket.

8. Calculate the subtotals of all the questions under all the objectives.

9. Calculate the totals. Your total should tally with the weightages of objectives and content units that you had already marked on the blue print. Fill in the summary of types of questions, Scheme of Sections and Scheme of Options.

5.2.8 Prepare questions based on the blue print

After the blue print is ready, questions are to be prepared according to the dimensions defined in the blueprint. For example, if there are essay type questions to be prepared to test the writing skills, one letter and one report and also a short answer question on writing a notice, the test constructor should prepare these three questions along with their options which may be comparable in terms of objectives to be tested, content areas, forms of questions and the difficulty level.

While preparing questions it must be kept in mind that the question :

1. is based on the specific objective of teaching as indicated in the blue print

2. relates to the specific content area as per the blue print

3. is written in the form as required by the blue print and satisfies all the rules for framing that form of questions

4. is at the desired level of difficulty

5. is written in clear, correct and precise language which is well within the comprehension of pupils

6. clearly indicates the scope and length of the answer.

Another thing to be kept in view while writing questions is to prepare the answers simultaneously because quite often the answers help in refining the questions.

5.2.9 Assembling the Question Paper

After the questions are prepared, they are to be assembled in a question paper form. For this, instructions are to be written. General instructions for the paper may be given on top whereas instructions for specific questions may be given just before the questions.

The order of questions is also to be decided while assembling the question paper. Sometimes it is according to the forms of questions, i.e., objective type questions may be put first, then very short answer, short answer and essay type questions or it may be according to the content as in the case of a language question paper where we may have structure questions first, then questions on unseen passage and then composition questions.

The assembling and editing of the question paper is important from the point of view of administration. For example, if the question is divided into two sections, one of which is to be collected within a specific time limit, clear instructions to do so should be mentioned and also the arrangement of questions should be such that both the sections are easily demarcated.

5.2.10 Preparing the Scoring Key and the Marking Scheme

Scoring key is to be prepared for objective type questions and the marking scheme for other questions.

The scoring key gives the alphabet of the correct answer and the marks carried by each question. The marking scheme gives the expected outline answer and the value points for each aspect of the answer.

Detailed instructions for marking are also worked out, e.g., in marking compositions, etc. It is specified as to how many marks are to be deducted for spelling mistakes or structural mistakes, or if the composition is to be graded, how it is to be done and on what basis.

The detailed marking scheme is necessary to ensure consistency and uniformity in scoring by different examiners. In other words it ensures reliability of scoring.

5.2.11 Preparing Question-wise Analysis

After the question paper and marking scheme are finished, it is desirable to prepare a question-wise analysis. This analysis helps in tallying the questions in the test with the blue print, It also enables us to know the strengths and weaknesses of the test better, eg., through the analysis we can know how many topics have been covered in the syllabus, what is the difficulty level of each question and what specifications are being tested by each question. The analysis is done on following points :

(*i*)   Number of the question.

(*ii*)   Objective tested by the question.

(*iii*)   Specification on which the question is based.

(*iv*)   Topic covered.

(*v*)   Form of the question.

(*vi*)   Marks allotted.

(*vii*)   Approximate time required for answering.

(*viii*) Estimated difficulty level.

Did u know?   Standardized testing has received criticism from psychologists, educators and parents. Criticism of academic testing often focuses on linguistic biases against minorities, the testing methods that may not work for all types of students and negative reinforcement of lower performing students.

The first recorded use of standardised testing occured in France at the and of the 19ᵗʰ century and the start of 20ᵗʰ century. France began to require all students attend some school and educators and administratives fore saw the problem of special education. To assess the abilities of children, the French Government Commissioned.

Therefore Simon and Affred Bient to study this issue. They came up with what we now call an "IQ Test".

5.3 Different Guidelines for Planning a Test

There are various guides that are helpful is focussing attention beyond the knowledge level. They are as follows :

5.3.1 Bloom Taxonomy

The learning competencies measured by a test should faithfully reflect the objectives of instruction. Thus requires the identification of those instructional objectives to be measured by a test and then clearly stating them in performance terms.

One useful guide is the cognitive domain of the *Taxonomy of Educational Objectives,* also known as *Bloom Taxonomy.* This is an older guide one. It contains range types of learning competencies to consider when planning a test. The cognitive domain involves knowledge and the development of intellectual skills. This includes the recall or recognition of specific facts, procedural patterns, and concepts that serve in the development of intellectual abilities and skills. There are six major categories as follows :

Category	Description	Keywords (verb)
Knowledge	Recall data or information	defines, describes, identifies, knows, labels, lists, matches, names, outlines, recalls, recognizes, reproduces, selects, states
Comprehension	Understand the meaning, translation, interpolation, and interpretation of instructions and problems. State a problem in one's own words	comprehends, converts, defends, distinguishes, estimates, explains, extends, generalizes, gives an example, infers, interprets, paraphrases, predicts, rewrites, summarizes, translates
Application	Use a concept in a new situation or unprompted use of an abstraction. Applies what was learned in the classroom into novel situations in the work place	applies, changes, computes, constructs, demonstrates, discovers, manipulates, modifies, operates, predicts, prepares, produces, relates, shows, solves uses
Analysis	Separates material or concepts into component parts so that its organizational structure may be understood. Distinguishes between facts and inferences	analyzes, breaks down, compares, contrasts, diagrams, deconstructs, differentiates, discriminates, distinguishes, identifies, illustrates, infers, outlines, relates, selects, separate

Synthesis	Builds a structure or pattern from diverse elements. Put parts together to form a whole, with emphasis on creating a new meaning or structure	categorizes, combines, complies, composes, creates, devises, designs, explains, generates,modifies, organizes, plans, rearranges, reconstructs, relates, reorganizes, revises, rewrites, summarizes, tells, writes
Evaluation	Make judgments about the value of ideas or material	appraises, compares, concludes, contrasts, criticizes, critiques, defends, describes, discriminates, evaluates, explains, interprets, justifies, relates, summarizes, supports

The categories which are listed in order above, starting from the simplest behaviour to the most complex. The categories can be thought of as degrees of difficulties. That is, the first one must be mastered before the next one can take place.

5.3.2

Another guide one is *Quellmalz Framework* which is similar to the *Bloom Taxonomy*. It represents a kind modified version of the *Bloom Taxonomy*, simpler and easier to use. The descriptions within each category are followed by an indication how the categories relate the *Bloom Taxonomy*.

Category	Description	Keywords (verb)	Bloom's Categories
Recall	Remembering or recognizing key facts, definitions, concepts, etc.; repeating verbatim or paraphrasing information that has already been provided to the student	Define, list, label, name, identify, repeat, who, what, when	Knowledge Comprehension
Analysis	Understanding relationships between the whole and its component parts and between cause and effect; sorting and categorizing; understanding how things work and how the parts of something fit together; understanding causal relationships; getting information from charts, graphs, diagrams, and maps.	Analyze, break down, relationship, how it works, how it's used, give an example	Analysis
Comparison	Explaining how things are similar and how they are different.	Compare, contrast, distinguish, alike different	Analysis

Inference	Reasoning inductively or deductively.	Hypothesize, synthesize, use evidence, apply a rule, generalize, create what if, infer, predict, conclude, apply, solve	Application Synthesis
Evaluation	Expressing and defending an opinion. Evaluation tasks require students to judge quality, credibility, worth or practicality using established criteria and explain how the criteria are met or not met	Judge, evaluate, best solution, justify, defend, critique defend	Synthesis Evaluation

The learning competencies to be measured by a test are most useful in test construction when they are stated as a terminal performance that is observable, include only objectives that can be tested and they are stated as general competencies.

This list of general competencies could, of course, be expanded by making the statements more specific, and in some cases it may be desirable to do so. The number of general learning outcomes to use is somewhat arbitrary. Typically, a shorter list is satisfactory for a unit of study, while a more comprehensive list is needed for summative testing at the end of a course.

When a satisfactory list of general learning competencies has been identified an clearly stated, the next step is to list the specific types of student performance that are to be accepted as evidence that the competencies have been achieved. The terms used to describe the specific learning competencies indicate student performance that can be demonstrated to an outside observer.

Action verbs should indicate precisely what the student is able to do to demonstrate achievement. Such vague and indefinite terms as "learns," "sees," "realizes," and "is familiar with" should be avoided, since they do not clearly indicate the terminal performance to be measured.

In defining the general learning competencies in specific performance terms, it is typically impossible to list all of the relevant types of performance. The proportion that need be listed depends to a large extent on the nature of the test.

5.4 Guidelines for Item Writing

There are a number of general suggestions that apply to the writing of all item types. These provide a general framework for writing items that function as intended and that contribute to more valid and reliable results.

1. *Select the type of test item that measures the intended learning competence most directly*. Use a supply-type item if supplying the answer is an important element of the task (e.g., writing). Use a selection-type item if appropriate, (e.g., identification) or if both types are equally appropriate.

2. *Write the test item so that the performance it elicits matches the* performance *in the learning task*. The intended learning competence specifies the learning task in performance terms and the test task should call forth the same performance.

3. *Writing the test item so that the test task is clear and definite*. Keep the reading level low, use simple and direct language, and follow the rules for correct punctuation and grammar.

4. *Write the test item so that it is free from nonfunctional material*. Material not directly relevant to the problem being presented increases the reading load and may detract from the intent of the item. Use extraneous material only where its detection is part of the task (e.g., in math problems).

5. *Write the test item so that irrelevant factors do not prevent an informed student from responding correctly*. Avoid trick questions that might cause a knowledgeable student to focus on the wrong aspect of the task. Use clear, unambiguous statements that maximize the performance to be measured and minimize all other influences. For example, word problems measuring mathematical reasoning should keep reading level and computational demands simple if an uncontaminated measure of reasoning ability is desired.

6. *Write the test item so that irrelevant clues do not enable the uninformed student to respond correctly*. Removing unwanted dues from test items requires alertness during item writing and reviewing the items after setting them aside for a while. The most common clues for each item type will be considered in the following chapters. It *is* also important to prevent the information given in one item from providing an answer to another item in the test.

7. *Write the test item so that the difficulty level matches the intent of the learning competence, the age group to be tested, and the use to be made of the results*. When difficulty is altered to obtain a wide spread of test scores, check to be certain that the added difficulty is relevant to the intended learning competence and that the item is free from sources of irrelevant difficulty (e.g., obscure materials, overly fine discriminations).

8. *Write the test item so that there is no disagreement concerning answer*. Typically the answer should be one that experts would agree is to correct or best answer. Most problems arise here when students are to provide the best answer (best procedure, best explanation). This involves a rm ter of judgment and to be defensible the answer must be clearly best identified as such by experts in the area. Where experts disagree, it may 1 desirable to ask what a particular authority would consider to be the be method, the best reason, and the like. When attributed to a source, the answer, can be judged as correct or incorrect.

9. *Write the test items for enough in advance that they can be later reviewed and modified as needed*. A good time to write test items is shortly after the material has been taught, while the questions and context are still clearl in mind. In any event, reviewing and editing items after they have been SE aside for a while can detect flaws that were inadvertently introduced during the original item writing.

10. *Write more test items than called for by the test plan*. This will enable you to discard weak or inappropriate items during item review, and make easier to match the final set of items to the test specifications.

Self Assessment

1. Fill in the blanks :

 (*i*) _____ are standardized tests that aim to determine person's _____ level.

 (*ii*) The first step in planning a test is Construct a _____ .

 (*iii*) The _____ is a three dimensional chart which shows the placement of each question in respect of objective and the contact area.

 (*iv*) Taxonomy of Educational objectives is also known as _____ .

5.5 Planning for Multiple Choice Test

A content valid test measures what it is intended to measure — the content as represented by the test blueprint. A reliable test yields the same scores over repeated administrations. A test that is not reliable cannot be valid. To ensure that tests are both valid and reliable, adequate planning and

careful item construction are necessary. In this section the construction of test questions is discussed. Guidelines for constructing test items are reviewed, and different item formats are presented.

The problems with most test items relate to reading load, clues embedded in the questions or ambiguity in the phrasing of test items. The use of unnecessarily difficult vocabulary can confuse students who otherwise could respond adequately. Tests with poorly constructed items are not accurate measures of what students know and may not rank students in the same way as a well constructed examination. The following guidelines provide examples of items with common flaws.

I.   Reducing Errors Due to Reading Difficulty or Reading Speed

• **Avoid writing items with an unnecessarily difficult vocaobulary**.	
Poor The promiscuous use of sprays, oils, and antiseptics in the nose during acute colds is a pernicious practice because it may have a deleterious effect on A. the sinuses. B. red blood cells. C. white blood cells.	**Better** Frequent use of sprays, oils and antiseptics in the nose during acute colds may results in A. spreading the infection to the sinuses. B. damage to the olfactory nerve. C. destruction of white blodd cells.

• **Avoid repeating words.**	
Poor Entomology is A. the study of birds. B. the study of fish. C. the study of insects.	**Better** Entomology is the study of A. birds. B. fish. C. insects.

Notes   What is Quellmalz framework ?

5.6 Planning for Essay Questions

Essay examinations are useful tools to evaluate students' thought processes if they are constructed to elicit more than a rehash of material presented in class. Unfortunately, essay examinations are often subjectively and inconsistently scored. These problems may be related to the construction of the essay items, reader bias, or to the scoring procedure. In this section, guidelines for improving the reliability and validity of essay scores are presented and examples of essay topics are critiqued. The following suggestions help to improve essay examinations :

1.   Frame questions so that the task is clearly defined.

2.   Specify the value and approximate time limit for each question.

3.   Do not employ optional questions, but do use several essay question or a combination of essay, short answer, and multiple choice questions.

4.   Write a trial response to each essay question.

5. Prepare a tentative scoring key in advance.

6. Alert students if penalties for spelling or grammar are given.

7. Evaluate essays anonymously by mixing papers and masking names.

8. Scan longer essays to presort into stacks of high, medium and low quality and then reread to assign marks.

9. Score all answers to one question before scoring the next question.

10. Decide whether essays are to be scored holistically or analytically. Holistic scores are awarded as a global measure of overall quality, i.e. a single mark is given on a 1 - 5 (A - E) or perhaps a 10 or 12 point scale. An analytic score is the sum of several different ratings of features about an essay. An example of analytic scoring is found on page 29.

When a combination of essay and multiple choice items are included in a test, the calculation of the exam score can be handled in several ways. Perhaps the easiest method is to assign weights to each section of the test. If the essay portion is worth 50 percent and the multiple choice worth 50 percent, the essay might be rated on a scale from 0 to 10 points, either holistically or analytically. The total number of essay points could be multiplied by 5 making a range of 0 to 50 possible points for the essay. The difficulty in using a single scale from 0 to 50 or more points to initially rate essays is that it is extremely difficult for a reader to reliably differentiate papers that receive, for example, 36 or 38 points. A second reading might well reverse the two scores.

The following discussion about essay examinations is particularly helpful in evaluating essay questions. It is adapted from a bulletin published at Michigan State University.

The Writing of Essay Questions

It is commonly recognized that the reading and marking of essay examinations is a difficult and time-consuming process; it is not always recognized that the preparation of essay questions requires time and effort. The weaknesses which have frequently characterized the essay question have, in part, resulted from a lack of awareness of or concern about these difficulties. Overemphasis on factual information and on memorized extracts of textbooks and lectures, ambiguity in meaning and an associated ambiguity of response, and variations in the standards used to judge student performance result in difficulties in reading and in unreliable marking. Essay questions can be markedly improved by conscious attention to the identification and elimination of these and other deficiencies.

Perhaps the most general suggestion for improving essay testing is that the task set for the student should first be clearly understood by the teacher and then clearly presented to the student. It is unreasonable to present the student with the combined tasks of reading the instructor's intent in the question and meeting his expectation as to the answer. On the other hand, to make an essay question highly specific and detailed is to reduce it to a type of objective test item involving a filling in of blanks implied in the question. The poser of essay questions must tread the path between highly general questions to which no adequate answer can be given in the time permitted—even assuming the meaning is clear—and highly specific questions which involve word or phrases answers. The question must state precisely what is wanted, and it should contain enough clues that students can readily infer the direction their answer should take. It is also desirable that the question should state or imply the approximate length and complexity of the desired response. Some examples may serve to clarify this and other points.

5.7 Planning for Diagnostic Testing

The essential steps in educational diagnosis are :

(*i*)   Identifying the students who are having trouble or need help.

(*ii*) Locating the errors or learning difficulties.

(*iii*) Discovering the causal factors of slow learning.

(*i*) Identifying the students who are having trouble or need help

First, one must know the learners who require help. For this you can administer a general achievement test based on the topics already taught. After evaluation you will be in a position to make lists of students who are below average, average or above average. Next, one has to locate the area where the error occurs in order to have a deeper insight into the pupils' difficulties.

(*ii*) Locating the errors or learning difficulties

After identifying the students who need help and visualising the necessity of additional instructional material to improve the quality of learning, your main role is to find out the area where the learner commits mistakes or which is the area where learning difficulties lie. For example we examine the pupils' responses as follows :

(*i*) $$\frac{3}{4}+\frac{5}{4}=\frac{8}{4}$$ (*ii*) $$\frac{2}{3}+\frac{5}{3}=\frac{7}{3}$$

(*iii*) $$\frac{7}{12}+\frac{4}{12}=\frac{11}{12}$$ (*iv*) $$\frac{4}{13}+\frac{5}{13}=\frac{9}{13}$$

(*v*) $$\frac{3}{4}+\frac{5}{4}=\frac{⑧}{4}$$ (*vi*) $$1+\frac{3}{4}=\frac{④}{4}$$

(*vii*) $$3+\frac{5}{7}=\frac{8}{7}$$ (*viii*) $$5+\frac{1}{7}=\frac{⑥}{7}$$

(*ix*) $$\frac{1}{2}+\frac{1}{3}=\frac{②}{⑤}$$ (*x*) $$\frac{3}{7}+\frac{4}{5}=\frac{⑦}{12}$$

From the above example you would realize that the learner has the knowledge of adding fractions having the same denominator but the concept of addition of fractions having different demoninators is not clear to him even through it was taught in the classroom. Thus you have located the area where the learning difficulty lies.

(*iii*) Discovering the causal factors of slow learning

In some cases of learning difficulties, the causal factors are relatively simple. A student may be inattentive during teaching-learning or may be committing errors due to insufficient practice or irregular attendance.

Sometimes the cause is ill-health or faulty work habits etc. It has also been observed sometimes that the basic cause of low achievement is a feeling of helplessness or the complexity of the subject-matter which perhaps is much above the level of their comprehension.

Sequential presentation in Figure 1 shows how diagnosis leads to improved quality of learning.

Teaching-learning ⟶ Evaluation ⟶ Unsatisfactory performance of pupils

▮

Identify students who need help ⟵ Administer a General Achievement Test or use the data of previous testing ⟵ A case for diagnosis

▮

Locate the area of difficulty ⟶ Identify causal factors ⟶ Provide remedial measures ⟶ Improved quality of learning
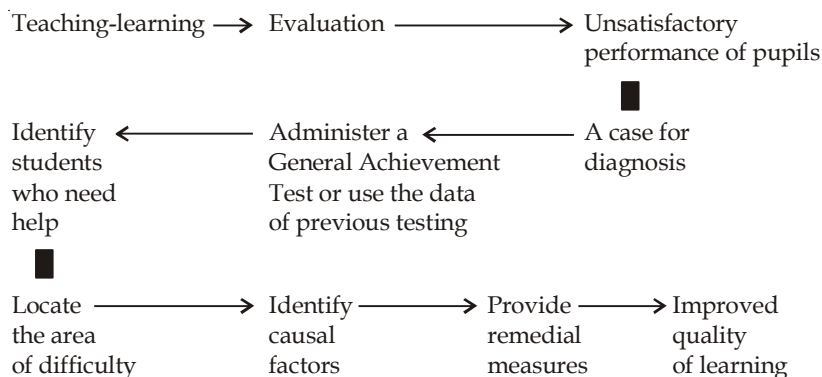
Figure : 1

Self Assessment

State whether the following statements are 'true' or 'false' :

(*i*)   Tests with poorly constructed items are not accurate measures etc for student knowledge.

(*ii*)   Essay examinations are useful tools to evaluate students thought processes.

(*iii*) A student may be attentive during teaching — learning due to insufficient practical or irregular attendance.

5.8 Summary

• From early education through high school, we constantly measure academic progress. Standardized testing, in many forms, is the most common way of measuring progress and intelligence.

• Intelligence Test

  (*i*) Intelligence tests are standardized tests that aim to determine how a person can handle problem solving using higher level cognitive thinking. Often just called an IQ test for common use, a typical IQ test asks problems involving pattern recognition and logical reasoning.

  (*ii*) Academic Progress

   • Standardized testing in schools for academic progress and intelligence are not the same, although they use similar questions and methodologies. Academic progress tests such as the Iowa Basic Skills Test give schools an idea of how their students perform on a national level in core areas and how well the school has taught certain subjects.

  (*iii*) College Entrace Exams

   • Colleges often require results from a standardized test, such as the SAT or ACT, to measure college readiness. College entrance exams are similar to other academic progress exams but require a higher level of reading and mathematics.

• Once the teacher of the test constructor is aware of the characteristics that a good test must possess, s/he can proceed to construct a test, which may be either a unit test or a full-fledged question paper covering all the aspects of the syllabus.

• Prepare a Design

• The first step in preparing a test is to construct a design. A test is not merely a collection of assorted questions. To he of any effective use, it has to be planned in advance keeping in view objectives and the content of the course and the forms of questions to be used for testing these.

• Weightage to Objectives

• To make a test valid, it is necessary to analyze the objectives of the course and decide. which objectives are to be tested and in what properties. For this marks are allotted to each objective to be tested according to its importance.

• Weightage to different areas of Content

• It is necessary to analyze the syllabus and allot weightages to different areas of content. This is again done to endure the validity of the test.

• Weightage to different forms of Questions

• After analyzing the objectives and the content, it is to be seen how they are to be tested. A particular objective and content can be tested more appropriately by a particular form of questions.

• Scheme of Sections

- The design of a question paper may also indicate the scheme of sections for the paper. For example, a question paper may consist of both multiple choice questions and supply type questions.

- Scheme of Options

- The design may indicate the pattern of options i.e. the complete elimination of overall options or retention of internal options within limits. No options are to be provided in case of multiple choice, short answer and very short answer questions.

- After deciding on the design of the test, the blue print is prepared. The blueprint is a three dimensional chart which shows the placement of each question in respect of the objective and the content area that it tests. It also indicates the marks carried by each question. It is useful to prepare a blue print so that the test maker knows which question will test which objective and which content unit and how many marks it would carry.

- After the blue print is ready, questions are to be prepared according to the dimensions defined in the blueprint.

- After the questions are prepared, they are to be assembled in a question paper form. For this, instructions are to be written. General instructions for the paper may be given on top whereas instructions for specific questions may be given just before the questions.

- Scoring key is to be prepared for objective type questions and the marking scheme for other questions.

- The scoring key gives the alphabet of the correct answer and the marks carried by each question. The marking scheme gives the expected outline answer and the value points for each aspect of the answer.

- After the question paper and marking scheme are finished, it is desirable to prepare a question-wise analysis. This analysis helps in tallying the questions in the test with the blue print, It also enables us to know know the strengths and weaknesses of the test better, eg.

- One useful guide is the cognitive domain of the *Taxonomy of Educational Objectives,* also known as *Bloom Taxonomy.* This is an older guide one. It contains range types of learning competencies to consider when planning a test. The cognitive domain involves knowledge and the development of intellectual skills. This includes the recall or recognition of specific facts, procedural patterns, and concepts that serve in the development of intellectual abilities and skills. There are six major categories as follows :

- Knowledge

- Comprehension

- Application

- Analysis

- Synthesis

- Evaluation

- Another guide one is *Quellmalz Framework* which is similar to the *Bloom Taxonomy.* The descriptions within each category are followed by an indications how the categories ralate to the *Bloom Taxonomy.*

- Recall

- Analysis

- Comparison

- Inference

- Evaluation

(*i*) Preparing the Test Specifications

The writing of test items should be guided by a carefully prepared set of test specifications. The function of the specifications is to describe the achievement domain being measured and to provide guidelines for obtaining a representative sample of test tasks.

(*i*) Selecting the learning competencies to be tested. The learning outcomes for a particular course will depend on the specific nature of the course, the objectives attained in previous courses, the philosophy of the school, the special needs of the students, and a host of other local factors that have a bearing of the instructional program.

(*ii*) Outlining the subject matter. The content of a course may be outlined in detail for teaching purposes, but for test planning only the major categories need be listed.

(*iii*) Making a two-way chart. The two-way chart describes the sample of items to be included in the test.

- Preparing a table of specifications includes the following steps :

- Identify the learning outcomes and content areas to be measured by the test.

- Weigh the learning outcomes and content areas in terms of their relative importance.

(*i*) Thus, whether a table of specifications is useful in preparing a formative test depends greatly on the nature of the achievement domain being measured.

(*ii*) The construction of a set of relevant test items is greatly simplified if the intended learning competences have been clearly defined and the test specifications carefully prepared. The quality of the test will then depend on how closely the test maker can match the specifications.

- Selecting the Types of Test Items to Use

- The items used in achievements tests can be classified as either *selection-type* items or supply-type items. The selection-type item presents students with a set of possible responses from which they are to select the most appropriate answer. The supply-type item requires students to create and supply their own answers. These two major categories can be used to classify the most widely used item types as follows.

- *Selection-Type Items*

(*i*) Multiple choice

(*ii*) True-false

(*iii*) Matching

(*iv*) Interpretive exercise

- *Supply-Type Items*

(*i*) Short answer

(*ii*) Essay (restricted response)

(*iii*) Essay (extended resporise)

- There are a number of general suggestions that apply to the writing of all item types. These provide a general framework for writing items that function as intended and that contribute to more valid and reliable results.

(*i*) Select the type of test item that measures the intended learning competence most directly.

(*ii*) Write the test item so that the performance it elicits matches the performance in the learning task.

(*iii*) Writing the test item so that the test task is clear and definite.

(*iv*) Write the test item so that it is free from nonfunctional material.

(*v*) Write the test item so that irrelevant factors do not prevent an informed student from responding correctly.

(*vi*) Write the test item so that irrelevant clues do not enable the uninformed student to respond correctly.

(*vii*) Write the test item so that the difficulty level matches the intent of the learning competence, the age group to be tested, and the use to be made of the results.

(*viii*) Write the test item so that there is no disagreement concerning answer.

(*ix*) Write the test items for enough in advance that they can be later re viewed and modified as needed.

(*x*) Write more test items than called for by the test plan.

- A content valid test measures what it is intended to measure — the content as represented by the test blueprint.

- This form is most effectively used when it is important for students to recall than recognize the correct answer, or as in the case of some mathematics problems, when the instructor wants the student to calculate an answer without the assistance of possible correct answers. For example :

- When the square root of 21 is computed, what number should appear in the second decimal place ?

- These questions are written as completion items or direct questions or statements to which students are directed to respond. Completion items (fill-in the blank) tend to be the most difficult to write clerly. If they are used, the blanks should be near the end of the sentence and the question needs to be precisely stated :

- Essay examinations are useful tools to evaluate students' thought processes if they are constructed to elicit more than a rehash of material presented in class. The following suggestions help to improve essay examinations :

  (*i*) Frame questions so that the task is clearly defined.

  (*ii*) Specify the value and approximate time limit for each question.

  (*iii*) Do not employ optional questions, but do use several essay question or a combination of essay, short answer, and multiple choice questions.

  (*iv*) Write a trial response to each essay question.

  (*v*) Prepare a tentative scoring key in advance.

  (*vi*) Alert students if penalties for spelling or grammar are given.

  (*vii*) Evaluate essays anonymously by mixing papers and masking names.

- The essential steps in educational diagnosis are :

  (*i*) Identifying the students who are having trouble or need help.

  (*ii*) Locating the errors or learning difficulties.

  (*iii*) Discovering the causal factors of slow learning.

5.9 Keywords

Plain	:	Something that somebody intend to do or achieve
Design	:	The general arrangement of the different parts of somethy that is made, like a book, test etc.
Comprehension	:	A exercise that trains students to understand a language

5.10 Review Questions

1   Explain the procedure of planning for different types of test

2.   How does a multiple choice test planned ?

3.   What are general guidelines for Item writing ?

4.   Give the suggestions help to improve examinations.

5.   Explain the essential steps in educational diagnosis test.

Answers : Self Assessment

1.   (*i*)   Intelligence tests, IQ   (*ii*)   design

     (*iii*)   blue print   (*iv*)   Bloom Taxonomy

2.   (*i*)   T   (*ii*)   T   (*iii*)   F

5.11 Further Readings

Books

1.   Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2.   Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3.   Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4.   Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 6 : Validity – Types , Methods and Usability

Objectives

The objectives of this unit can be summarised as below :

- To discuss about the validity
- To describe about the types of validity
- To explain the different methods of validity
- To describe the usability of validity

Introduction

The tests in the field of education are generally constructed to measure some specific qualities or abilities (variables) of the students. If a test measures the qualities or abilities for which it was constructed, then such a test is called valid, and this quality of the test is called validity. For example, take an achievement test. If this test measures the knowledge, skill and sentiments for which it was made, then it will be called a valid test and its this quality will be called validity.

It has been seen that no test comes fully true on the criterion of validity. A test neither has validity fully, nor it is absent altogether. So now it is said that the more a test is able to measure the qualities or abilities for which it is constructed, the more valid it is. **Cronback** has defined it in the following words :

Validity is the extent to which a test measures what it purports to measure.   —*Cronback*

That is the length of the test should be increased 6 times. Thus, it order to increase its validity from .60 to .90, the number of questions will have to be increased from 50 to 300, that is, 250 questions will have to be added to it. We will discuss about validity of test in this unit.

6.1 Validity

6.1.1 Meaning and Nature of Validity

Validity is an extremely essential attribute of a test. By validity is meant that characteristic of a test under which it is understood that it measures that element or attribute for which it has been constructed. The extent to which a test measures what it purports to measure is called its validity.

The following formula will prove useful in remembering validity in brief :

Adequacy + Objectivity $\rightarrow$ Reliability + Aim $\rightarrow$ Validity

For example, if a test has been constructed to measure intelligence, then the extent to which it will measure intelligence will be its validity. On the contrary, if an intelligence test comprises of such questions which arouse his emotions, and he commits error in solving questions, then the influence of emotion in the test will decrease its validity. In the same way, if in an interview for the selection of a post of a psychologist, such questions are asked as : Who is Sachin Tendulkar ? Who is Laden ? What are the objectives of the five-year plans ? What is 10 + 2 + 3 system ? Who is Dhanraj Pillai ? etc. Such a test will have a very low reliability, because it contains those elements which do not measure those factors or qualifications which are necessary for the post of a psychologist.

Psychological tests are constructed keeping many objectives in mind. The extent to which the purpose or objective of a test is realized is called its validity. A test will be valid only from the standpoint of its specific reference and realization of its specific objective. This may be valid in reference to one objective, and it may not be so in reference to another objective; that is, only one test cannot be valid for all objectives at the same time. Therefore, if we have to know how valid is a test, it is also necessary to know the purpose for which it is valid. If it is an IQ test, it cannot be valid for measuring achievement in subjects. It is evident that validity of a test can be ascertained in many ways and in many contexts. Validity is a very important characteristic of a test. If a test is not valid, then the data represented by it will not be correct in the context of its purpose for which it was made.

The following factors should be taken into consideration while evaluating validity of a test :

1. Validity is the characteristic of test results, and not of the test.

2. Validity is not a common characteristic of a test, rather it is used in the measurement of specific abilities.

3. We can see about the validity of a test whether it is highly valid, least valid or ordinarily valid. A test can never be fully valid or fully invalid.

Some famous scholars have defined validity as follows :
1. **"The validity of a test or of any measuring instrument depends upon the fidelity with which it measures what it purports to measure."**   **— Garrett (1959)**

2. "I**n general, the term validity refers to effectiveness of a test in achieving specific purposes."**
                                                                    —Stodola and Stordahi

3. **"In a broad sense, validity has to do with the question of what test scores measure and what they will predict."**   **—Guilford**

4. **"A measurement procedure is valid in so far as it correlates with some measurement of success in the job for which it is being used as a predictor."**   **—Thorndike**

5. **"Validity is defined as the proportion of true variance that is relevant to the purposes of the testing."**   **—Frederick G. Brown**

6. **"Validity is the extent to which a test measures what it purports to measure."** **—Cronback**

7. **"It is the correlation of the test with some criterion."**   **—Gulikson**

8. **"Perhaps the commonest definition given for the term validity is that it refers to the extent to which a test or a set of operation measures what it is supposed to measure."** **—Ghiselli**

If validity is defined on the basis of variance, it becomes clear that the entire variance is formed of two parts of test scores : 1. true variance and 2. error variance; that is :

$$S_X^2 = S_T^2 + S_E^2$$

Where, S_X^2 = Total Variance

S_T^2 = True Variance

S_E^2 = Error Variance

True variance too can be further divided into two parts : (a) valid variance and (b) reliable but irrelevant variance; that is :
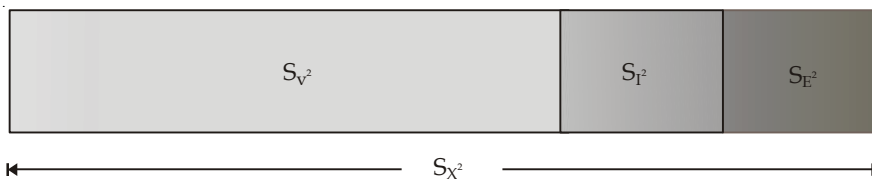
$$S_T^2 = S_V^2 + S_I^2$$

Where,

= True Variance

= Relevant or Valid Variance

= Irrelevant Variance

Thus we can say that $\quad S_X^2 = S_V^2 + S_I^2 + S_W^2$

It can be understood more clearly by the following diagram :

It is clear from the above formula that when error variance is lower, that is, a test is more reliable, then the ratio of valid variance increases, but because the remaining systematic variance can be relevant or irrelevant, so a lower number of errors does not mean that its validity will be higher. This is reason that it is said : **"Reliability is a necessary, but not a sufficient condition for validity."**

6.1.2 Criterion for Validity

An independent criterion is needed for determining validity. In mental tests, the measurable elements cannot be seen in their concrete forms, they can be perceived only through the medium in the form of activities. Electricity cannot be seen, but it can be known by a lighted bulb or running machine. In the same manner, intelligence cannot be seen, but it can be found out by the activities of an individual. If an individual lacks intelligence, he cannot understand or perform even small activities; and an intelligence individual can perform even the most difficult tasks. Therefore, in order to establish validity of a test, some such activity will have to be found out from the external environment in which the measurable element is manifested.

6.2 Types of Validity

According to the **National Committee on Test Standards**, validity of a test can generally by done by the following three types of validity :

1. Content or curricular validity,

2. Criterion-related validity, and

3. Construct validity.

1. **Content or Curricular Validity :** By content validity of a test we mean each item in the test should be a sample of that knowledge and performance for which it has been constructed. Content validity may be defined as the extent to which a test measures a representative sample of the subject matter and the behavioural changes under consideration.

    According to **Anastasi, "Content validity involves essentially the systematic examination of the test content to determine whether it covers a representative sample of the behaviour**

domain to be measured.” Explaining the meaning of representative sampling, **Frederick G. Brown** has said, **“Representative sampling means selecting items in proportion to their emphasis or importance.”**

The tests constructed for evaluating content knowledge, skill or other teaching objectives should be established with content validity, and this is a very important task, The chief purpose of achievement tests is to evaluate students’ knowledge and skills etc. relating to the content. “Content validity is the chief characteristic of this type of tests. Content validity indicates how far a test evaluates that knowledge, that is, how far a student has acquired learning in that subject, how much his understanding is, what skills he has acquired, etc.

By content is meant both subject matter and instructional objectives. By subject matter is meant behavioural changes which are produced in students by different units of knowledge and instructional objectives. How far a test evaluates these two, can be found out from its content validity. Subject matter can be used to find out validity only when it is helpful in realizing teaching objectives, and not merely for teaching in the classroom. According to the **National Committee on Test Standards (1966), “In the case of an educational achievement test, the content of the test may be regarded as definition of the one or more educational objectives.”**

(a)   To list out all units of the whole knowledge under that subject matter and to select those important units on which the test has to be based.

(b)   To list all those important and desirable behaviours which are necessary to be produced in students through instruction, and to select only those of them on which the test will be based.

(c)   Of the above two lists, to determine how much importance to be given to each list.

(d)   To prepare table of specification.

(e)   To determine the number of questions or problems or items of the test on the basis of table of specification.

Despite the fulfillment of the above characteristics, content validity has its certain limitations. According to **Frederick G. Brown,** its chief limitation is the absence of quantitative index which can summarize the degree of validity. In his words, **“The major limitation of content validity is the lack of quantitative indices that summarize the degree of validity.”** Secondly, this type of validity cannot be used in finding out the validity of ability tests and personality tests. Also, it is not a simple task to select the representative samples.

2.   **Criterion-related Validity :** When we desire to predict about future working efficiency or performance of some individuals on the basis of some test, or we want to estimate the performance of some individuals in a certain field, then the test has to be based on criterion-related validity. For example, if we want to know, how far the intelligence level of students can correctly predict their success in some future examination, then we will have to study criterion-related validity. This type of validity is called predictive validity, because it indicates to the fact about what can be said about the achievement or success of some individuals in a future test.

In the second case, the purpose of a test may be to measure the present ability, skill or behaviours in a certain field today. How far a test measures these present characteristics of individuals correctly, is called its concurrent validity. These two types of validity come under criterion-related validity.

According to **Grounlund, “Criterion-related validity may be defined as the extent to which test performance is related to some other valued measure of performance.”**

In predictive validity, this ‘other measure of performance’ is available in the future, and it indicates to the fact how for the success in a future task their success can be predicted.

In the words of **Frederick G. Brown, “Predictive validity refers to situations in which criterion data are collected at some future time.”**

Time

Test ←————————————————→ Criterion
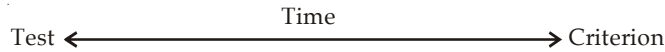
In concurrent validity, this 'other measure of performance' is available in the present, and it indicates to the fact how far the present ability or working efficiency can be measured accurately by that test.

The major problem faced in finding out criterion-related validity is the selection of the suitable criterion. Generally, appropriate and valid criterion is not available on the basis of which predictive validity or concurrent validity of a test under construction can be found out. This type of validity is established on the basis of certain criteria, and if these criteria are not valid themselves, the tests constructed on their basis cannot be reliable too. This is an significant problem. So, **Ebel** says that validity of psychological tests is a misleading characteristic. He also says that it is not necessary to find out validity of all tests, nor can it be. So, content validity or face validity of a test is very important, and it can be relied upon in a greater degree.

3. **Construct Validity :** According to **Cronback** and **Meehl,** when a test has to be analyzed in the form of a measurer of such attributes which can be defined operationally, in such a situation, construct validity is needed. This type of validity has been termed by Garrett as factoral validity, and this has been termed by Guilford as relevant validity. Validation is especially related with the application of tests. They point out to what extent a test evaluates knowledge of a subject or future efficiency of individuals of a specific field.

    From psychological viewpoint, each test evaluates psychological abilities of an individual, therefore the scores obtained on a test can be analyzed in the psychological context also. For example, the scores of a science test indicate the extent of knowledge in science; at the same time, it can point to the extent of reasoning ability in respect of science. In other words, test scores express those behaviours or expressions of an individual which become the basis of success in different fields. These psychological elements as manifested in test scores are the basis of construct validity.

Did u know? **Garrett** has explained this factor by the example of watch. Supposing you have set your watch faster by ten minutes, it will indicate 10.10 when it is 10 o'clock, but this time will not be valid because of its difference from the standard time. Of course, the watch will be reliable certainly.

According to **Anastasi, "The construct validity of a test is the extent to which the test may be said to measure a theoretical construct or trait."**

According to **Cronback** and **Meehl, "With construct validation both the theory and the test are being validated at the same time."**

The following procedure has to be followed to establish construct validity :

(a) To determine the psychological elements which are behind the test performance.

(b) To construct hypotheses on the basis of those elements.

(c) To verify these hypotheses.

The following methods are used for finding out construct validity :

(a) To compare the scores of different groups of individuals.

(b) To establish correlation with any other test.

(c) After having ascertained the number of test items, to determine what mental faculties or psychological elements related to the subject become active on responding to these items.

(d) To verify the result with the help of factor-analysis technique.

"The factoral validity of a test is the correlation between that test and the factors common to a group of tests or other measures of behaviour." —A. Anastasi

(e) To compare scores of individuals before placing them in the specific situation and after placing them in the specific situation; for example, to administer an intelligence test on a group of students before training and obtain their scores, and then administer an intelligence test on this group after training, and then go on to compare the two scores. If there is no variation in scores, then it can be accepted that the test measures intelligence correctly.

Face Validity

When a test appears to measure that element, trait or subject knowledge for which it has been constructed then it is said that it possesses face validity. In face validity, a test appears to be valid by the form of its items. This type of validity can be found out by mere looking at the items. Generally, validity of achievement tests and personality tests is found out in this way. For example, if we have to measure history-related achievement, then its items can be seen to say that it measures achievement in history or not. In the same way, a personality inventory which has been constructed to measure introvert and extrovert aspects of personality, should comprise of such items which can measure the two aspects equally well. Statistical methods cannot be used too calculate face validity. For this, the assistance of subject specialists or field specialists is taken. On reading of items, an effort is made to find out what attribute they measure correctly. Therefore, only qualitative statements can be given about face validity; for example, high, low, average or poor etc. **Mosier** has mentioned four types of face validity : by assumption, by definition, by appearance and by hypothesis.

Notes Factor-analysis is a type of special statistic technique in which the minimum factors existing in tests are found out with the help of correlation coefficients obtained in different tests. The credit for developing this technique goes to **Thomson** (1960) and **Burt** (1941).

Self Assessment

1. *Fill in the blanks :*

(*i*) The extent to which the purpose or objective of a test is realized is called its _____ .

(*ii*) According to _____, Validity of a test are of _____types.

(*iii*) The major limitation of content validity is the lack of _____that summarize the degree of validity.

(*iv*) In _____, a test appears to be valid by the form of its items.

(*v*) The major problem faced in finding out _____is the selection of suitable criterion.

6.3 Methods of Validity

The chief aim of validity determination is to know whether a test is measuring that trait to the required extent and with accuracy, so that meaningful hypotheses can be formulated on its basis, and important decisions can be taken on its basis. Following are the important methods of calculating validity :

1. Simple correlation method.

2. Bi-serial correlation method.

3. Point bi-serial correlation method.

4. Tetrachoric correlation method.

5. Expectancy table method.

6. Cut-off score method.

7. Differential predictors method.

8. Factor-analysis method.

1. **Simple Correlation Method :** This method is the most prevalent method to calculate validity coefficient. This is a very simple method. In this, students' scores are compared with the correlation efficient scores obtained on criterion. In calculating correlation coefficient, the product moment method is used. Sometimes, the scatter diagram method is also used if the number of students is large, whose formula is as follows :

$$r = \frac{N\Sigma\, fd_x d_y - \Sigma\, fd_x \Sigma\, fd_y}{\sqrt{\left[N\Sigma\, fd_y^2 - \left(\Sigma\, fd_x\right)^2\right]\left[N\Sigma\, fd_y^2 - \left(\Sigma\, fd_y\right)^2\right]}}$$

If the amount of correlation efficient is positive and of high level, then the level of a test's validity too is considred high. On the basis of correlation coefficient, some important conclusions can be drawn :

(a) What is the correlation coefficient of the test ?

(b) From the change occurring in a variable from the calculation of r^2, the effect of change on another variable can be estimated; for example, supposing correlation coefficient is .50, then $r^2 = (.50)^2 = .25$, or the difference of 25% which has occurred from Y-variable, has occurred due to the difference of scores in X-variable.

(c) Just in this way, we can use the regression equation in order to predict the scores of an individual in one subject (X) on the basis of his scores in another subject (Y). The concerned formula is :

$$Y = r\frac{\sigma_y}{\sigma_x}(X - M_X) + M_Y$$

Besides, standard error of estimate in this prediction can be calculated with the help of the following formula :

$$Y = r\frac{\sigma_y}{\sigma_x}(X - M_X) + M_Y$$

(d) On the basis of correlation coefficient too, the forecasting efficiency of a test can be estimated, for which the following formula is used :

$$\sigma\left(esty\right) = \sigma_y\sqrt{1 - r^2}$$

For example, supposing correlation efficient between test and criterion is .80, then prediction efficiency of the test will be .40 or 40%.

$$E = 1 - \sqrt{1 - (.80)^2}$$
$$= 1 - \sqrt{1 - .64}$$
$$= 1 - \sqrt{.36}$$
$$= 1 - .60 = .40$$

2. **Bi-serial Correlation Method :** Under certain situations when the individuals of a group fall under a variable, then we can measure in the score form, but under the second variable, we can divide into categories only those individuals, and the correlation calculated between these two variables is called bi-serial correlation, and the variable which is divided into two categories is called dictotomous variable.

For example, if we want to know correlation between mental age and social adjustment, then we can measure mental age of students under the first variable, but under the second variable, we will have to divide students into two categories. One category will be socially adjusted and the second will be socially maladjusted.

Some examples of dichotomous variables can be as follows :

(a)   Likes and dislikes.

(b)   Radical and conservative.

(c)   Socially-minded and mechanically-minded.

(d)   Pass and fail.

(e)   Athletic and non-athletic.

(f)   Drops-outs and stay-ins.

This correlation is calculated by the following formula :

$$r_{bias} = \frac{m_p - m_q}{\sigma} \cdot \frac{pq}{y}$$

Where, m_p = mean of the group in the first category in which the group shows higher ability

m_q = mean of the group of the second category

σ = SD of the entire group

p = proportion of the entire group in category one

q = proportion of the entire group in category second, i.e.

$q = 1 - p$

and y = the ordinate of the normal curve at the point of division of p and q.

Bi-serial correlation has its own certain limitations too, which are as follows :

(a)   Bi-serial correlation is less reliable than product moment correlation. The standard error of r_{bias} obtained from the same sample is more than Pearson correlation efficient (r). From this standpoint, Pearson (r) is better than r_{bias}.

(b)   Though r_{bias} is correlation between two variables, yet it cannot be applied in the regression equation, nor can its error of estimate by predicted.

(c)   Though r_{bias} is an estimate of product moment (r), yet its value is not limited within plus-minus 1, but it can exceed 1. therefore, this coefficient cannot be compared with the other coefficient. Finally, r_{bias} is still suitable for finding out validity in achievement tests.

3. **Point bi-serial Correlation Method :** The chief difference between bi-serial correlation method and point bi-serial correlation method is that in the point bi-serial correlation method, the dichotomous variable is in fact discrete. Therefore, when some such situation arises, then the point bi-serial correlation method should be used. It has been seen that some variables are such which are not dichotomous in fact, and their distribution from the standpoint of population is also homogeneous. In such a situation from the viewpoint of convenience, these variables itself are taken as dichotomous. Supposing, an item has been scored correctly and incorrectly. Now the truth remains that the students who have solved this question correctly, their abilities within the group is not similar. If the ability of these students is measured fully, it can be seen that it is continuously distributed, but in practice, we divide these students into two groups, and thus

make a rough prediction. From the standpoint of scoring, the students who solve the item correctly are given 1 mark and those who are not able to do so are given 0 mark. Thus, we have only two classes, and the correlation calculated between these two is equal to r_{pbi} of Pearson coefficient.

The value of r_{pbi} can be calculated by the use of a formula, which is as are as follows :

$$r_{pbi} = \frac{m_p - m_q}{\sigma} \cdot \sqrt{pq}$$

Where, M_p = Mean of the first group

M_q = Mean of the second group

p = Ratio of sample in the first group.

q = Ratio of sample in the second group.

σ_1 = Standard deviation of the entire group.

4. **Tetrachoric Correlation :** This method is used in such a situation when both variables X and Y are dichotomous. If they are dichotomous, rbias cannot be calculated, and in their place, tetrachoric correlation is calculated. In some appropriate conditions, the value of r_1 comes equal to Pearson (r). Therefore, r_1 is supposed be approximation of Pearson (r).

The following formula is used to calculate this type of correlation :

$$r_1 = \cos\left\{ \frac{180}{1 + \sqrt{\dfrac{ad}{bc}}} \right\}$$

Here, a, b, c, d frequencies are arranged as per the following table

X-Variable

Y-Variable	–	+	Total	Porportions	
+	– +	+ +	$a + b$	P
–	––	+ -	$c + d$	Q
Total	$b + d$	$a + c$		1.00
Proportions	P′	Q′	1.00	

Both variables have been divided into two categories : X and Y. Cell A has + + entries and Cell B has – –. If correlation is fully positive, then all cases will be set in a and d cells, and if correlation is fully negative, all cases will be set in b and c cells. If frequencies are similar in all four cells, then correlation will be zero. On applying the value of a, b, c, d in the formula, when the bracket is opened, a number is received which is in the form of a fraction angle. The cos value of this angle is found out. This value expresses the extent of r_t. When the value of ad is more than bc, then the value of r_t will be positive and when the value of bc is more than ad, then r_t will be negative. When the value of bc is equal to ad, then the value of the angle will be 90°, and the value of r_t will be equal to cos 90° or zero. When angle is 0°, then the value of r_t will be +1 and when it is 180°, then the value of r_t will be –1.

The values of more than cos 90° are not given in the table. When the angle is between 90° and 180°, then it is deducted from 180° and its cosine value is referred and the obtained value is given the prior mark.

Tetrachoric correlation is found advantageous when we want to see the correlation between two attributes, and it is not possible to measure these two variables in the form of scores, but

these variables can be divided into two categories. For example, if want to calculate correlation between school attendance and current employment. Here, we shall divide the individuals into two categories. In the first category will be the individuals who have passed high school and the second category will have individuals who have not passed high school; on then we shall divide these individuals into employed and unemployed categories, and then go on to calculate correlation. In the same way, if we want to see correlation between intelligence and social maturity, then we will have to make such categories, as above average, below average, socially mature and socially immature etc. in order to calculate r^t.

INTELLIGENCE

	Above Average	Below Average
MATURITY Poorly Adjusted SOCIAL	*a*	*b*
Well Adjusted	*c*	*d*

Assumptions : The assumptions of the tetrochoric correlation are as follows :

(a)   Both variables X and Y should be continuous.

(b)   Both variables should be normally distributed.

(c)   Both variables should be lineally related.

Supposing, the following two questions were given to students :

(a)   Do you find pleasure in meeting others ?

(b)   Do you find pleasure in working with others ?

Here, the assumptions related to continuity and normality will be as follows :

The students who have written Yes in these questions, but have not done it with the same degree of consistency, and those who have written NO and have not done so with the same degree of consistency. So, each question presents a continuous variable, and this continuous scale has consistence on one end and lack of consistency on the other end, or approval on one end and disapproval on the other end. When a variable is taken as continuous, then NPC can be taken for a psychological trait.

5.   **Multiple Correlation Method :** This method can be used when more than one scales are used for some hypothesis. In this, scores of different scales are used jointly in the forms of statistics, and it is correlated with the third scale in order to know its multiple correlation. Multi coefficient expresses correlation between one distribution of scores and two or more co-distributions. It indicates how the co-distributions are correlated with some definite hypotheses.

If we want to select trainees for some training, such as engineering, medicine or teaching etc., then we have to predict their success in the given profession; for example, we can use the intelligence test, study habits test and vocabulary test for admission to college. In the same way, before buying a car, we shall contact such people who already possess a car or some mechanic or will test-drive the car, and shall decide about its purchase only when satisfied. More than one predictor make our prediction more accurate.

Supposing we want to predict success of students in a college. For this, the vocabulary test was used, and supposing its criterion (test scores) is .5, then it would mean that 25% variance of the evaluation has been described by the vocabulary test.

X		Y
Vocabulary Test	25%	Grade Agerage

Now, supposing college students are also administered the intelligence test in addition to the vocabulary test, and the correlation between intelligence test and grade average is .5 and 0 by the vocabulary test. The intelligence test explains 25% variation of average variance. Thus, vocabulary and intelligence tests explain 50% variance together.

X		Y		X_2
Vocabulary Test	25%	Grade Agerage	25%	Intelligence Test

This joint variance which is described by predictor tests, is used in calculating correlation coefficient. This variance is equal to the square of multiple correlation. In this examination, the value of multiple correlation will be equal to .50 or .71, and it will be displayed by R_{123}. In its full form, following is the formula for multiple correlation :

$R_{1.234.....n}$

This symbol displays correlation of all tests 2, 3, 4,.... with test-l.

Thus, if we link several tests in the test battery, whose mutual correlation is zero and they are correlated with the criterion, then multiple correlation will be equal to the total sum of squares of the correlation of criteria of predictors, that is :

$R^2_{1.23}...n = r^2_{1.2} + r^2_{1.3} + ...+ r^2_{1.n}$

General Definition of Multiple Correlation :

"Multiple correlation is the correlation between scores on one variable and the composite scores on several other tests while these scores are optimally weighted."

Table : Semesters Completed in Percents

Decile	2	3	4	5	6	7	8
X	97	94	93	91	89	88	86
VII	93	86	84	80	78	77	77
VI	91	81	78	76	72	71	71
II	84	70	65	63	60	58	57
I	80	66	57	50	49	48	45

It is clear here that the students who are given X item (maximum), there is possibility that 70% students will complete the semester; while in the lowest decile group, there are only 45% students who will be able to pass the semester.

Prediction can also be displayed by another method, called bar diagram. By it, the predictive ability of a test is calculated. Under this, a test battery which has been used for selecting workers for pilot training, the percentage of students on stanine has been displayed. It is clear from it that on the basis of test, the individuals who obtain 9th stanine are least likely to be discharged from their work (only about 4% pilots were discharged due to inefficiency), while those obtaining 1st stanine, 70% of them were discharged due to inefficiency. Thus, explanation can be given on the basis of other stanines.

7. **Cut-off Score Method :** Cut-off score is a test score which is used as the demarcation point for accepting or rejecting students. For example, supposing the following scores indicate adjustment in some adjustment list, and we consider 9 as the cut-off score, then it would mean that the students who have scored 9 or less on the adjustment inventory will be accepted, and those

scoring below it will be rejected. The following table displays the number of accepted and rejected candidates and the number obtained on the basis of cut-off score on the basis of a psychiatric test :

Cut-off Score	600 Accepted	400 Rejected
7	28%	86%
13	13%	74%
23	4%	50%

It is clear from the table that 86% candidates were rejected after interview and they can be rejected on the basis of test too. Also, the 28% candidates who are accepted on the basis of interview, can be rejected on the basis of this cut-off score.

This method can be used for determining predictive validity.

This method is used when the vacancies are few and the number of candidates is large. The chief aim of this method is to select suitable candidates and rejection of unsuitable candidates. However, since every test Jacks complete validity, and any one test cannot measure all aptitudes and skills of candidates, so it may be possible that some suitable candidates may be rejected while some unsuitable candidates may be selected. Despite some limitations, this method is considered a good objective method.

Task   What is prediction in the Content of Correlation ?

---

8.   **Differential Predictors Method :** Validity of a test is evaluated on this basis also that its correlation with two different criteria is different. Many tests are criticized on this ground that their predictive validity is equal for two different criteria. For example, supposing test-A's correlation with students' scores in social studies is .55, and it is .45 in physics; it means that the test does not discriminate between two types of abilities, that is social studies and physics. However, it is a totally misleading statement, because human abilities, capabilities and skills are interrelated. Besides, it is also true that the achievement of an individual in educational or other fields is not dependent merely on intellectual or psychomotor abilities, rather it is also influenced by his motivational and personality factors.

9.   **Factor-analysis Method :** This is a method on the basis of which psychologists discover those psychological elements which can be independent units. Human brain is a very complex blending of many types of elements and characteristics. It is a large network of these. The effort to distinguish these elements and characteristics into independent units is the chief objective of factor-analysis. According to **Spearman**, intelligence cannot be only one mental faculty. If it was so, there would have been complete correlation among all tasks that he may undertake; that is, if an individual succeeds in one task, he would have been equally successful in doing another task as well, and if an individual fails in one task, he would have been equally failure in doing another task as well. However, in fact it is not so. On this basis, it is said that intelligence or complete mental ability is a blending of two elements.

The basis of factor-analysis is correlation. If two activities has some common or combined factor, then there will be correlation between two activities. In his book 'Factoral Analysis for Non-mathematicians', **C.J. Adcock** has given a very appropriate example for this situation.

6.4 Usability of Validity

Educational assessment should always have a clear purpose. Nothing will be gained from assessment unless the assessment has some validity for the purpose. For that reason, validity is the most important single attribute of a good test.

The validity of an assessment tool is the extent to which it measures what it was designed to measure, without contamination from other characteristics. For example, a test of reading comprehension should not require mathematical ability.

It is fairly obvious that a valid assessment should have a good coverage of the criteria (concepts, skills and knowledge) relevant to the purpose of the examination. The important notion here is the purpose. For example :

- The PROBE test is a form of reading running record which measures reading behaviours and includes some comprehension questions. It allows teachers to see the reading strategies that students are using, and potential problems with decoding. The test would not, however, provide in-depth information about a student's comprehension strategies across a range of texts.

- STAR (Supplementary Test of Achievement in Reading) is not designed as a comprehensive test of reading ability. It focuses on assessing students' vocabulary understanding, basic sentence comprehension and paragraph comprehension. It is most appropriately used for students who don't score well on more general testing (such as PAT or e-as TTle) as it provides a more fine grained analysis of basic comprehension strategies.

Self Assessment

2. State whether the following statements are 'true' or false'.

(i) Simple Correlation method is the most prevalent method to calculate validity coefficient.

(ii) Multiple correlation method can not be used when more than one scales are used for some hypothesis.

(iii) Cut off score is a test score which is used as the demarcation point for accepting or rejecting students.

(iv) Prediction can be displayed by another method called bar diagram.

(v) The basis of factor analysis is coefficient.

6.5 Summary

- Validity is an extremely essential attribute of a test.

- The following formula will prove useful in remembering validity in brief :

  Adequacy + Objectivity $\rightarrow$ Reliability + Aim $\rightarrow$ Validity

  (i) "The validity of a test or of any measuring instrument depends upon the fidelity with which it measures what it purports to measure." — Garrett (1959)

  (ii) "In general, the term validity refers to effectiveness of a test in achieving specific purposes." —Stodola and Siordohl

- Criterion for Validity : An independent criterion is needed for determining validity. In mental tests, the measurable elements cannot be seen in their concrete forms, they can be perceived only through the medium in the form of activities.

- According to the National Committee on Test Standards, validity of a test can generally by done by the following three types of validity :

  (i) Content or curricular validity,

  (ii) Criterion-related validity, and

  (iii) Construct validity.

- According to Anastasi, "Content validity involves essentially the systematic examination of the test content to determine whether it covers a representative sample of the behaviour domain to be measured."

- The tests constructed for evaluating content knowledge, skill or other teaching objectives should be established with content validity, and this is a very important task, The chief purpose of achievement tests is to evaluate students' knowledge and skills etc. relating to the content.

- Criterion-related Validity : When we desire to predict about future working efficiency or performance of some individuals on the basis of some test, or we want to estimate the performance of some individuals in a certain field, then the test has to be based on criterion-related validity.

- According to Grounlund, "Criterion-related validity may be defined as the extent to which test performance is related to some other valued measure of performance."

- Construct Validity : According to Cronback and Meehl, when a test has to be analyzed in the form of a measurer of such attributes which can be defined operationally, in such a situation, construct validity is needed. This type of validity has been termed by Garrett as factoral.

- According to Anastasi, "The construct validity of a test is the extent to which the test may be said to measure a theoretical construct or trait."

- When a test appears to measure that element, trait or subject knowledge for which it has been constructed. then it is said that it possesses face validity. In face validity, a test appears to be valid by the form of its items.

- The chief aim of validity determination is to know whether a test is measuring that trait to the required extent and with accuracy, so that meaningful hypotheses can be formulated on its basis, and important decisions can be taken on its basis. Following are the important methods of calculating validity :

  (*i*) Simple correlation method.

  (*ii*) Bi-serial correlation method.

  (*iii*) Point bi-serial correlation method.

  (*iv*) Tetrachoric correlation method.

  (*v*) Expectancy table method.

  (*vi*) Cut-off score method.

  (*vii*) Differential predictors method.

  (*viii*) Factor-analysis method.

- Simple Correlation Method : This method is the most prevalent method to calculate validity coefficient. This is a very simple method. In this, students' scores are compared with the correlation efficient scores obtained on criterion.

- Bi-serial Correlation Method : Under certain situations when the individuals of a group fall under a variable, then we can measure in the score form, but under the second variable, we can divide into categories only those individuals, and the correlation calculated between these two variables is called bi-serial correlation, and the variable which is divided into two categories is called dictotomous variable.

- Point bi-serial Correlation Method : The chief difference between bi-serial correlation method and point bi-serial correlation method is that in the point bi-serial correlation method, the dichotomous variable is in fact discrete.

- Tetrachoric Correlation : This method is used in such a situation when both variables X and Y are dichotomous. If they are dichotomous,rbias cannot be calculated, and in their place, tetrachoric correlation is calculated.

- Multiple Correlation Method : This method can be used when more than one scales are used for some hypothesis. In this, scores of different scales are used jointly in the of statistics, and it is correlated with the third scale in order to know its multiple correlation. Multi coefficient expresses correlation between one distribution of scores and two or more co-distributions.

- Differential Predictors Method : Validity of a test is evaluated on this basis also that its correlation with two different criteria is different. Many tests are criticized on this ground that their predictive validity is equal for two different criteria.

- Factor-analysis Method : This is a method on the basis of which psychologists discover those psychological elements which can be independent units. Human brain is a very complex blending of many types of elements and characteristics. It is a large network of these. The effort to distinguish these elements and characteristics into independent units is the chief objective of factor-analysis.

Educational assessment should always have a clear purpose. Nothing will be gained from assessment unless the assessment has some validity for the purpose. For that reason, validity is the most important single attribute of a good test.

The validity of an assessment tool is the extent to which it measures what it was designed to measure, without contamination from other characteristics. For example, a test of reading comprehension should not require mathematical ability.

6.6 Keywords

Validity : The state of being legally or officially acceptable

Adequacy : Enough inquantity, or good enough in quality

Criterion : a standard or principle by which something is judged

6.7 Review Questions

1. Explain the meaning and nature of validity.

2. What are the types of validity ? Explain then.

3. Discuss the face Validity.

4. What are the methods of Calculating Validity ?

5. Give the usability of validity

Answers : Self Assessment

1. (*i*) Validity   (*ii*) National Committee on test standards, three
   (*iii*) Quantitative indices   (*iv*) face Validity
   (*v*) criterion related validity
2. (*i*) T   (*ii*) F   (*iii*) T   (*iv*) T   (*v*) F

6.8 Further Readings

Books

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 7 : Reliability – Types , Methods and Usability

CONTENTS

Objectives

Introduction

7.1 Reliability

7.2 Types of Reliability

7.3 Methods of Reliability

7.4 Usability of Reliability

7.5 Summary

7.6 Keywords

7.7 Review Questions

7.8 Further Readings

Objectives

The objectives of this unit can be summarised as below :

- To discuss about the reliability
- To describe about the types of reliability
- To explain about the methods of reliability
- To discuss about the usability of reliability

Introduction

Reliability is one of the most important elements of test quality. It has to do with the consistency, or reproducibility, of an examinee's performance on the test. For example, if you were to administer a test with high reliability to an examinee on two occasions, you would be very likely to reach the same conclusions about the examinee's performance both times. A test with poor reliability, on the other hand, might result in very different scores for the examinee across the two test administrations. If a test yields inconsistent scores, it may be unethical to take any substantive actions on the basis of the test. There are several methods for computing test reliability including test-retest reliability, parallel forms reliability, decision consistency, internal consistency, and interrater reliability. For many criterion-referenced tests decision consistency is often an appropriate choice.

7.1 Reliability

7.1.1 Meaning of Reliability

By reliability is mean to rely or trust. Generally we trust many things in our daily life without giving much thought to them. This trust is not formed over one or two days, but a strong assumption is created after having perceived those objects over a long period of time. This is the reason that a shopkeeper or friend becomes reliable and his opinion is accepted without any debate or argument.

Suppose, a cloth length has been measured at 3 *m* by a scale. Now the question arises, is this scale reliable ? If we measure that cloth length by any other scale and its length comes to 3 *m*, then we can easily say that it is reliable. However, if the cloth measures 3 *m* at one instance and 3.5 *m* at another instance, then this scale cannot be called reliable. The same thing applies to mental tests too. If a mental test measures a child's IQ at 110, and the same scale measures it at 110 again, then we shall call the first test as reliable. It is clear that by reliability of a test is meant is constancy of the score of measurement. It refers to the consistency of the measurement. That is, if the test is repeated under the same set of known conditions and the obtained scores should remain the same, then the test is said to be reliable.

Supposing, an individual is given a test today, and then the same test is administered to him after an interval of time, and his scores are equal in both cases, then the test will be said to be reliable. In the same manner, if a test is administered on several individuals repeatedly, and its scores are equal like the first instance, then the test will be considered reliable.

According to **Anastasi**, **"The reliability of a test refers to the consistency or scores obtained by the same individuals on different occasions or with different set of equivalent items."**

Thus, reliability is that attribute of a test which is responsible for making its results equal. On the contrary, if the scores of a student or a group of them change sufficiently, then the reliability of this test will be considered to be low. The more reliable is a test, the more consistent will be the scores of a student or a group of them. It is not necessary that these scores will be exactly equal under all equal circumstances, there can be minor variation between them due to some other causes, but this variation should not be excessive. In theory, reliability is considered the measurer between observed scores and true scores. According to **Guilford**, an obtained score is the sum of true score and error score. That is :

$$X = T + E$$

Where, X = Obtained score

T = True score

E = Error score

This thing has been explained by **Helmesteder** using the following formula :

$$X_i = t_i \times e_i$$

Where, X_i = Score by a certain student

T_i = True score by a certain student

E_i = Error score by a certain student

7.1.2 Reliability Defined in Terms of Variance

The formula used for reliability $X = T + E$ tells about the relation of obtained score, true score and error score; however, because reliability is an attribute of testing, not of the scores of a certain individual; so the following formula can prove to be more meaningful :

$$SX^2 = ST^2 + SE^2$$

Where, SX^2 = Total variance

ST^2 = True variance

SE^2 = Error variance

That is, on the basis of the above formula, total variance is found in two types of variances : true variance and error variance. True variance is found due to systematic sources, that is, personal distinction of individuals; and error variance occurs due to unsystematic sources, such as guess, recording error or scoring error etc.

According to **Frederick G. Brown, "Reliability can be defined as that ratio of true variance in a set of test scores to the total or obtained variance."**

$$R_{xx} = \frac{St^2}{Sx^2}$$

Where, R_{xx} = Reliability co-efficient of the test

St^2 = Variance of true score

SX^2 = Variance of observed or obtained score

The more is true variance, the lesser will be error variance, and the more reliable will the test be considered. In fact, measurement of true variance cannot be done directly, it can only be estimated. Therefore reliability is a imaginary attribute of test scores.

Some scholars have defined reliability in the following words :

1. **"Briefly, reliability is the proportion of the true variance in obtained true scores."**
    —*Guilford*

2. **"...defined reliability of measurement as the extent of unsystematic variation in scores of an individual to some trait when that trait is measured a number of times."** —*Ghiseli*

3. **"The reliability of a test of any measuring instrument depends upon the consistency with which it gauges the ability to whom it is applied."** —*Garrett*

4. **"The reliability of a test can be defined as the correlation between two or more sets of scores on equivalent tests from the same group of individuals."** —*Stodola and Stordahl*

5. **"The term reliability refers to the extent to which a test is internally consistent and the extent to which it gives consistent result on testing and retesting**." —*Freeman*

It is essential to consider the following points while examining reliability of a test :

1. Reliability is a trait of the test results and not of the test itself.

2. Obtained scores of a test are not equally reliable, that is, the obtained scores will be equal under all circumstances, this is not meant by reliability, rather it means that they will be equal under certain and specific circumstances, such as after an interval of time or after the test is administered on the one-half or one-forth of the group.

3. Reliability of a test affects its validity too, that is, reliability is a necessary but not a sufficient condition for validity.

4. If the reliability of a test is low, then its validity too will be low; however, it may not be true that the validity of a test will be in full measure if its reliability is in full measure. It may be possible test may be 100% reliable but not fully valid.

Notes   Test reliability is the aspect of test quality concerned with whether or not a test produces consistent results.

7.2 Types of Reliability

Decision Consistency

In the descriptions of test-retest and parallel forms reliability given above, the consistency or dependability of the *test scores* was emphasized. For many criterion referenced tests (CRTs) a more useful way to think about reliability may be in terms of examinees' *classifications*. For example, a typical CRT will result in an examinee being classified as either a master or non-master; the examinee will either pass or fail the test. It is the reliability of this classification

decision that is estimated in decision consistency reliability. If an examinee is classified as a master on both test administrations, or as a non-master on both occasions, the test is producing consistent decisions. This approach can be used either with parallel forms or with a single form administered twice in test-retest fashion.

Internal Consistency

The internal consistency measure of reliability is frequently used for norm referenced tests (NRTs). This method has the advantage of being able to be conducted using a single form given at a single administration. The internal consistency method estimates how well the set of items on a test correlate with one another; that is, how similar the items on a test form are to one another. Many test analysis software programs produce this reliability estimate automatically. However, two common differences between NRTs and CRTs make this method of reliability estimation less useful for CRTs. First, because CRTs are typically designed to have a much narrower range of item difficulty, and examinee scores, the value of the reliability estimate will tend to be lower. Additionally, CRTs are often designed to measure a broader range of content; this results in a set of items that are not necessarily closely related to each other. This aspect of CRT test design will also produce a lower reliability estimate than would be seen on a typical NRT.

Interrater Reliability

All of the methods for estimating reliability discussed thus far are intended to be used for objective tests. When a test includes performance tasks, or other items that need to be scored by human raters then the reliability of those raters must be estimated. This reliability method asks the question, "If multiple raters scored a single examinee's performance, would the examinee receive the same score. Interrater reliability provides a measure of the dependability or consistency of scores that might be expected across raters.

Self Assessment

1. *Fill in the blanks :*

    (*i*) According to _____ , "The reliability of a test refers to the _____ of scores obtained by the same individuals on different occassions or with different set of equivalent items.

    (*ii*) _____ demonstrates the extent to which a test is able to produce stable, Consistent scores across time.

    (*iii*) The _____ measure of reliability is frequently used for norm referenced tests.

    (*iv*) _____ reliability provides a measure of the dependability or consistency of scores that might be expected across raters.

7.3 Methods of Reliability

The chief basis of ascertaining reliability is correlation. The relation between two equivalent or parallel test scores is called reliability coefficient, which indicates the reliability of the test. Following are the chief methods of finding out reliability coefficient :

1. Parallel form or alternate method,

2. Method of rational equivalence or K.R. formula,

3. Test-retest method,

4. Split-halt method.

1. Alternate Method or Parallel Form Method : Under this method, a test is constructed in two parallel formats. The questions in these two parallel test are different from each other, but they are equal from all other views. It essential in this method that both tests are parallel. The parallel form 'A' and 'B' of many standardized tests are already available. In order to

find out their reliability, they are administered to certain group of students or individuals, and then the scores obtained on these two parallel tests are used to calculate correlation coefficient (*r*). This correlation multiple indicates their reliability.

Under this method, individual differences of experience, practice and students do not influence reliability, because the material in the two tests is different. But if there is much similarity between the questions and materials of the two tests, then their reliability will be enhanced. Generally, if two tests are administered with a sufficient time interval, then experience, memory and other factors do not influence scores. It is suitable to have a time interval of four weeks between the two tests.

Following points should be kept in view while preparing parallel formats :

   (a)   The items should be distributed equally from the standpoint of difficulty.

   (b)   The items should be homogeneous.

   (c)   The administration and scoring methods of the two tests should be equal.

   (d)   The number of items in the two formats should be equal.

   (e)   Type of items, content, difficulty level and samples should be equal in the two formats.

Limitations

   (a)   Tests have to be conducted twice.

   (b)   This testing produces the problem of standardization, which in itself is a very complex and expensive process.

   (c)   If the time interval between the administration of two tests is longer, then the error of testing-retesting method will be repeated in this too.

   (d)   Exercise has its influence in this method, because the form of items is almost equal or similar.

   (e)   It is a difficult task to prepare two equal and similar formats for testing.

Merits

   (a)   This method consumes far less time as compared to testing-retesting method.

   (b)   This method is the amended form of testing-retesting method.

   (c)   This has least influence of exercise and memory.

   (d)   This method can be used for follow-up purpose too.

   (e)   This method can be used to ascertain reliability of speed test too.

   2.   **Method of Rational Equivalence or K-R Formula : Kuder** and **Richardson** presented a method to ascertain reliability, which is called method of rational equivalence or K-R formula. Under this method, the limitations of all other methods have been eliminated. Under it, the test has to be administered only once, and correlation between question items is found out in order to see similarity between them. So, the reliability coefficient obtained from this method is also called coefficient of internal consistency. The chief characteristic of internal consistence is that different items of the test have high correlation with each other. The chief assumption of the application of this method is that all items included in this test should be homogeneous, else reliability coefficient will be less than the split-half method. It entails that a test has the same type of questions, because the aim of all these questions is to measure the elements which influence ability and personality.

   Under this method, two points are given special importance, these are inter-correlation of the items and correlation of items with the while test.

   By inter-correlation of questions is meant to measure the consistency of reactions in a subject which occur for different items. That is, how consistent are one statement in relation

to another. This consistency is influenced by two types of errors. First, how the test questions have been selected, whether the questions are right or not, and second, heterogeneity of behaviour, that is, if the same type of behaviour has been studied. If a test studies different types of behaviours at the same time, then it would be less reliable than a test which studies the same type of behaviour. For example, if a test contains only multiplication questions and another test has addition, subtraction, division and multiplication questions, then the first test will have more inter-item consistency as compared to the second one, because if a student scores 10 in the second test, it would be difficult to tell if a student is good in addition, or subtraction, or division or multiplication. So, all subjects cannot be compared using heterogeneous tests. Now the question arises to ascertain whether the test which we are going to use for prediction is homogeneous or heterogeneous. Though a homogeneous test is good, yet a question arises : Can a homogeneous test predict about a heterogeneous test ? Therefore, if we have to measure heterogeneous behaviour, then we have to take a heterogeneous test, and the amount of error variance will have to be controlled in it. supposing, if an intelligence test is heterogeneous and can predict well, but if it is in the battery form and measures only one ability at one time, and in the sum total, also measures many abilities. There are many formulae for calculating reliability coefficient in this method, but the following formula is the most prevalent and useful. It is called K-R formula :

$$r_{11} = \frac{n}{n-1} \times \frac{\sigma^2 - \Sigma pq}{\sigma^2}$$

Where, $\quad r_{11}$ = Reliability coefficient of the whole test

n = Number of items in a test

σ = Median deviation of scores in the test

p = Proportion of students solving each question correctly

$q = (1 - p)$ Proportion of students solving each question incorrectly

In 1951, **Cronback** standardized this formula mathematically and arrived at this conclusion : **"K-R coefficient is actually the mean of all split-half coefficients resulting from different splittings of a test but, unless the test items are highly homogeneous, the K-R coefficient will be lower than the split-half reliability."**

That is, the reliability of K-R is equal to that of split-half, but the only condition is that the test should be homogeneous. This method is not good enough to know reliability of a heterogeneous test.

In order to find out reliability coefficient, at first, it is found out how many percent students have solved each question correctly and how many percent students have solved each question incorrectly. They are displayed by p and q respectively. Thus, the value of p and q is found out for each question and they are multiplied with each other, which is called pq. In the end, the pq values of all questions are added which becomes Σpq. After that, the scores of each student in the whole test are taken and their standard deviation (σ) is calculated and squared, which is called (σ^2). Then, these values are set in the above formula and thus reliability coefficient is calculated.

Limitations

(a) The reliability coefficient obtained from this method is somewhat less than that obtained by other methods.

(b) The K-R formula is based on the basic assumption that the difficulty level of all items will be the same, but in practice, it is not possible.

(c) This method cannot be used in speed tests, because the parts of the test are not independent.

Merits

(a) The scores are obtained after administering the test only once under this method, which results in saving time and energy.

(b) This method is quite wide.

3. **Test-Retest Method :** In the test-retest method, a test is administered to a definite group of students or individuals. After some time, that test is once again administered to that group. This is reliability coefficient. This is the simplest method. The higher the reliability coefficient (*r*) between two tests, the more reliable is the test. The assumption behind the test-retest method is that if the test is reliable, then the scores obtained by the group will be repeated the next time also. There may be some variation, but less the variation, the higher the reliability.

This method too has certain limitations. The first demerit is that if the test is repeated soon after the first test, then students will continue to remember the answers of the first test and they will make an effort to write them in the second time. Thus, the scores in the retest will not be independent, rather they will be influenced by the first test. Consequently, the contact between the two scores will influence the reliability of the test. Its second demerit is that if the test is administered after a long interval of time, such as six months or more, then the second time scores will be influenced by the factor of time, because six months is a sufficient period to effect much mental development in students. This development has its influence on students' performance, and in this case, their scores will be more than those of the first time. Thus, score coefficient (*r*) is decreased, and the reliability of the test takes a back seat. Its third demerit arises due to practice and experience. When students have solved such types of questions once, they have gained experience and practice, and they are benefited by it when they are administered the test the second time, which helps to increase their scores. It decreases the reliability of the test. It is clear that the method is certainly simple, but it is not more accurate.

4. **Split-half Method :** In the above two methods, test-rest method and parallel method, the test has to be administered twice, but it is possible that a test may not be administered twice or two formats of the test may not be available. In such a situation, split-half method is considered' the most suitable. In this method, the test is not administered twice, rather : he test items are divided into two parts on the basis of the same test, and then separate scores of students are found out in order to ascertain their correlation coefficient.

There can be several methods of diving a test into two parts; however, the commonest of them is the even-odd method. In this method, all the test are arranged in the ascending order of difficulty, and then even numbered items, such as 2, 4, 6, 8, etc. are segregated from the odd numbered items, such as 1, 3, 5, 7, etc. Thus, the test is divided into two parts. Then, the coefficient between two parts is calculated. It is not necessary that only odd and even numbered items be segregated. There is yet another method to do it which is called first half-second half method. In this method, the first half items are kept in the first test and the second half items are kept in the second test. But this method is not suitable, because if the questions have been arranged in the ascending or descending order of difficulty, then one part will have all the simple questions and another part all the difficult ones, and thus, the two parts will not be homogeneous, rather these two parts will become two different tests. Also, the amount of coefficient will be decreased.

This problem is also faced when questions are divided on even and odd numbers, if the questions have been arranged in ascending order of difficulty; for example :

Odd	Even
1	2
3	4
5	6
7	8
"	"
"	"
"	"
"	"
"	"
99	100

Because, the second question is more difficult than the first one, the fourth is more difficult than the third one; the sixth question is more difficult than the fifth one, and so on. Therefore, the even-numbered test will be more difficult than the odd-numbered test. This problem is solved by using the following way :

First Part	Second Part
1	2
4	3
5	6
8	7
9	10

Thus, *a* 10-question test has been divided into 1, 4, 5, 8, 9 and 2, 3, 6, 7, 10.

The correlation coefficient (*r*) between the two will be its reliability. To know the reliability of the whole test, the Spearman-Brown prophecy formula is used, which is are as follows :

$$r_{11} = \frac{2r_{1/2'1/2}}{1 + r_{1/2'1/2}}$$

Where, R_{11} = Reliability coefficient of the whole test

$R_{1/2'1/2}$ = Reliability coefficient of one-half test

Supposing, the reliability coefficient of two parts of a test is *r* = .75, on this basis, the reliability coefficient of the entire test will be :

$$r = \frac{2 \times .75}{1 + .75} = \frac{1.50}{1.75} = .9 \text{ (approx)}$$

That is, reliability coefficient of the entire test will be .9 (approx).

Limitations

(a)   A limitation of this method is that chance error can arise in the test, which increases the amount of *r* in two parts.

(b)   It is not possible to use these tests in speed tests because the difficulty level of questions there is equal. **Nunnly** too has said : **"It Is particularly misleading to use the sub-divided test method on a highly speeded test."**

(c) A test is split half using several methods. Therefore, reliability coefficient can be different for splitting half a test using different methods. This type of coefficient has no unique value.

(d) Sometimes, the two parts of a test may not be homogeneous.

(e) In this test, the basis of reliability is half-test, though its reliability is counted for the full-length test, so from this view, splitting half a test may reduce its reliability.

Merits

(a) The chief characteristic of this method is that it does not consume much time and energy. A group is administered the test only once. Therefore, in such a situation where a retest is not possible, nor is it possible to prepare the two parallel or homogeneous tests, then this method becomes very important. Such a situation arises when reliability of a performance test or personality test has to be calculated, because these two types of tests cannot be constructed in parallel or homogeneous form, nor can they be given for retest; for example, Rorschach test or TAT test. Besides, this method is also suitable for calculating aptitude and interest.

(b) As the two parts are administered at the same time, there is no influence of time interval, practice, memory etc. on scores.

Hoyt's Reliability

Like Kuder-Richardson, **Hoyt** (1941) too laid down a method to find out reliability coefficient. According to this method, differences or deviation occurring in achievement of an individual from one item to another are not errors, but they are actual differences which we call intra-individual difference.

Hoyt has defined reliability as follows :

"... total variation observed is conceived to be made up of three components, true inter-individual differences, intra-individual differences (measured by item variance) and error inter-individual differences."

The above definition can be described in the following equation :

$$\sigma_x^2 \;=\; \sigma_t^2 + \sigma_i^2 + \sigma_e^2$$

Where, σ_x^2 = Variation in scores

σ_i^2 = Item variation

σ_e^2 = Error variation

σ_t^2 = Total variation

Therefore, according to Hoyt, the suitable definition of reliability is as follows :

$$VX^i \;=\; \frac{\sigma t^2}{\sigma x^2 - \sigma i^2} = \frac{\left(\sigma x^2 - \sigma i^2\right) - \sigma e^2}{\sigma x^2 - \sigma i^2}$$

On applying variance method, the calculation of reliability can be found out by the following equation :

$$VXX^i = \frac{MS_{ind} - MS_{resi}}{MS_{ind}}$$

Where, MS_{ind} = Average class of deviation of individual's average scores

MS_{resi} = Average class of variation of remaining items after deducting items and individual variation

---

Did u know? A test does not become a speed test by merely specifying the time limit, because if all students are able to complete the test within the specified time limit, then there is no reason for a student to increase his speed, and on the contrary, if a student is not able to complete the test, then it is not possible to measure his speed accurately.

7.4 Usability of Reliability

The **reliability** of an assessment tool is the extent to which it consistently and accurately measures learning.

When the results of an assessment are reliable, we can be confident that repeated or equivalent assessments will provide consistent results. This puts us in a better position to make generalised statements about a student's level of achievement, which is especially important when we are using the results of an assessment to make decisions about teaching and learning, or when we are reporting back to students and their parents or caregivers. No results, however, can be completely reliable. There is always some random variation that may affect the assessment, so educators should always be prepared to question results.

Factors which can affect reliability :

- The length of the assessment — a longer assessment generally produces more reliable results.

- The suitability of the questions or tasks for the students being assessed.

- The phrasing and terminology of the questions.

- The consistency in test administration — for example, the length of time given for the assessment, instructions given to students before the test.

- The design of the marking schedule and moderation of marking procedures.

- The readiness of students for the assessment — for example, a hot afternoon or straight after physical activity might not be the best time for students to be assessed.

Task What is K-R Coefficient ?

Self Assessment

2. *State whether the following statements are 'true' or false'.*

   (*i*) The chief basis of ascertaining reliability is correlation.

   (*ii*) The reliability coefficient obtained from the method of rational equivalence is also called coefficient of internal consistency.

   (*iii*) The higher the reliability coefficient (*r*) between two tests, the less reliable is the test in test retest method.

7.5 Summary

- According to Anastasi, "The reliability of a test refers to the consistency or scores obtained by the same individuals on different occasions or with different set of equivalent items."

- Thus, reliability is that attribute of a test which is responsible for making its results equal.

- In the descriptions of test-retest and parallel forms reliability given above, the consistency or dependability of the *test scores* was emphasized. For many criterion referenced tests (CRTs) a more useful way to think about reliability may be in terms of examinees' *classifications*.

- Internal Consistency

  - The internal consistency measure of reliability is frequently used for norm referenced tests (NRTs). This method has the advantage of being able to be conducted using a single form given at a single administration.

- All of the methods for estimating reliability discussed thus far are intended to be used for objective tests. When a test includes performance tasks, or other items that need to be scored by human raters then the reliability of those raters must be estimated. This reliability method asks the question.

- Parallel form or alternate method.

- Method of rational equivalence or K.R. formula,

- Test-retest method,

- Split-halt method.

- Alternate Method or Parallel Form Method : Under this method a test is constructed in two parallel formats. The questions in these two parallel test are different from each other, but they are equal from all other views. It essential in this method that both tests are parallel.

- Method of Rational Equivalence or K-R Formula : Kuder and Richardson presented a method to ascertain reliability, which is called method of rational equivalence or K-R formula. Under this method, the limitations of all other methods have been eliminated.

  There are many formulae for calculating reliability coefficient in this method, but the following formula is the most prevalent and useful. It is called K-R formula :

  $$r_{11} = \frac{n}{n-1} \times \frac{\sigma^2 - \Sigma pq}{\sigma^2}$$

  Where, r_{11} = Reliability coefficient of the whole test

   n = Number of items in a test

   σ = Median deviation of scores in the test

   p = Proportion of students solving each question correctly

   q = $(1 - p)$ Proportion of students solving each question incorrectly

- Test-Retest Method : In the test-retest method, a test is administered to a definite group of students or individuals. After some time, that test is once again administered to that group. This is reliability coefficient.

- Split-half Method : In the above two methods, test-rest method and parallel method, the test has to be administered twice, but it is possible that a test may not be administered twice or two formats of the test may not be available.

  There can be several methods of diving a test into two parts; however, the commonest of them is the even-odd method. In this method; all the test are arranged in the ascending order of difficulty, and then even numbered items, such as 2, 4, 6, 8, etc.

- Hoyt has defined reliability as follows :

  - "... total variation observed is conceived to be made up of three components, true inter-individual differences, intra-individual differences (measured by item variance) and error inter-individual differences."

  - The reliability of an assessment tool is the extent to which it consistently and accurately measures learning.

  - When the results of an assessment are reliable, we can be confident that repeated or equivalent assessments will provide consistent results. This puts us in a better position to make generalised statements about a student's level of achievement, which is especially important when we are using the results of an assessment to make decisions about teaching and learning, or when we are reporting back to students and their parents or caregivers.

7.6 Keywords

Reliability   :   The State in that somebdy can rely on something or somebody.

Consistency   :   The quality of always behaving in the same way

7.7 Review Questions

1.   Explain the meaning of reliability.

2.   What are the different types of reliability ?

3.   Give the methods of finding reliability.

4.   What is the usability of reliability ?

Answers : Self Assessment

1.   (*i*)   Anastari, Consistency   (*ii*)   test retest reliability

      (*iii*)   internal consistency   (*iv*)   interrater

2.   (*i*)   T   (*ii*)   T   (*iii*)   F

7.8 Further Readings

Books

1.   Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2.   Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3.   Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4.   Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 8 : Test Construction

CONTENTS

Objectives

Introduction

8.1 Meaning of test

8.2 Developing the blueprint of Question paper

8.3 Framing of questions and developing marking schemes

8.4 Formating, Question wise analysis and Review

8.5 Moderating the test

8.6 Summary

8.7 Keywords

8.8 Review Questions

8.9 Further Readings

Objectives

The objectives of this unit can be summarised as below :

- To discuss about the meaning of the test
- To explain about the developing the blueprint of question paper
- To describe the framing of questions and developing marking schemes
- To discuss about the formatting, question wise analysis and review
- To explain about moderating the test

Introduction

A test is usable or practicable if it is easy to construct, administer, score and interpret. MCQ are easy to score but difficult to construct. ETQ are easy to construct but difficult to score. If a Q.P. has two sections, one for MCQ and the other for open-ended questions (OEQ), difficulty is of having two separate question papers. Longer test requires more time for administration. Complex operation, summation of scores, analysability of scores, provision of marking scheme, development of format, mode of indicating answers etc. affect the administration and interpretation of test. Application of scoring key; use of correct formulate, machine or manual scoring; and arrangement of questions in the form of objective, content or form have a direct bearing on the interpretability of scores.

8.1 Meaning of Test

Test : A set of questions or exercises evaluating skills or knowledge. One valid approach to assessment is to observe everything that is taught. In most situations this is not possible, because there is so much information to be recorded. Instead, one has to select a valid sample from the achievements of interest. Since school learning programmes are expected to provide students with the capability to complete various tasks successfully, one way of assessing each student's learning is to give a number of these tasks to be done under specified conditions. Conventional

pencil-and-paper test items (which may be posed as questions) are examples of these specially selected tasks. However other tasks may be necessary as well to give a comprehensive, valid and meaningful picture of the learning. For example, in the learning of science subjects practical skills are generally considered to be important so the assessment of science subjects should therefore include some practical tasks. Similarly, the student learning music may be required to give a musical performance to demonstrate what has been learned. Test items or tasks are samples of intended achievement, and a test is a collection of such assessment tasks or items.

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

8.1.1 Definition of Test

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

8.1.2 Factors affect the Test

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

(*i*) Number of Papers

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

(*ii*) Total Time

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

(*iii*) Total Marks

A question paper carrying 100 or 50 marks has different implications. In the former it is possible to include more open-ended essay-type questions (time remaining the same) by increasing the number of short-answer and objective type questions. With less marks, as in the latter, possibility of including long-answer questions decreases and less number of questions would be included, thereby lowering the reliability of the test. Inclusion of more short-answer and objective-type questions would become necessary for better coverage of syllabus. Total marks therefore have its impact on the quality of question papers.

(*iv*) Scheme of Sections

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

> *Notes* One important implication of having sections on content basis is that you can impose restrictions on the number of questions to be attempted from each section, thereby ensuring better sampling of the content.

(*v*) Scheme of Options

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

(*vi*) Medium of Testing

Decisions about medium of testing and number of versions to be developed has also to be taken before commencing to write the questions. Framing of questions directly in a language and getting them translated from questions written in another language is quite different. Validity of many questions is lost during their translation from English to other languages. Unless the framer of the question is cognizant of this fact during construction of questions some good questions that cannot be properly translated into another language have to be sacrificed at the altar of validity of the question.

8.1.3 Coverage of Examination Syllabus

The syllabus is to be delineated into convenient units of testing, neither too big nor too small. Small units may hinder inclusion of essay-type questions, whereas bigger units may lead to neglect of some topics while framing the questions. Each of these units is given proportional weightage, depending upon the length of each unit in the text-book, importance of the topic and the density of content elements in each unit. A table may be prepared representing the weightages to various units, say 8 in all.

Table 1 : Showing weightage to different units

Units of syllabus	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5	Unit 6	Unit 7	Unit 8	Total
Marks allotted	4	12	8	20	12	24	10	10	100

This table indicates that :

(*i*) questions are to be set on all units of the syllabus;

(*ii*) questions are to be set in terms of proportional marks allotted to different units; and

(*iii*) adequate coverage of syllabus is ensured by proper sampling of content and giving proportionate weightage to different units of syllabus.

For languages the broad units may be orthography, structures, vocabulary and composition, which may similarly be given proportionate weightage for a particular examination. Similar type of table can be prepared by teachers for their annual examination or pre-board examination, usually conducted by the schools. This table is a step towards ensuring adequate sampling of content for proper coverage of examination syllabus to take care of content validity of the test, i.e. the question paper.

8.1.4 Coverage of Assessment Objectives

In a written test there are three major objectives that are usually tested, viz. knowledge, understanding and application objectives, besides of course the psychomotor objective, e.g. the drawing skill in various content subjects like sciences, social sciences and mathematics. For languages it is the knowledge, comprehension and written expression that are tested in a question paper. These objectives are also tested in terms of proportionate weightage given for the purpose of paper setting. These weightages vary or may vary from one examining board to another, depending upon the background of examinees, curriculum prescriptions, instructional strategies in vogue and the performance standard of students over the years. When applied to unit-level testing for various topics, the weightages to various assessment objectives would differ from unit to unit, depending upon the nature and scope of content elements in each unit. A table (Table 1) indicating the weightages given to various objectives was prepared in one of the school boards at one time.

Table 2 : Weightage to assessment objectives in content subjects

Assessment objectives				
Objectives	Knowledge objective	Understanding objective	Application objective	Skill (Drawing) objective	Total
Marks allotted	40	30	20	10	100

Table 3 : Weightage to assessment objectives in language subjects

Assessment objectives			
Objectives	Knowledge of functional grammar	Reading comprehension	Written expression	Total
Marks allotted	24	36	40	100

These weightages can vary after a few years, especially when new curricula are prescribed and emphasis on various assessment objectives is changed. This table indicates that :

(*i*)   all assessment objectives are being included in test papers;

(*ii*)   proportionate weightage is given to the three assessment objectives; and

(*iii*)   questions are to be set on all assessment objectives in accordance with the weightages indicated in the table.

Thus we find that necessary step is taken to ensure the needed curricular validity by covering the assessment objectives proportionately while developing the question papers.

8.1.5 Use of Different Forms of Questions

A comprehensive test like the boards' external examination or an annual examination of the school must utilise all forms of questions, viz. essay type, short answer and objective type, to ensure wider coverage of skills, abilities and content elements. Background of students, needed objectivity, extent of acceptability, reliability, nature of subject and the total time and marks stipulated do limit the scope of various types of questions for inclusion in a question paper. There is now a clear shift towards inclusion of more and more short-answer and objective-type questions. Depending upon the state of art in different boards, weightages to different forms of questions can be given to ensure proper coverage of the assessment objectives and also wider coverage of syllabus.

Table 4 : Weightage to different forms of question

Forms of questions					
Forms of questions ⟶	Essay type or long answer	Short answer (Free response)	Very short answer (Fixed response)	Objective type	Total
Marks allotted	30	40	10	20	100

This table indicates that :

(*i*) four types of questions are to be used in the question paper;

(*ii*) different types of questions would receive the stipulated weightage in setting; and

(*iii*) total marks for each type of question will be the basis for deciding the number of questions.

8.1.6 Difficulty Level of Questions

Provision must be made in the design to include questions that cater to all the ability levels of students in heterogeneous groups of students, as we normally have in public examinations. Difficulty level can be estimated for each question by paper setters on the basis of their previous experience and expert judgement. It becomes still easier if the constructor has an access to some item bank in a subject and has the choice to select items of required difficulty level. Anyhow, assuming a big mixed-ability group of examinees comprising bright, average and poor students we may fix the difficulty level that approximates the normal distribution, as shown in Table 1.

Table 5 : Estimated difficulty level

Estimated difficulty level	Difficult questions	Average questions	Easy questions	Total
Marks (%)	20	60	20	100

There could be some variations like 15, 70, 15% for difficult, average and easy questions or 25, 50, 25% depending upon previous standard and experience. In school examinations judgment can better be made by teachers to fix difficulty level of the paper in accordance with the concurrent performance standard, objectives to be tested and the homogeneity or heterogeneity of the group. In board examinations, however, differential level of difficulty of questions has to be maintained to reflect and assume normal distribution curve of expected achievement.

8.1.7 Section-wise Distribution of Time and Marks

Whenever paper setters are not well trained in the technology of paper setting, it is all the more important to stipulate some restrictions on the number of questions and give them a framework of distribution of time. This is still more essential when there are two different sections, out of which one is of objective questions to be administered separately. What type of questions can be included in each sections and what sort of distribution of marks of different forms of questions could be is indicated in the design as shown in Table 1

Table 6 : Distribution of time and marks

Section	Form of questions	No. of questions	Marks allotted	Form-wise marks	Section-wise marks	Time
Section A	• Objective type	20	20 × 1	20 ⎫	30	30 minutes
	• Very short answer	10	10 × 1	10 ⎭		
Section B	• Short answer	10	10 × 4	40 ⎫	70	2.30 hours
	• Essay type	03	3 × 10	30 ⎭		
	Total	43	30 + 70	100	100	3 hours

Sometimes even the time for various types of questions can also be indicated to guide the paper setters to develop questions in terms of time estimated for various types of questions. Such a table provides direction to paper setters to dovetail each question of particular form to the intended time needed to be spent on answering a question, thereby ensuring that question paper remains neither under-timed nor over-timed but well balanced in time.

8.1.8 Flexibility in Designs

Design includes all the decisions that are to be taken by the examining agency like the school boards to be stipulated for paper setters to plan their question papers within the framework provided in the form suggested in Tables 1 to 7. sThese tables are only suggestive for illustrative purpose. Who is to prepare the design ? In schools it is the school teachers of particular subjects who are to prepare the design. In all external examinations it is the examining agency like the school boards who would provide the paper setters with the design, on whose basis paper setters construct the question papers. There is a lot of flexibility in the weightages to be given to content units, assessment objectives, form of questions etc. from board to board. Such weightages would also change from time to time when new syllabi are developed, or other changes are to be introduced with reference to assessment objectives, form of questions, setting of standards etc.

> *Notes* Comparative study of designs question of papers of various subjects for different boards would give us an insight into the nature and quality of expected attainment levels in different subjects at secondary and senior secondary stages.

8.2 Developing the Blueprint of Question Paper

8.2.1 Concept of Blueprint

Once the design is ready, the next step is to develop blueprint by the paper setter. A blueprint is a three-dimensional chart, showing distribution of questions reflecting numerical weightages in terms of emphasis to be given to : (i) different units or topics of examination syllabus, (ii) assessment objectives (K-U-A etc.), and (iii) forms of questions. These weightages are reflected strictly in conformity with the prescribed weightage in the design. Some people develop two-dimensional grid, indicating distribution of questions corresponding to various units of syllabus, testing different objectives. If Bloom's taxonomy is followed, all the six objectives, viz. knowledge, comprehension, application, analysis, synthesis and evaluation, are to be given proportionate weightage. But in Indian context after decades of work and studies in the NCERT the use of

three-tier taxonomy of knowledge, understanding and application has become popular and functional at school-level and board-level examinations. Of course, the fourth objective of psychomotor skill (mainly drawing) is also testable in our written examinations, whereas knowledge, comprehensive and written expression form the three assessment objectives in language subjects.

If a blueprint is prepared with care, it reveals the complete horoscope of the 'question paper to be', indicating the various 'nuts and bolts' of the question paper. How best to develop a blueprint and translate the design into an action plan is not only a technical task but also a skill requiring academic acumen and insight into the teaching of the subject. Paper setters must have a clear vision of the nature and hierarchy of content elements and the intended learning outcomes of each unit, to take decision about the nature and type of questions to be included in a particular unit and the related objectives of assessment.

Did u know? Besides reflecting weightages to content units, assessment objectives and forms of questions, a blueprint can also reflect the scheme of options and nature of sections or parts of the paper.

8.2.2 Technique of Development

Assuming an EVS question paper of 3 *hr* duration carrying 100 marks for class VI comprising 8 units and 4 assessment objectives and use of 4 types of questions (ETQ, SAQ, VSQ and OTQ), a blueprint is prepared as given in Table. The small letters a, b, c, ...j given in the blueprint are not a part of blueprint but are shown for guidance of teachers to follow the sequential guidelines given below.

In this table, 'a' indicates the first step to be followed while preparing a good blue-print (of course based on the design provided or developed), followed by 'b', the second step; 'c', the third step; and so on.

8.2.3 Steps Involved in Developing Blueprint

(a) Insert total marks (100) at the bottom right-hand corner box.

(b) Insert column-wise marks to various objectives (40-30-20-10), as per design.

(c) Insert marks against each unit, row-wise (4, 12, 8, 20, 12, 24, 10, 10).

(d) Enter ETQ where suitable $\begin{pmatrix} (\text{U-2Kn}) \\ 10 \end{pmatrix}$ $\begin{pmatrix} (\text{U-4-Un}) \\ (8+2) \end{pmatrix}$ $\begin{pmatrix} \text{U-6Kn} \\ 10 \end{pmatrix}$

with internal option in;

(e) Enter questions testing skill (U-4 part of ETQ-2 marks for diagram), U-5 (4 marks,) for skill and U-8 (4 marks) for skill

Table : Development by blueprint of the question paper   **Notes**

Blueprint

(bassd on prescribed design)

Subject : EVS   Class : VI

Total marks : 100   Time : 3 *hr*

Assessment Objectives		Knowledge 1 Kn				Understanding 2 Un				Application 3 Ap				Skill 4 Sk		5
Unit No.	Form of Questions	E	S	VS	O	E	S	VS	O	E	S	VS	O	E	S	Row wise Totals
U-1	Looking at Maps.			1(1) j	1(1) j			1(1) i					1(1) f			4 c
U-2	The Indian Ocean	10(1) d		1(1) j				1(1) i	i							12 c
U-3	Our Physical Enviro-nment		4(1) j		1(1) j				1(1) i			1(1) g	1(1) f			8 c
U-4	Water					8(1) d	4(1) i		1(1) i		4(1) h		1(1) f		2*(−) c	20 c
U-5	Air				1(1) j				1(1) i		4(1) h		2(2) f		4(1) e	12 c
U-6	Our People	10(1) 10[a](1) d	4(1) j	1(1) j	2(2) j		4(1) i	1(1) i	1(1) i		1(1) g					24 c
U-7	Our Country			2(2) j	2(2) j		4(1) i	1(1) i	1(1) i							10 c
U-8	Living Things								1(1) i		4(1) h		1(1) f		4(1) e	10 c
Sub-Totals Column-wise		20(2)	8(2)	5(5)	7(7)	8(1)	12(3)	4(4)	6(6)	−	12(3)	2(2)	6(6)	2(-)	8(2)	100(43)
Totals Column-wise		40			b	30			b	20			b	10	b	100 a

(f)   Enter OTQ under AP objective (U-1, 3, 4, 5, 8).

(g)   Enter VSQ under AP objective (U-3, 6).

(h)   Enter SAQ under AP objective (U-4, 5, 8).

(i)   Proceed unit-wise entries for OTQ, VSQ, SAQ simultaneously, as shown in :

U-l. VS-O (Kn)-VS(Un)-O(Ap) : U-2- VS(Kn)-VS(Un) :

U-3. S-O (Kn)-O(Un)-VS-O(Ap) : U-4. S-O(Un)-S-O(Ap)

U-5. O(Kn)-O(Un)-S-O(Ap) : U-6. S-VS-O(Kn)-S-VS-O(Un)-VS(Ap)

U-7. VS-O(Kn)-S-VS-O (Un) : U-8.O(Un)-S-O(Ap)

(*i*)   Count the number of E, S, VS, O questions to verify with the design. Rectify if necessary.

(*ii*)   Make subtotals row-wise and column-wise to check weightages to different forms of questions to tally with unit-wise weightage.

(*iii*) Except for ETQ testing, understanding objective—which tests skill objective also (2 marks)—all other questions test only one objective.

(*iv*)   Insert all entries in the same manner—number of questions within brackets and marks for those outside the brackets, e.g. 10(1), 4(1) 2(2) 1(1).

(*v*)   In U-6 two entries 10(1) 10(1)° show two alternative questions, out of which one is to be attempted; 10(1)° indicates option.

(*vi*)   Under column 2 (understanding) against unit 4, ETQ with 10 marks is split into 2 parts, testing understanding (8 marks) and skill (2 marks for diagram).

8.2.4 Judging the Quality of a Blueprint

Once the blueprint is ready, paper setter may like to check whether :

(a)   it reflects the same weightages to assessment objectives, content units and forms of questions as prescribed in the design;

(b)   predisposing conditions, scheme of options, format, and the number and form of questions are met;

(c)   one type of questions is not crowded in a particular unit and under particular objective;

(d)   internal options if provided are balanced in terms of content area, assessment objective, difficulty level and time requirement; and

(e)   on inspection you find that questions of different types are not crowded on one side but are well spread, providing a balanced view of distribution of questions.

Blueprint is not simply a 2- or 3-dimensional chart showing distribution of various forms of questions to be set on different units of content for testing various objectives. It is indeed to be an architectural plan of the paper setter developed within the framework of the prescribed design of the question paper, forming the basis for framing questions. By translating the requirement of the design into an action plan, blueprint provides the needed basis for maintaining validity and reliability of the evaluation tool (question paper). It focuses the attention of the frames of questions to construct various types of questions strictly in accordance with the blueprint. Nevertheless, a good blueprint does not necessarily lead to development of a good question paper; it is the quality of the questions based on the blueprint that determines the quality of question paper, which is our next concern in setting a better question paper.

Self Assessment

1. Fill in the blanks :

(*i*)   _____ is a set of questions or exercises evaluating skills or knowledge of student individually.

(*ii*)   There are three major objectives that are usually tested _____, understanding and _____ objectives

(*iii*) A Comprehansive test like the _____ or an annual examination of the school must utilise all forms of questions.

(*iv*) _____ Can be estimated for each question by paper setters on the basis of their provious experience and expert judgement.

(*v*) A _____ is a three dimensional, chart showing distribution of questions reflecting numerical weightage.

8.3 Framing of Questions and Developing Marking Schemes

8.3.1 Purpose and Scope of Various Questions

The place of various types of questions and their construction has already been discussed. Regarding marking scheme, it should be prepared simultaneously while preparing a question. Variations in marking open-ended questions looms large in grading of essay-type questions. In SAQ also many answers that are acceptable were not intended by the paper setter. In VSA questions again the answers expected by the paper setter are different from the one examinees write. Even sometimes in OTQ the key is doubtful or more than one answer is correct. This happens where questions are not well worded and the scope of intended response is not clear. Unless key, model answer and marking schemes are clearly worked out by making the language of the question precise, unambiguous and comprehensible, we cannot ensure scoring objectivity. Three main purposes of developing model answers and detailed marking schemes are :

(a) to improve quality of the question by dovetailing it with the marking scheme;

(b) to ensure maximum possible objectivity in scoring the response; and

(c) to maximise scoring uniformity among the examiners.

A good marking scheme when developed without making any modification in the wording of a question is indeed an acid test of a good question. Maximum scoring objectively through a well-developed marking scheme by making the needed changes in the wording of original question is the proof of using marking scheme for improving the quality of a question. A functional scheme facilitates examiners to grade responses as objectively as possible. For development of a good marking scheme the following basic principles may be followed.

8.3.2 Basic Principles of Development

Multiple-Choice Questions

(*i*) Is the correct or best answer only one and distinctly one ?

(*ii*) Is it completely correct and not partially correct ?

(*iii*) Is the answer correct under all conditions, i.e. it requires no further qualifiers ?

Short-Answer Questions

(*i*) Word the questions to indicate clearly the scope of intended answer.

Poor : What is parasitism ? (6-7 varieties of answers were observed in one seminar).

(2 marks)

Better : Give one example each of a partial and total parasite. (2 marks)

In the first case workable marking scheme is not possible whereas in the second the intend scope of answer is quite clear and it lends to objective scoring.

(*ii*) Write the answer yourself as you expect the examinees to write. Work out all acceptable answers considered relevant.

(*iii*) Identify the main value points or credit points that you intend to test. Write them in sequential order if one exists. These points form the basis of marking.

(*iv*) Allocate credit-wise (value point-wise) marks proportionately according to the potential of different credit points.

(*v*) Allocate marks to question in terms of value points, which should be functionally operative. (3 marks for a SAQ with two credit points is not workable)

8.3.3 Development of Outline Answers and Marking Schemes

(*i*) Varieties of ETQ

In ETQ it is very difficult to outline answer and marking scheme because of unwarranted extent of variations in developing the marking scheme. The following varieties of ETQ are in use :

(*i*) Long answer with extended response (ETQ)

(*ii*) Long answer with restricted response (ETQ)

(*iii*) Structured essay-type questions (SEQ)

(*iv*) Diagram-based, essay-type questions like picture composition (DEQ)

Marking scheme will be more flexible in (i) and (iv) than in (ii) and (iii).

(*ii*) Basic Principles of Developing a Marking Scheme of ETQ

(*i*) Develop a comprehensive but flexible marking scheme to include all possible acceptable responses considered desirable.

(*ii*) Use the needed criteria on whose basis answers could be rated on a grading system of 3, 4, 5 points representing qualitative evaluation, followed by numerical marking using proportionate weightage to the criteria used.

(*iii*) Develop relevant procedure or mechanics to score a question as objectively as possible by ensuring better comparison of responses by markers (e.g. by marking one question by the same examiner and second question of all scripts by another examiner).

(*iv*) In DEQ develop criteria of assessment relating to skill testing (accuracy, symmetry, neatness, display of contents etc.) and the related functional understanding (accuracy of details, correct and methodical labelling etc.)

(*v*) For SEQ, guidelines as described in under SEQ may be followed.

(*iii*) Training of Markers

With a few exceptions, the markers, coordinators, head examiners or quality controllers are seldom given training in marking. Basic qualifications and teaching experience are usually considered adequate to ensure uniformity in scoring and the needed scoring objectivity. The practice of training the examiner is necessary. More than desired number of markers, quality controllers or head examiners and coordinators should be invited to such training programme, which consists of these major activities.

(*iv*) Developing and Finalising Criteria

(*i*) On the basis of a parallel model paper, the format of the question paper is explained.

(*ii*) Trainees are asked to write answers to each question one by one in the paper, as a student.

(*iii*) Key of MCQ is checked after discussion, as there may sometimes be some problem in the key.

(*iv*) In open-ended questions, participants' answers are written on the blackboard and each response is discussed to arrive at the final list of acceptable answers.

(*v*) Writing of marks to be awarded to correct and partially correct answers.

(*vi*) Listing even the non-relevant and non-correct answers to be awarded zero marks.

(*vii*) Deciding the mode of deducting marks (if any) for mistakes in spelling etc.

(*viii*) Deciding how to treat the irrelevant, redundant and unwanted responses.

(*v*) **Practice Marking**

Once the marking scheme is finalised, it is printed and given to participants the next day. Trainees are asked to mark the given number of scripts using the marking scheme finalised the previous day. They work in different groups of 5 to 10, which are formed in advance, keeping in view the number of questions. If 10 questions are there, 10 people sit round a table and start marking question-wise. Each marker marks the same one question in all the scripts. The subject coordinator coordinates this work. Whole group meets again to sort out the problem (if any) of marking a particular question.

(*vi*) **Selection and Appointment of Markers**

On the basis of quality of marking, the Mauritius Examination Syndicate rejects : (a) some people who cannot be appointed as markers, (b) selects some as markers, and (c) appoints the best one as quality controllers for different groups. This is done about a month before the actual examination. Thus through well-designed activity yielding empirical evidence about the markers, appointments of markers (examiners) and quality controllers (head examiners) are made.

(*vii*) **Illustrations of Marking Scheme**

   (*i*) VSQ

   Q. Complete the sentence using the right form of the verb given in brackets.

   Nowhere......I ... such beautiful temples. (see)

   Marking scheme provides the following acceptable answers :

   had I seen/have I seen/would I see/will I see

   (*ii*) SAQ

   Diagram (Figure 1) shows an electric circuit. Observe it carefully and then answer Q. 1.
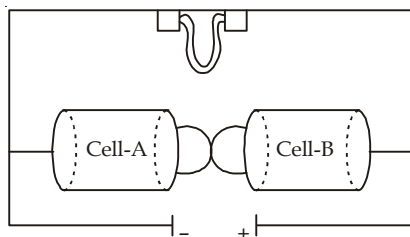
Figure 1 : Electric circuit using two batteries

Q 1. The bulb in the circuit will not light up. Why ?   (2 marks)

   (a) Completely acceptable answers—2 marks (full marks)

   (*i*) Two + ve ends are joined together.

   (*ii*) Similar ends are joined together.

   (*iii*) One end of the bulb is not joined to + ve end of one battery.

   (*iv*) Two ends of batteries are not put in right order.

   (*v*) The circuit is not correctly connected.

  (b) Partially acceptable answers   (1 mark to be awarded)

   (*i*) Circuit is not complete.

   (*ii*) Two batteries are not placed properly.

Notes

(c) Non-acceptable or incorrect answers

    (i)   There is no switch for connection.

   (ii)   Batteries are not charged.

  (iii)   There could be more similar unacceptable or incorrect answers.

Q. 2. What are the reasons that many textile factories are located in rural areas of Mauritius?

(10 marks)

Table 1 : Development of marking scheme

S. no.	Value or credit point	Part-wise marks	
1.	Government policy of decentralisation	1	
	(a) Creation of industrial estates	½	2
	(b) Facilities like building are given	½	
2.	Job creation or labour access	1	
	(a) Employment opportunities	½	2
	(b) Meets demands of foreign investors	½	
3.	Infrastructure	1	
	(a) Better facility for water and electricity	½	2
	(b) Good road network etc.	½	
4.	Land availability	1	
	(a) Lower cost of land	½	2
	(b) Non-availability of land in town	½	
5.	No municipal taxes	1	2
	(a) Octroi tax	½	
	(b) Other taxes	½	2
6.	Rural development	1	
	(a) Increase in income	½	2
	(b) Better quality of life	½	
7.	Reduce worker community	1	
	(a) Solves transport problem	½	2
	(b) Reduced travel to work	½	

Total (10 marks)

Any five reasons like the ones given may be considered correct and marks awarded accordingly.

(Illustration adapted from SAQ of MES to make it an ETQ)

The proof of a good marking scheme lies in :

(a)   improving the quality of a question;

(b)   defining the nature and scope of expected answer;

(c)   delineating the value on credit points;

(d)   indicating fully or partially the acceptable responses;

(e)   reflecting non-answers;

(f)   allocating part-wise or credit-wise marks;

(g)   mode of deduction of marks for some omissions; and

(h)   any other direction that facilitates scoring or enhances scoring objectivity.

Importance of developing a marking scheme along with each question can be appreciated further if as a paper setter you write a complete answer and ask 3-5 good students to write the answer and get it examined by 2-3 examiners of good standing. Quality of a question is judged independently by its own attributes and the extent to which it yields the intended responses when used in a given context. However, it is also true that development of a functional marking scheme for a given question that ensures the needed uniformity and objectivity in scoring is the proof of a quality question.

Task   What is practice marking ?

8.4 Formating, Question-wise analysis and review

### Consolidation of Questions

Questions are to be grouped in a manner to give a particular format that facilitates administration, scoring, analysing and interpreting the scores. It may include the following :

(*i*)   Consolidation of questions into different groups like MCQ, VSQ, SAQ and ETQ, and sequencing of questions within each group.

(*ii*)   Arranging questions objective-wise and content-wise within each form of questions, depending upon the intended mode of analysis of results.

(*iii*)   Preparing general instructions for examinees, indicating the number of questions to be attempted, mode of attempting MCQ and other relevant instructions. All instructions must be relevant, adequate, unambiguous and precisely worded.

(*iv*)   Preparing specific directions for different groups of questions along with mode of attempting, and indicating answers and options, if any.

(*v*)   Coding of paper, subject, year, marks, time, papers I or II etc. For grouping of questions, and providing for in-built mechanism to improve readability of the paper.

(*vi*)   Ensure that within time limit the examinees are able to attempt maximum number of questions to score maximum marks.

(*vii*)   Persuade examinees to attempt at a stretch one particular group of questions requiring the same or similar mental set.

(*viii*)   Form-wise grouping of questions helps in administration and mechanics of writing. Content-wise grouping helps the examinees attempt and concentrate at one time on the same type of content, but the problem of administration of OTQ and openended questions remains. Objective-wise grouping is good for analysis of objective-wise performance, but the disadvantage of administration of objective and open-ended question still remain.

(*ix*)   When 20% or more OTQ are included which are speeded tests, separate administration of OTQ and non-OTQ becomes almost a necessity and can be accepted as a more usable classification.

(*x*)   Within a particular question form follow most familiar to less familiar forms, lower-level content elements to higher-order content elements and lower-level objective (knowledge) to higher-order assessment objective. (Application)

Thus we find that the proper arrangement and grouping of questions should take care of : (a) administrative mechanics of usability, (b) homogeneity and content hierarchy, and (c) complexity of assessment objectives, i.e. from simple and easy knowledge-level questions to more difficult questions of application level.

The following diagrammatic sketch indicates the order and grouping of questions that may be useful to follow.
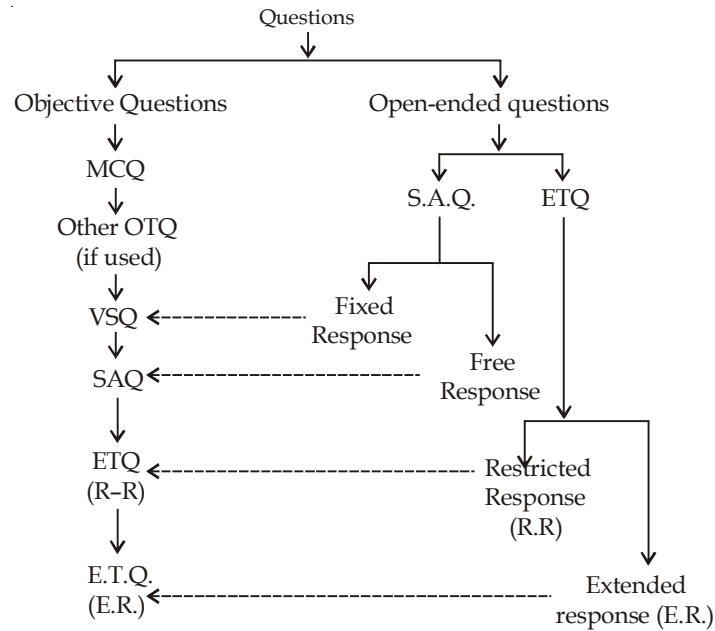
8.4.1 Number of Questions and Marks

The purpose of numbering the question and marks is to facilitate examinees to attempt questions in sequential order and the examiners to indicate marks more easily on the title cover of the answer book and proforma provided. Numbering is usually numerical, which may be Roman numerals or Arabic numerals. Use of alphabets (a, b, c) or small Roman numerals (i), (ii), (iii)... is possible when grouping is done at two or three levels within the same question paper. Different patterns are usable :

Pattern I	MCQ	VSQ		SAQ	ETQ	Each question is numbered
	Q.l-20	21-30		31-40	41-43	in arabic numerals
Pattern II	Q. 1	Q.2	Q.3	Q.4	Q. 5-7	Each question under 1-5
	All OTQ	All VSQ	SAQ	SAQ	ETQ	or I-V can be numbered as
	(20)	(10)	(5)	(5)	(3)	a, b, c; i, ii, iii or 1.1, 1.2, 1.3

For allocation of marks, each question has to be given marks against it if differential marks are given for various questions. If marks to each question in a set or subset are the same, common direction for marks may be given or total marks for that set can be given, e.g. 10 marks for a set of 5 SAQ.

8.4.2 Preparation General Instructions

(*i*)   Instructions for the whole paper. (All questions are compulsory etc.)

(*ii*) Instructions for one section of OTQ. (In Q. 1-20 only one answer is correct; encircle the serial number against the correct response)

(*iii*) Instruction or direction in the question itself. [Describe in not more than 100 words the method of preparing oxygen gas in the laboratory, with the help of a diagram. (Directional word, length, diagram)

(*iv*) Are instructions relevant, unwarranted or redundant to the question paper ?

(*v*) Are instructions adequate, inadequate or omit significant instructions ? (negative marking)

(*vi*) Are instructions clear, unambiguous, comprehensible and worded properly ?

8.4.3 Question-wise Analysis and Review

Though question paper is now ready, it is better to have final look at it. As a wrapping up exercise it is better to undertake quick review by question-wise check with reference to different attributes of a question using a proforma shown in Table 1.

Table 1 : Question-wise analysis

	(a)	(b)	(c)	(d)	(e)	(f)
Q. no.	Content area	Content unit	Assessment objective	Competency tested	Form of question	Wording of question
1.	Anatomy	Stem	Understanding	Interpret	MCQ	Definite
2.	Physiology	Osmosis	Application	Predict	SAQ	Appropriate
3.	Evolution	Evidence	Knowledge	Recall	ETQ	Indefinite

	(g)	(h)	(i)	(j)	(k)
Q. no.	Language used	Estimated difficulty level	Time estimate	Marks allotted	Remarks
1.	Unambiguous	Average	1 *min*	1	....
2.	Precise	Difficult	5 *min*	3	....
3.	Ambiguous	Easy	20 *min*	10	Directional word need to be changed
⋮	⋮				

Like this all questions may be analysed and data recorded. Each of these entries as to checked, as shown below.

(a) Is the content tested the same as indicated in blueprint ?

(b) Is the unit tested the same as indicated in blueprint ?

(c) Does the question really test the intended objective ?

(d) Does the question measure the intended competency ? (interpret)

(e) Is the question set in the same form as shown in blueprint ?

(f) Does wording evoke the intended response ?

(g) Does language admit the scope of answer ?

(h) Is question easy, average or difficult ?

(i) What approximate time is required to attempt ?

(j) How many marks are allotted ?

This analysis should be done either by an expert other than who set the question paper or by the paper setter himself, without seeing the individual item sheet or item card on which questions were written or recorded. Only after this proforma is filled up can the analysis be compared with the entries in the blueprint. If there is any discrepancy, the question can be modified to fit with the blueprint. As for wording, language and other attributes, the needed changes or modifications may be made. Inferences like the following may now be made :

(*i*) Is there any question that does not test the intended objective and needs revision ?

(*ii*) Are various competencies well sampled ?

(*iii*) Is there any discrepancy in reflecting the stipulated form of question between the intended and the observed ?

(*iv*) Do some questions need improvement in wording and language ?

(*v*) Is there appreciable difference in the percentage of easy, average and difficult questions from the one indicated in the design ?

(*vi*) Does the total time allotted as estimated in this analysis conforms to the time allotted in the paper ?

Review by question-wise analysis of the whole paper is the final task for paper setter to check whether all questions reflect the intended attribute indicated in the blueprint and item sheet that characterises the item on all those attributes. Congruence between the intended and observed attributes shows the quality of the question paper.

8.5 Moderating the test

Moderation means re-examination of the questions, key, marking scheme, model answers, instructions or directions for the examinees besides the format and mechanics of paper setting. The purpose is to ensure quality, relevance, reliability and practicability of the question paper in a given situation, i.e. school-based examinations or public examinations like those of the boards. Moderation is done to improve the quality of the question paper in terms of requirements reflected in the design and the blueprint of the question paper. Moderation is necessary, as some omissions or commissions might have been inadvertently crept in due to personal bias or lack of some technical know-how. It is therefore essential to moderate it before the paper is accepted as final for print to make it a more valid and reliable instrument.

8.5.1 Nature and Scope

Moderation is not only a technical process that demands expertise in setting good question paper, but also critical acumen to edit, form rational judgment and take right decisions. Therefore proper selection of experienced subject experts is essential, who are well conversant with the technology of paper setting and instructional objectives.

It demands ensuring the right standards, relevance to conditions stipulated by examining agency and its practicability in terms of administration, scoring and processing of results. Moderation may be pre-examination or post-examination, concerning moderation of marks and processing of results. Here we are concerned only with the pre-examination moderation to improve the quality of question papers. Thus the task involved in moderation of question papers relates to the moderation of :

(a) questions;

(b) key or outline answers;

(c)   marking schemes;

(d)   instructions;

(e)   format and organisational pattern of questions;

(f)   language and wording of questions; and

(g)   mechanics of writing the questions.

⚠️ *Caution* A moderator who is not well trained in the technology of paper setting and role of instructional objectives may spoil a good question paper set by a trained evaluator.

8.5.2 Moderators

(a)   School-based Examinations

In term tests or annual examinations conducted by the school the moderation can be done in different ways.

Moderation by Paper Setter Himself : Paper setter himself is the moderator and therefore it is a quick review rather than a thorough moderation. Quality-control process lies with the producer of the question paper. It is acceptable in unit testing and other class tests, where the purpose of testing is formative and not summative. To improve the situation a check-list can be prepared with the help of experts, for use in moderating the tests.

Moderation by a Co-class Teacher : Here the subject teacher of another section or class can moderate the paper. Similar check list can be used by him for moderation. This brings some externality in moderating the paper. Advantage of involving two teachers teaching the same class or section is to utilise their experiences of teaching the same class.

Moderation by Senior Subject Teacher : A senior teacher of senior secondary classes can moderate the paper of another teacher teaching the middle or secondary classes. Senior teacher can discuss, ask for clarification, and get modifications accepted through mutual discussions. Moreover, he acts as an external moderator.

8.5.3 External Public Examinations

Board examinations have high stakes of confidentiality because of significant role the public examinations play in certification and selection of students for admissions and scholarships. This requires appointment of external moderators more judiciously. For selection of moderator in such external examinations, the moderator should possess the following qualities :

(*i*)   must be a different person from that of a paper setter;

(*ii*)   must be a subject specialist of long standing :

(*iii*)   well conversant with the examination syllabus and instructional objectives;

(*iv*)   trained in technology of paper setting; and

(*v*)   a good item writer and item critic of proven ability.

8.5.4 Panel of Moderators

Sometimes a panel of 2 or 3 moderators is recommended when secrecy is not at stake. In that case:

(*i*)   subject specialist reviews the subject matter at the first-stage;

(*ii*)   evaluation expert reviews the items to judge quality of items at the second stage; and

(*iii*) language expert undertakes editorial review to ensure correct use of language, style, format and mechanics of writing, at the third stage.

A moderator may spoil a good question paper if he does not know what a good question paper is and how to develop one.

8.5.5 Tasks Implied in Moderation

Some moderators think that their job is to go through the paper and make some inevitable editorial corrections to make the paper print-worthy. Other moderators think they can make any change or replace any question they like, being a senior specialist in that subject. This is a wrong notion with most of the moderators. They think it is 2-3 *hr* job and their stamp is sufficient to make the paper perfect. The following guidelines if followed in letter and spirit would help them moderate in a scientific way :

(a) Study the design prescribed by the board to understand the basic framework and predisposing factors.

(b) Verify the blueprint prepared by paper setter to judge if it conforms to the prescribed design. If not, carry out necessary changes to make it consistent with the design.

(c) Verify each question in terms of blueprint and entries, and apprise yourself whether or not the question is based on the same content unit, testing the same objective, using the same form of question.

(d) Judge each question in terms of other criteria applicable to a good question, as indicated in the proforma of question-wise analysis.

(e) Check key, and outline answers and marking scheme in terms of the intended response to ensure uniformity in scoring and scoring objectivity.

(f) Judge the suitability of organisational pattern of consolidating questions with reference to sections, directions, lay-out, mode of indicating answers etc.

Self Assessment

2. Multiple choice Questions

Choose the correct option

(*i*) Since unwarranted extent of variations in developing the marking scheme, it is very difficult to outline answer and marking scheme in _____

    (a) Essay type question   (b) Short answer type question

    (c) Multiple choice type question   (d) Objective type question

(*ii*) In _____ participants answer are written on the black board and each response is discussed to arive at the final list of acceptable answers.

    (a) Essay type question   (b) Open ended question

    (c) Multiple choice questions   (d) Objective type questions

(*iii*) _____ is not only a technical process that demands expertise in setting good question paper, but also critical acumen to edit, form rational judgement and take right decisions

    (a) test   (b) Evaluation   (c) moderation   (d) Examination

8.6 Summary

- Setting of good question papers is not everybody's cup of tea. It requires a long, consistent practice of writing, criticising, editing and improving the items. It demands thorough knowledge of the criteria of a good question paper (test), i.e. the technical qualities like

validity, reliability and usability, which make an evaluation instrument dependable for ranking, selecting and certification of students' achievement. Thorough understanding of technology of paper setting involves preparation of a functional design of question paper that provides framework for setting the question paper. Such a framework includes identification of content units or areas of syllabus, assessment objectives to be included for testing, form of questions to be used, scheme of sections if any, and difficulty level of question type of options to be provided, besides taking cognizance of some predisposing factors like total time, total marks, number of papers, assumption to be made etc. Giving proportional weightage to various content areas, objectives and forms of questions is also necessary in designing. The next step is development of blueprint conforming to the prescribed design by the board or the examining agency. This blueprint provides distribution of questions over various content units in relation to assessment objective, along with indication of marks. Then it is necessary to frame the questions in conformity with the blueprint requirement, along with key, outline answers and marking scheme. Consolidation of questions to give question paper a particular format and organisational pattern is followed. Writing of general instructions and directions for various parts comes next. Editing and review of questions is undertaken by means of question-wise analysis, on whose basis the needed modifications and improvements in questions are made and paper is finalised to make it press-worthy. However, for a final look an expert review is needed to ensure further improvements through moderation, which is undertaken by different subject experts and trained evaluators. Sometimes a panel of paper setters and moderators is used to make question papers more valid, reliable and usable through various modifications and improvements made by the moderators, who do a final review to make the question paper perfect. Thus paper setting is a long-drawn process that demands technical know-how, hard labour and critical acumen.

8.7 Keywords

Construction   :   The process or method of building or making something

Predispos   :   to influence somebody so that they are likely to think or behave in a particular way

Blue Print   :   A plan which shows what can be achieved and how it can be achieved

8.8 Review Questions

1. Explain the Concept of test and predisposing factors which affect the test.

2. What is blue print of test. Give the steps involved in developing blue print.

3. What is the purpose and scope of framing the questions of various types and their marking?

4. How is it analysed the test ?

5. Explain the process of moderation of the question paper.

Answers : Self Assessment

1. (*i*)   Test   (*ii*)   knowledge, application   (*iii*) board's external examination

    (*iv*)   difficult level of test   (*v*) blue print

2. (*i*) a   (*ii*) b   (*iii*) c

8.9 Further Readings

Books

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 9 : Test Standardization

Objectives

The objectives of this unit can be summarised as below :

- To describe about the Construction and Validation of items

- To explain about the equating of two or more forms of test

- To discuss the derivation of test norms

- To describe about the establishment of final validity and reliability

- To explain the practical uses of standardised test

- To describe about the planning the testing programme

- To discuss about the limitations of standardised tests

Introduction

We have already discussed the Construction of an achievement test used in school-based assessments and also in external examinations. Different steps like designing, blueprinting, construction of questions, formating the test and moderation of the question paper were described for developing good question papers. When we talk of standardisation of an achievement test, it is the process of desiring comparative norms that distinguishes the formal standardised test from the informal objective test of achievement. But any programme of standardisation demands additional requirements like :

(*i*)   more critical analysis of subject-matter;

(*ii*) more careful formulation of test material;

(*iii*) more exacting refinement of test items;

(*iv*) more critical standards of test forms and quality of items;

(*v*) more rigid statistical analysis than the usual informal objective tests; and

(*vi*) derivation of the set of norms.

9.1 Construction and Validation of Items

One problem for developers of standardised test is the examination of content or syllabus and the group of students they have not specifically taught. Therefore the developers must be reasonably sure that the content selected for test items is likely the one that has received instructional emphasis. Selection of content must be general enough to fit into any school situation where the course is taught. The difficulty arises in selection of content due to the nature of knowledge, understandings, skills, concepts or applications to be tested. In subjects where instructional objectives are clearly stated in terms of intended learning outcomes, it is easier to develop test items that sample the content adequately, as in mathematics where facts and skills are well known. In other fields, where knowledge, skills, attitudes and other outcomes are of more indefinite nature, it is more difficult to validate the test items. Validity of test content in most of the subjects is difficult to establish by acceptable objective means and statistical methods. It is practically impossible in some cases. Particular validation procedures may be effective and acceptable in one subject and completely unsuitable in another. That is why developers of standardised tests have used different types of validation procedures in different subjects.

Selection of Test Items

Principles and guidelines for constructing objective test items were discussed. Therefore methods used for construction of items for standardised tests and actual construction of test items on different outcomes of learning are briefly discussed here.

(*i*) **Validity of a Test : It depends on :** (a) validity of the content in general, and (b) validity of the individual test items of which the test is composed of. Validity of the items before they are actually administered depends on the ability of the test constructor to :

(*i*) select the right form of the objective items; and

(*ii*) skill to construct the items on the pre-decided intended learning outcomes identified while designing the test, besides avoiding the various types of weaknesses highlighted in the earlier chapter on construction of objective-type questions.

Real evidence about item validities is found only by :

(*i*) actual administration of the test in preliminary form to a large group of typical pupils, representative of the population, and

(*ii*) detailed statistical analysis of results item-wise.

On the basis of this analysis many items are rejected, modified, revised or replaced. For this reason in the preliminary form of the test there may be many more items than envisaged in the final form.

(*ii*) **Objectivity :** It is so important an element in the reliability of measurement that one can-not think of a standardised test made up of items which are not characterised by objectivity. Usually objectivity is determined by the form of test item the framer uses. Though the precise form of objective technique that fits best the subjects to be tested is difficult, it is usually taken care of by experimentation and previous experience.

(*iii*) **Content Analysis :** It is the next problem once instructional area to be covered is decided. It refers to the content elements like terms, facts, concepts, principles, processes and other generalisations or intended outcomes that form the basis of item writing. These elements are to be stated in some objective form that reflects the nature of instructional objectives or

intended outcomes of the course, keeping in view the maturity level of pupils being tested and the best item type or item form that suits best for the content elements.

Three other factors that characterise the objective test item.

(*iv*) Ensuring uniformity of response

Each test item be worded in precise, unambiguous and comprehensible language so that the intended response is clearly demonstrated. This is particularly necessary when supply-type items are used. It is difficult to expect uniformity in responses, if distractors in multiple-choice item are not plausible; the task is not properly set in the stem of the item; key is doubtful; more than one answer is possible when needed qualifiers are not used in the stem etc.

(*v*) Avoiding clues and suggestions

These clues are found in the stem of the item, in the key and even in distractors. Associational clues, verbal clues and determiner clues are commonly found. Use of prefixes like an, un; article clue (like a, an, the) and use of determiners like always, never, seldom etc are to be avoided.

(*vi*) Freedom from ambiguity

Use of ambiguous language, imprecise wording, unfamiliar vocabulary, vague directional words, incongruent mechanics of writing the item that leads to grammatically incorrect statements (stem of the item not tallying well with each option)—all lead to ambiguity in terms of intended response.

(*vii*) Reasonable Difficulty Level : Determination of optimum difficulty level is a serious problem, on which the experts do not agree. However consensus is that test as a whole should have about 50% difficulty for average pupil.

Likewise, brighter section of pupils need not waste more time on easy items, which are in the beginning. Thus gradual, continuous difficulty of items if made the basis of scaling of items in the test helps the less-abled and the brighter students to score the maximum.

Did u know? Modern practice of arranging test items in a standardised test is to cover wide range of difficulty in ascending order from easy to difficult. It ensures less-abled pupils to attempt maximum number of items before encountering more and more difficult items.

(*viii*) Acceptable Discriminating Power : Discriminating power of a test item refers to the quality or magnitude of response that may be expected from individuals along a defined scale in accordance with difference in their achievement due to varying degrees of abilities. In other words, superior-ability pupils should answer the item correctly more often than those with inferior ability. This suggests a method by which the power of a test item to discriminate between groups of pupils may be determined. For calculating the discriminating index (D.I.), Kelley used the method of grouping pupils on the basis of scores into three grades. Upper group of 27% making the highest scores, lower group of 27% making the lowest scores and the middle groups of 46%, which are not considered for calculating D.I. The next step is the count of all test items in the test. For each item the number of students from upper or higher group (27%) and those from lower group (27%) can be counted who attempted the item correctly, as shown in Table 1. D.I. can be calculated by the formula :

$$\text{D. I.} = \frac{RH - RL}{NH}$$

where RH = the number of students attempting the item correctly from the higher group (10); RL = the number of students attempting the item correctly from the lower group (4); NH = the number of students in the High or Low group (27).

$$D1 = \frac{10-4}{27} = \frac{6}{27} = \frac{2}{9} = 0.22$$

Examine Table 1. of analysis showing discrimination indices of items.

Table 1 : Analysis showing discrimination indices of items

Item no. group (27%)	Higher group (27%)	Lower Index	Discrimination
1	10	4	0.22
9	6	4	0.08
20	8	13	– 0.18
32	9	16	– 0.26
44	21	12	0.33
58	20	20	0.00
69	25	22	0.11
75	24	9	0.55
86	23	11	0.44
100	27	9	0.66

(*i*) Item 1 is difficult whereas item 9 is still more difficult, attempted correctly by only 10 students and both have limited power of discrimination (0.22 and 0.08)

(*ii*) Items nos 20 and 32 are also difficult, attempted by 21 and 25 pupils respectively.

In both cases however more number of pupils attempted the item correctly from lower group than those from upper group, which show negative indices of –0.18 and –0.26. There must be some defect in these items, which should not be retained in the test. Item no. 58 has 0 D.I., as equal number of pupils attempted this item correctly, from upper and lower groups. Items 75,86 and 100 have positive indices of 0.55,0.44 and 0.66 and are probably good enough to be retained in the test. This method of determining discriminating power of test items is widely used in critical analysis of test items for standardised tests.

9.2 Equating of Two or More Forms of Test

While standardising the test, two or more parallel forms are to be developed. These forms are considered equated where the same pupils or different pupils of the same ability make identical scores on each test form. This means that the test items in both forms of the test are closely parallel in terms of skills and abilities tested and the difficulty level. Equality of item difficulty in alternate forms can be obtained by means of following ways :

1. Prepare large number of items, 2 or 3 times the desired number, covering the entire range of intended outcomes to be tested. Suppose, the sufficient number of items at different difficulty level may permit pairing of items of equivalent difficulty level in the alternate forms of the test. Such forms of the test could be taken roughly equivalent in difficulty, though ensuring only general and limited equivalence of content.

2. Develop two or more test items on the same concept, principle, law or generalisation. In fact when a blueprint is developed on the basis of design of the test, it provides good basis to prepare two or more items of the same form, using the same concept, testing the same objective or learning outcomes, as reflected in such a table of specification. This ensures much better equivalence of items.

3. Another procedure is to use derived scores for establishing comparable forms of tests; though the complexity of statistical techniques makes it impracticable at this point. There are widely used derived scores that have constant meaning, whether or not they are obtained on the same form of the test of from the same pupil group.

9.3 Derivation of Test Norms

Norms are tables of information necessary for interpretation of test scores and are obtained by giving the particular test to a large and representative sample of pupils in the same grades with which teachers will use the test. Establishment of norms that furnish reliable and useful basis for interpretation depends on the extent to which sample used in obtaining the norms is distributed over a large population in typical school situations and the conditions under which tests are to be administered are rigidly followed by teachers using the tests. Norms provide the users of a standardised test a basis for practical interpretation and application of the results. Existence of norm is the most distinctive feature of standardised tests, though not the only characteristic feature.

9.3.1 Types of Norms

The form in which norms for a test are provided depends largely on the level in the school system where the test is used. It is also conditioned by the nature of the test itself. Tests designed for elementary school grades are usually accompanied by age norms and grade norms and also sometimes by percentile norms based on grade placement. Tests for use at secondary stage are more frequently provided by percentile and grade norms only, because the age norms are not considered useful since growth curve at 16th and 17th years appears to flatten out rapidly.

9.3.2 Grade Norms

These are based on median scores obtained by giving the tests to a large groups of pupils within each grade. It is a common but not a universal practice to express these norms in terms of end of the year's achievement. These norms clearly indicate the period they are designed to cover. They help in expressing the progress of pupils through grades by converting their raw scores or standard scores into grade-equivalent scores. If seventh grade end of the year norm of a test was 120 points and the eighth grade end of the year norm is 140 points, then a score of 130 points will be treated as representing achievement half way through 80 grade or 8.5 grade equivalent. In most of the tests composed of several parts, raw scores are frequently changed into standard scores before establishing grade norms (Iowa Language Abilities Test). Raw scores on each subtest are changed into standard scores. Total score on all the parts of the test is represented by the median of the several standard scores.

9.3.3 Age Norms

Age norms appear to provide more adequate basis for the interpretation of individual pupil achievement at elementary school level than is possible with grade norms or percentile grade norms alone. It involves re-grouping of all pupils used in grade tabulation into chronological age groups regardless of the grade location or school progress. Test scores of these chronological age groups are then tabulated and the means or medians computed, which becomes the basis for setting up tables of scores corresponding to several age groups. Factors like overageness, retardation and acceleration do influence the average achievement of pupils grouped in grades. For example,

though average chronological age of a 7th grade student may be 13 years and 6 months at the end of the school year, the average test score of pupils of 13 years and 6 months is not at all the same as the end of the year score for the 7th grade. Actual achievement of underage pupils is significantly superior to that of overage pupils in a given grade. For developers of standardised tests therefore the need for norms that take into account the wide differences in maturity, mental ability or school progress within the grade cannot be undermined, to avoid interpretations that are likely to be misleading.

9.3.4 Age-at-grade Norms

In the process of test standardisation, establishment of age-at-grade norms involves difficulties like : (a) availability of population, and (b) statistical procedures resulting from inadequate population groups. For example, the number of 6th grade pupils who are 10 to 11 years old would represent a large portion of normal 6th grade population, whereas the number of underage pupils in 6th grade who would be 9, 8 or 7 years old and the number of overage pupils who would be 11, 12,13 or 14 years old would fall off rapidly. Therefore in order to get reliable age-at-grade norms for all ages within the grade, very large number of pupils have to be tested to secure adequate population in the fringe areas, failing which estimation by extrapolation has to be made. In the IOWA Basic Skill Test-C, language raw point scores are directly converted into grade-equivalents as shown below :

6th grade : 1st Semester, 13 yr and 6 months old pupil has grade-equivalent score of 4.8

6th grade : 1st Semester, 10 yr and 6 months old pupil has grade-equivalent score of 6.3

6th grade : 2nd Semester, 13 yr and 6 months old pupil has grade-equivalent score of 5.3

(Typical Child) 9th grade : 2nd Semester, 13 yr and 6 months old pupil has grade-equivalent score of 10.1.

It is apparent that age-at-grade norms offer very useful means for interpretation of individual accomplishment, especially for pupils who are over-age or under-age for that grade.

Notes   It is interesting to note that younger pupils within the grade make higher grade-equivalent scores up to a certain point and the older pupils make lower.

9.3.5 Percentile Grade or Subject Norms

This is done by turning raw scores or standard scores into percentile scores by computing percentile values from the frequency tables for each grade and assigning percentile equivalents for each score. Percentile norm labels show a wide sampling of pupils in a certain grade or course : (a) the percentage of pupils exceeding each score or each of a number of equally special scores, or (b) the score below which certain percentages of pupils fall. Percentile scores corresponding to specific raw or standard scores may be reported by grades by test parts and totals, or only the raw scores or standard score equivalents for specified percentiles, quartiles and deciles may be shown in more compact tables. It is suggested that when percentile grade norms are provided, these values should be used in the interpretation of individual pupil scores rather than the grade equivalents from usual grade norms. Overlapping of the distribution of scores from grade to grade on most of the standardised tests at elementary level is quite great so that the differences between successive grade medians are so slight that grade equivalents may exaggerate differences, leading to unsound interpretation. A 7th grade pupil assigned a grade equivalent of 9.5 does not belong to 9th grade. It is more accurate to use percentile scores to describe a pupil's achievement as superior in relation to other 6th grade pupils.

9.3.6 Percentile Norms for School Averages

In the schools participating in state-testing programmes or in school system in a large city it is desirable to interpret result in terms of school averages. Comparison of one school with another school in the same city or one system with another school system (Government, K. V., Public, N. V.) is not possible through usage of norms based on individual pupil scores, because variability of individual scores is much greater than the variability of school averages. Percentile norms for school averages are worked out in the same manner as other percentile norms are derived, except that averages are substituted for individual scores in the grade distribution. Percentile norms used in interpretation of school results should not be used in interpreting individual pupil scores, because it would introduce sessions error in test interpretation by confusing norm for school average with percentile grade norms used in interpreting individual pupil achievement.

9.3.7 Norms vs Standards

The words standards and norms are sometimes used carelessly as synonymous. Standardisation is a process of gathering the data for critical analysis of tests and derivation of suitable norms. The term 'standard' when used to refer to a level of pupil achievement implies the ultimate goal or intended level of achievement towards which to strive, but may not actually be reached by any individual. On the contrary, norms represent the level of achievement that the typical pupils actually attain. The process of securing these comparative scores, known as 'norms', is called normalising.

The standard in arithmetic accuracy is 100% as computation errors are useless. But actual norm of arithmetic accuracy of computation on well-known tests may be from 70 to 75% for high school grades. It means that instead of desired ultimate goal of 100% accuracy, the pupils of these grade work out arithmetic examples with an accuracy of 70 to 75%. Norm does not necessarily reflect satisfactory level of achievement. This is particularly true of schools where instruction and school environment are superior and pupils' intake is from good homes, having superior abilities and other support services.

Standards are of two types. One, these may be minimum essential that are generally accepted by schools, e.g. reading, spelling and arithmetic. They represent minimum quality and perhaps speed of performance that equip the pupils for post-school life. On Ayres Scale of Handwriting, 60 is widely accepted standard at the rate of 70 letters per minute. Though 60 at the given rate is approximately the norm for pupils who completed the sixth grade, it is also taken as standard or minimum ability that should be acquired by all pupils before they finish school. Second, standards in any school subject of pupils' achievement can be formulated, probably subjective or as vaguely conceived idea in the mind of principal or teachers concerning his expectations of the pupils. In this case standards are quite variable and differ from teacher to teacher, school to school and year to year for the same teacher. Modern emphasis on providing each child the type of instruction that suits best to his abilities, interest and his future needs rather than moulding all pupils in the same pattern has reduced the reliance on standards. Attempt is now made to provide maximum aid to each child to develop his potentialities and evaluate his achievement in terms of himself as an individual.

Self Assessment

1. *Fill in the blanks :*

   (*i*) _____ are the tables of information necessary for interpretation of test scores.

  (*ii*) _____ help in expressing the progress of pupils through grades by Converting their raw scores or standard scores it grade equivalent scores.

 (*iii*) _____ of a test item refers to the quality or magnitude of response.

 (*iv*) _____ is a process of gathering data for critical analysis of test and derivation of suitable norms.

9.4 Establishment of Final Validity and Reliability

Once the test is ready and given to representative group of pupils and norms are derived on the basis of scores, final step is to establish validity and reliability of the test. Though preliminary steps taken during construction of test should make reasonably certain to satisfy these criteria, as a final check it is essential to report to users of the test the validity and reliability of these test to enable them to evaluate it.

9.4.1 Validity of the Test

Validity of the test depends on the efficiency with which it measures what it attempts to measure, i.e., it serves the purpose the user has in mind. Tests cannot be considered as valid in general terms, but only in connection with their intended use and at the intended ability level of the pupils. Three types of validity are : (a) curricular validity, (b) statistical validity, and psychological and logical validity.

9.4.2 Curricular Validity

Basis for determining curricular validity are :

(*i*)   Adequate sampling of content or course of study

(*ii*)   Reflection of instructional objectives or outcomes of learning

(*iii*) Recommendations of committees, subject and test specialists.

9.4.3 Statistical Validity

It is calculated by determining correlation between test scores with criterion measure, which is considered as measurement standard. Commonly used methods are :

(*i*)   Correlation with school marks

(*ii*)   Correlation with ratings of expert judges

(*iii*) Correlation with other known concurrent measures, (teachers' ratings etc.)

(*iv*)   Correlation with measures of future outcomes (predictive).

9.4.4 Psychological and Logical Validity

In complex fields such as language and reading and study-skills, subjects are made up of many interrelated abilities, in contrast to practical-skill areas where tested performance is either an exact representation of, or a very similar substitute for, the instructional outcomes sought. Analysis of desired learning outcomes (Intended outcomes) and of the proposed test using logical and psychological methods can reveal sufficient commonality or similarity to justify that the test is a valid measure of the intended outcomes. This method is followed quite frequently in complex fields like language and reading skills or study skills.

9.4.5 Reliability of Test

A test is said to be reliable if it functions consistently or the efficiency with which the test measures what it does measure. What it does test or what it attempts to measure is the 'validity' of the test. How efficiently does a test measure What the user attempts to test (whatever it measures) is the reliability of the test. A test may be reliable without being valid, but it cannot be valid unless it is reliable. Thus reliability is really an aspect or a phase of validity. Reliability has two aspects : adequacy and objectivity. It is usually expressed in terms of coefficient of correlation. It is internal consistency or self-consistency, which is evaluated for estimating the reliability coefficient. The following methods or statistical procedures are used in obtaining various coefficients.

9.4.6 Reliability Coefficient

This requires correlating scores on two equivalent forms of the same test, given simultaneously to the same group of students using the same procedure. This measure is the coefficient of realibility. The following types of coefficients are commonly used.

Retesting Coefficient : When only one form of test is available, it is given to the group of pupils twice under similar testing conditions. Retesting coefficient is the correlation coefficient between two sets of scores. However, second administration should neither follow the first too quickly to avoid significant increase of scores that may result from memory, nor be delayed too much lest forgetting operates to a large extent.

Chance Half Coefficient (Split Half) : The test is given to a group of pupils and their scores are then obtained for two halves of the test. Two halves can be made as :

(*i*) odd and even numbered items; and

(*ii*) obtaining separate scores on items 1, 4, 5, 8, 9, 12, 13 etc. and on items 2, 3, 6, 7, 10, 11 etc. to equalise the difficulty of the two halves when items are scaled in a scaled order of difficulty. Correlation coefficient obtained between the two sets of scores indicates the degree of conformance between the two chance halves of the test. Reliability coefficient of the test is then found by estimating the correlation by using Spearman-Brown Prophecy formula :

$$r_{12} = \frac{2r\frac{1}{2}\frac{1}{2}}{1+r\frac{1}{2}\frac{1}{2}}$$

Foot Rule Coefficient : This may be an underestimate but never overestimate of the reliability coefficient. It is not the most accurate method. It requires use of three facts and measures from the test in a simple formula - the arithmetic means, standard deviation of scores and the number of items in the test. Owing to sufficient accuracy and simplicity, this method is recommended for use by teachers in estimating reliability of their informal objective examinations. The formula used is given below :

$$r\mathrm{H} = \frac{\overline{x}\left(k - \overline{x}\right)}{k\left(\mathrm{SD}^2\right)}$$

where $r\mathrm{H}$ = coefficient correlation;

$\overline{x}$ = mean;

k = number of items in the test; and

SD^2 = variance (Standard deviations squared).

Estimate of reliability coefficient often results in high or low test reliability. It must be based on known and appropriate range of ages or grade placement of pupils if it is to mean what it purports to mean. Hence reliability coefficient is neither an entirely adequate device, nor for that matter the only method of indicating the internal consistency of a test.

Standard Error of Measurement (SEM) : The other popular device by which test reliability can be estimated is the standard error of measurement. Standard error indicates the degree of accuracy existing in the test score, obtained for each pupil on a test.

Here the accuracy refers" to magnitude of sampling errors. Since SEM isnot ecTed by range of talent of the pupil group on which it is based (as in reliability coefficient), it is recognised as a more concrete way of indicating test reliability.

Adequacy and Objectivity in Test Reliability : "Adequacy" is the degree to which test samples sufficiently widely into the subject so that the resulting scores are representative of relative total

performance in the areas measured. If sampling is small, scores are likely to be unfair to some pupils. If sampling is ample or adequate, the scores are likely to be fair to all pupils.

"Objectivity" in a test makes for the elimination of the opinion, bias or judgment of the person who scores it. Both 'adequacy' and objectivity are essential to test reliability and both are therefore aspects or phases of reliability, though independent.

9.5 Practical Uses of Standardised Tests

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.

9.5.1 Instructional Uses of Standardised Tests

(*i*)   For class analysis and diagnosis.

(*ii*)   To gather advance information about proficiency of his class in certain subjects.

(*iii*) To identify weaknesses and strengths in certain areas.

(*iv*)   To ascertain the background of pupils in terms of general competencies and for mastery of specific knowledge outcomes.

(*iv*)   To check progress from time to time if the class moves together or that certain groups of pupil need special attention.

(*vi*)   To furnish justifiable basis for dividing the class into sections for remedial treatment

(*vii*) To make diagnostic use of test results, as basis for remedial teaching and basis for preventive work

---

Task   What is Foot rule Coefficient ?

9.5.2 For Individual Pupil Diagnosis

Besides the deficiency of the class in general there is also possibility of knowing the difficulties of individual pupils along with their specific weaknesses and strengths. Test results should be studied in the light of each individual's attainments and points of difficulty. Critical analysis of test scores of each pupil may reveal unexpected trouble spots, which otherwise would continue to hamper his proper progress. Identification of such hard spots or trouble spots in learning can then become the basis of remedial instruction or remedial material, which may help the pupil in correcting his own weaknesses. In fact development and use of diagnostic tests based on such hard spots of learning reveal the cause of such problems and difficulties, which in turn becomes the basis for remedial instruction and other correctives to be applied. In fact diagnostic profile chart on different skills and abilities can be prepared indicating areas of strengths and weaknesses.

9.5.3 Guidance Use of Standardised Tests

Standardised tests in different areas can be utilized for guiding pupils to select the right subjects in accordance with their mental abilities, aptitudes, interests etc. It is necessary to identify special abilities of pupils and providing challenge to them for greater efforts; directing children away from fields in which they are not interested or have no aptitude. Test records obtained only in the institution enable teachers and principals to help pupil adjust in school environment. Guidance centre serves also as a test bureau. Some of the items in the record will be all types of readily interpreted results of physical and mental tests, aptitude tests, progressive tests, personality inventories, survey, and the analytical and diagnostic achievement test. Guidance specialist must therefore be well trained in the use and interpretation of measurement and evaluation data.

9.5.4 Administrative Uses of Standardised Tests

With the expansion of education at various levels, classes are increasing in size, curricular offerings are expanding, educational costs are going up, and the public interest in educational efficiency is getting more demanding. Parents and teachers are becoming more critical of the methods of evaluating pupils' accomplishment and the marking system. Community demands school survey to know about the efficiency of their schools. Standardised test can be used to establish adequacy of system of assigning and reporting marks. Efficiency of teaching may require experimental evaluation. Efficiency of school system in terms of pupils' growth per unit-test may require demonstration.

Three major administrative uses of such tests are :

(*i*)   For gradation and placement of pupils

(*ii*) For comparison of performance of group or class with norms of subject and grade

(*iii*) For measuring the efficiency of learning vis-a-vis instructional effectiveness.

9.5.5 Standardised Tests for Decision Making

Standardised tests are likely to be useful in the following cases :

(i)   Since placement decisions call for a broader appraisal in an area, standardised tests can be used to identify entry level of performance on a scale of uniform scores.

(*ii*) Standardised tests help in making normative comparisons when counselling and guidance decisions are to be taken.

(*iii*) Selection decisions tend to imply comparison with others, for which adequate norms are important, and are available in standardised tests.

(*iv*) Standardised tests can play as important role when curriculum decisions between alternative programmes imply broad-based comparison, supplemented by measures developed locally for special objectives.

(*v*)   Since public and policy decisions call for a comprehensive and comparative view of the performance of a school, broad survey and comparison have significant value, as permitted by the standardised test.

9.6 Planning the Testing Programme

Planning of a testing programme includes the following steps :

(a)   Be cognizant of the teaching problem for whose solution the test results are essential.

(b)   Determine the type of test data relevant to the problem.

(c)   Select the best available tests for the purpose.

(d)   Make the needed preparation and administer the test.

(e)   Score the tests quickly, accurately and economically.

(f)   Tabulate, analyse and interpret results.

(g)   Use results and interpretations for improving the conditions related to the problem.

9.6.1 Selection of Tests

(a)   There is need for care in selecting the tests

(b)   Test the rating scales or device that affect the test quality, i.e. assignment of point values to different features of the test (validity, reliability, practicability, comparability and utility).

9.6.2 Administering the Tests

It involves the following steps :

(a)   Preparing for the testing.

(b)   Administration of tests.

(c)   Responsibility of teachers.

9.6.3 Scoring the Test

Its requirements are :

(a)   Hand-scored test-manual.

(b)   Self-scoring tests.

(c)   Machine-scoring devices.

(d)   Use of separate answer sheets.

9.6.4 Analysing and Interpreting Results of Testing

The results of testing are analysed and interpreted through :

(a)   Use of statistical techniques for analysis.

(b)   Use of norms.

(c)   Use of derived scores.

9.7   Limitations of Standardised Tests

There are following limitations of standardised tests.

1.   *As a measure of student achievement*

In this connection the limitations refer to :

   (*i*) lack of complete relevance to instructional objectives of a given course;

   (*ii*) lack of perfect score reliability owing to limited number of items or poor quality of items;

   (*iii*) susceptibility to coaching the students;

   (*iv*) tendency to lose validity with repeated use;

   (*v*) possibility of improper administration and scoring; and

   (*vi*) possibility of inappropriate or unreliable norms.

2.   *As a basis for inferring teacher competence*

It refers to :

   (*i*) sensitivity to conditions other than teacher competence that affect learning;

   (*ii*) insensitivity to teacher's contributions that do not directly foster learning; and

   (*iii*) imperfection as a measure of students' achievement.

Self Assessment

2. *State whether the following statements are 'true' or 'false'.*

   (*i*) Validity of the test depends on the efficiency with which it measures, what it attempts to measure.

   (*ii*) Analysis of desired learning out comes and of the proposed test using logical and psychological methods can not reveal sufficient commonality.

   (*iii*) A test is said to be reliable if it functions Consistently.

9.8 Summary

- A standardised test is expertly and carefully planned, constructed and administered; it provides tables of norms for interpretation of scores. It provides for comprehensive coverage of achievements for comparable scores in different areas. Profiles of test scores indicate pupils' general level of achievement and their specific strengths and weaknesses. Subset scores may be quite unreliable for effective use. Standardised achievement test scores are useful in facilitating inspection and in evaluating its results. Standard test scores provide information that contribute to evaluation of teacher competence, but should not be used as the sole basis for this purpose. In general, locally designed testing programmes are more relevant to local needs, although these are less expertly designed than external testing programme; which in turn are criticised for emphasising only verbal and quantitative skills, penalising creative thinkers. Essential functions served by tests of technical or professional competence would be better served if those who develop such tests were better trained end had paid more attention to the problems of test validity.

- Most of the, standardised tests are broad survey measure of areas of achievement. Such survey tests are connected primarily with selection, placement, guidance, curricular decision and public policy decisions. Since diagnostic testing is a time-consuming enterprise, their development and use is reserved for the students who face serious problems or weaknesses in some specific area of achievement. At secondary level, centrally produced achievement tests   provide evidence of outstanding competence for advanced placement and of minimal competence as a basic outcome of schooling. State-wise programme of testing vacillates between attempt to make objective referenced interpretation of achievements in local schools and global appraisal of effectiveness of different schools and school systems.

9.9 Keywords

Standard   :   The level of quality that is normal or acceptable for a particular person.

Validation   :   To prove that something is true

ambiguity   :   The state of having more than one possible meaning

9.10 Review Questions

1. Write short notes on

    (a) Validity of a test   (b) Objectivity   (c) Content analysis

2. Give the types of norms

3. Explain the reliability Coefficient

4. What the practical uses of Standardised tests?

5. Explain about the planning of testing programme.

Answers : Self Assessment

1. (*i*)   norms   (*ii*) grade norms   (*iii*) Discriminating power

    (*iv*) standardisation

2. (*i*)   T   (*ii*) F   (*iii*) T

9.11 Further Reading

Books

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 10 : Item Analysis : Item Difficulty, Discrimination index, Effectiveness of Distractors

Objectives

The objectives of this unit can be summarized as below :
- To discuss about the meaning of item analysis
- To describe about the need of item analysis
- To explain about the process of item analysis
- To describe about the Difficulty index and Calculation of difficulty value of items
- To explain about correcting difficulty indices for chance errors.
- To describe about the discrimination index and calculating discrimination value by formula
- To explain about the relationship between difficulty index and discrimination Index
- To discuss about the effectiveness of distractors

Introduction

After having administered a test and marked it, most teachers would discuss the answers with students. Discussion would usually focus on the right answers and the common errors made by students. Some teachers may focus on the questions most students performed poorly and the questions students did very well. However, there is much more information available about a test that is often ignored by teachers. This information will only be available if the *Item Analysis*

is done. What is item analysis ? Item analysis is a process which examines the responses to individual test items or question in order to assess the quality of those items and the test as a whole. Item analysis is especially valuable in improving items or questions that will be used again in later tests, but it can also be used to eliminate ambigous or misleading items in a single test administration. Specifically, item analysis is the process of finding out the *difficulty index, discrimination index* and the *effectiveness of distractors.* [*We will discuss each of these components of item analysis in detail later*].

The quality of a test is determined by the quality of each item or question in the test. The teacher who constructs a test can only roughly estimate the quality of a test. This estimate is based on the fact that the teacher has followed all the rules and conditions of test construction. However, it is possible that this estimation may not be accurate and certain important aspects have been ignored. Hence, it is suggested that to obtain a more comprehensive understanding of the test, item analysis should be conducted on the responses of students. Item analysis is conducted to obtain information about individual items or questions in a test and how the test can be improved. It also facilitates the development of an item or question bank which can be used in the construction of a test (Anastasi, 2001).

10.1 Meaning of Item Analysis

The success of a test depends on two factors. First, how successfully a test measures our prescribed objectives, and second, can each item in the test discriminate between bright and dull students or not. If a test is not capable of measuring the given objectives meaningfully and discriminating students, then it cannot be called a successful test.

When a test is constructed, it is begun by noting down the desirable behavioural or personality trait, and the statements or problems relating to them. As a common rule, the number of these statements or test items is far larger than the desirable number in the test. It is done so that the tests found not fit for final inclusion can be deleted. Therefore, an important task of a test-maker is to verify the test items and select only those which are good enough, reliable, valid and which fulfil the objectives. This whole process is called item-analysis.

As far as item-analysis is concerned, **Freeman** has opined that two aspects should be chiefly considered while undertaking item-analysis : first, the difficulty level of items and second, discriminating value of test items. According to **Guilford,** before the final form of the test is constructed, the method of item-analysis must be applied. The different attributes of the whole test are dependent on the items included in a test, such as mean variance, form of score distribution, reliability and validity.

According to **Lamark, "The procedure used to judge the quality of an item is called item-analysis."**

10.2 Need of Item-Analysis

The chief objective of item-analysis of a test is to select those items which are suitable, reliable and valid for that class of students or individuals for whom the test has been constructed. A good test requires good test items. If the test items of a test are objective-oriented, and are also reliable and valid, then the test becomes useful. Selecting such items is the purpose of item-analysis.

In brief, item-analysis is needed for the following purposes :

1.   Rejecting the items which are weak and have double meaning.

2.   Each selected item should represent the whole test.

3.   Determining the number of items to be included in the test.

4.   Expressing the defects in the responses of the students.

According to **Harper** and **Chauhan, " Test analysis is a general term which covers a wide range of statistical and analytical techniques which may be applied for the purpose of improving tests, examinations, quizzes and other mental measurement devices."**

— Harper, Proceedings of the 62nd Indian Science Congress, Part 2nd Presidential Address (1974)

Criteria for Item-analysis : The purpose of item-analysis is to select suitable and good test items from the many items the test-maker has made. For this, each item is evaluated. Besides, it is determined on the basis of a certain criterion which item is good and should be selected, and which item is not good and should not be selected. This decision is taken on two bases :

1. Difficulty index, and

2. Discrimination index or validity index.

If the difficulty index and discrimination index of an item are of the desired level, then that item is selected. There is no definite formula or principle to determine what should be the difficulty and discrimination indices of an item. This depends on the objective of the test and the insight and vision of the test-maker. If a test is survey-type, then the questions with difficulty index between 20-80 and with discrimination index. 3 or more, are considered suitable. If a test has been made with a view of selection, then its difficulty index should be between 10-50 and discrimination index should be about .4 or .5. To calculate difficulty index of a question, it is found how many people in the sample have solved the item correctly, and then this is simplified, and if it is lower, then it is made more difficult, Besides, each item's discrimination index is correlated with the entire test. It is assumed that the whole test measures that trait for which it has been constructed. Therefore, if the correlation of an item is more than the whole test, then it is understood that item measures that trait, that is, the item is valid.

Notes   Item-analysis is also called 'evaluation of the test' or 'try-out stage of test construction'. This is an extremely important step of the test, and no test can be completed without this process.

10.3 Process of Item-Analysis

To undertake item-analysis, at first the test is administered to a sample of those individuals or students for whom it has been constructed. After that, all test papers are marked and the following processes are sequentially followed :

1. Test-papers or answer-sheets are arranged in the ascending order of scores.

2. To prepare item-analysis chart. In this chart, at first, the serial number of items is written in the columns. On the left, the names or serial number of students are written in rows. Thus, the chart is divided into many rows and columns, and when lined, they make several cells.

3. After this, the answer-sheet of each student is taken and it is seen whether he has solved it correctly or not, and a right tick mark is given for the right response and a cross mark is given for the wrong response. Thus, all test items are marked on the chart. At the end of the chart, two rows are made where the difficulty index and discrimination index are written.

4. In each column of the chart from top to bottom, the right tick marks are counted and are changed into percentage in order to write them in the form of difficulty index. Thus, difficulty index is calculated in respect of all test items.

5. To calculate discrimination index, we divide the analysis chat into almost three equal parts from top to bottom.

For example, if one hundred students were administered the test, then the first and third part will have 33 students and the second part will have 34 students. The second part will be covered with a white sheet and stapled. Now, the total of the right responses in the first part and third part will be written in the white sheet. Now, the number of students in the third part who have attempted a particular item correctly will be deducted from the number of students in the first part who have attempted that particular item correctly. The difference thus obtained will be divided by N/3. The coefficient thus obtained will be called its discrimination index. Thus, discrimination index will be calculated for each item, and it will be written in the last row of the chart.

Following is the formula for calculating discrimination index :

$$\text{V. I.} = \frac{T-B}{N/3}$$

Where,   T = Top group

        B = Bottom group

        N = sample size

6.   Now, each item of the test in the chart is studied to see whether its difficulty index is between 20-80 or 10-50 as per the objective of the test and whether its discrimination index is above .3 or not. The items which pass these two criteria are selected. As many items are selected and written separately as are needed for the test. Thus, the final form of the test is readied, which is further verified for its reliability, validity and norms in order to make it standardized.

Some scholars do not take one-third of the samples as above, but they take only top 27% and bottom 27% of the sample. It is tabulated which tells the above requirements easily. According to the opinion of **Kelly,** the formula of top-bottom 27% is considered valid and suitable because on its basis it can be decisively said whether the top group is superior and the bottom group is inferior. If a smaller top-bottom group is taken, then these groups are too small to decide something. Still, there is no definite rule in this regard. The higher is the number of candidates in the top-bottom group, higher will be the discrimination index.

External Criterion for Item-analysis : In the above discussion, the basis for validity of each item was internal, that is, it was accepted as the criterion measure of the whole test, and its correlation was calculated in respect of the whole test. This type of assumption and logic is rational. On this basis, we can say that each item evaluates that trait which is evaluated by the test as a whole, but we cannot say how far it evaluates that trait which it ought to. Generally we take an external criterion in order to ascertain that each item in the test is evaluating the trait for which it has been included in the test, and thus each item is validated on the basis of that external criterion. In the view of Ryens, no specific assistance is offered by the external criterion.

Steps In Item Analysis

Let us take an example of teacher who has administered a 30 item multiple-choice objective test in geography to 45 students in a secondary school classroom.

Step 1 :   Obviously, upon receiving the answer sheet, the first step would to mark each of the answer sheets.

Step 2 :   Arrange the 45 answer sheets from the highest score obtained until the lowest score obtained. The paper with the highest score is on top and the paper with the lowest score is at the bottom.

Step 3 : Multiply 45 (the number of answer sheets) with 0.27 (or 27%) which is 12.15 and rounded up to 12. The use of the value 0.27 or 27% is not inflexible. It is possible to use any percentage between 27% to 35% as the value.

Step 4 : Arrange the pile of 45 answer sheets according to scores obtained (highest score to the lowest score). Take out 12 answer sheets from the top of the pile and 12 answer sheets from the bottom of the pile. Call these two piles as "high marks" students and "low marks" students. Set aside the middle group of papers (21 papers). Although these could be included in the analysis, using only the high and low groups simplifies the procedure.

Step 5 : Refer to Item # 1 or Question # 1,

- Count the number of students from the "high marks" group who selected each of the options (A, B, C and D),

- Count the number of students from the "low marks" group who selected the options A, B, C or D.

Question 1 : Geography Test					
What is the fertile area in the desert in which the water reaches the ground surface called ?					
A. mirage					
B. oasis					
C. water hole					
D. polder					
Item Analysis					
Options	A	B	©	D	Blank
High Marks Group (n = 12)	0	11	1	0	0
Low Marks Group (n = 12)	3	2	3	4	0

Table 10.1 Item Analysis for One Item or Question

From the analysis, 11 students from the "High Marks" group and 2 students from the "Low Marks" group selected 'B' which is the correct answer. This means that 13 out of the 24 students selected the correct answer. Also, note that all the distractors; A, C, & D) were selected by at least one student. However, the information provided in Figure 8.1 is insufficient and further analysis has to be conducted.

10.4 The Difficulty Index

Using the information provided in you can compute the '*Difficulty Index*' which is a quantitative indicator with regards to the difficulty level of an individual item or question. It can be calculated using the following formula :

$$\text{Difficulty Index} = \frac{\text{Number of Students with the Correct Answer (R)}}{\text{Total number of Students who Attempted the Question (T)}}$$

$$= \frac{R}{T} = \frac{13}{24} = 0.54$$

What does a Difficulty Index (p) of 0.54 mean ? The difficulty index is a coefficient that shows the percentage of students who got the correct answer compared to the total number of students in

the two groups who answered incorrectly. In other words, 54% of students selected the correct answer. Although our computation is based on the high and low scoring groups only, it provides a close approximation to the estimate that would be obtained with the total group. Thus, it is proper to say that the index of difficulty for this item is 54% (for this particular group). Note that, since difficulty refers to the percentage getting the item right, the smaller the percentage figure the more difficult the item. Lein (1980) provides these guidelines on the meaning of the difficulty index as follows :

> 0.70	= Easy
0.30 – 0.69	= Moderate
< 0.29	= Difficult

Table. 8.4 Interpretation of the Difficulty Index (*p*)

If a teacher believes that the achievement 0.54 on the item is too low, he or she can change the way he or she teaches to better meet the objective represented by the item. Another interpretation might be that the item was too difficult or confusing or invalid, in which case the teacher can replace or modify the item, perhaps using information from the item's discrimination index or distractor analysis.

Activity
A teacher gave a 20 item science test to a group of 35 students. The correct answer for Question # 25 is 'C' and the results are as follows :

Options	A	B	C	D	Blank
High Marks Group (*n* = 10)	0	2	8	2	0
Low Marks Group (*n* – 10)	2	4	3	2	1

(a) Compute the difficulty index (*p*) for Questions # 25.

(b) Is Question # 25 an easy or difficult question ?

(c) Do you think you need to improve Question # 25 ?

10.5 Calculation of Difficulty Value of Items

The ratio of students' group that solves a question correctly or knows it actually is called its difficulty index. Difficulty index assists to arrange items in sequence, and indicates what item will be placed first and what item will follow it or will come in the end. There is no definite suitable method to ascertain difficulty index, still following are the chief methods used to calculate difficulty index :

1. By right response of the 27% top group and 27% bottom group.

2. 35% Harper's facility index method.

3. By right response of the 50% top group and 50% bottom group.

4. By right response of the 33% top group and 33% bottom group.

5. By formula method.

Generally, 27% top and 27% bottom group method is used, but 50% top and 50% bottom group is taken in case of a smaller group (N < 100) Here we shall discuss only the first four methods only :

10.5.1 50% Top and 50% Bottom Group Right Response Method

For example, a subject teacher wants to prepare a weekly objective test containing 10 items. He has included 16 items in the initial draft and administered it on 10 students. In this case, the difficulty index and discrimination index of this test will be calculated as follows :

Item Analysis Table

Item No.	Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Score
1.	Usha Pande	✓	✓	✓	×	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	×	✓	13*
2.	Mukta Gulati	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	×	✓	14*
3.	Jnana Prakash	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×	×	12*
4.	Imran	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	×	×	✓	13*
5.	Karuna	×	×	×	×	×	×	✓	×	×	×	×	✓	✓	✓	✓	✓	6
6.	Jitendra	✓	✓	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×	4
7.	Sunil Tomar	✓	✓	✓	✓	✓	✓	×	×	×	×	✓	✓	✓	✓	✓	✓	11*
8.	Rajendra	✓	✓	✓	✓	✓	×	×	×	×	×	×	×	×	×	×	×	5
9.	Amandeep	✓	✓	✓	×	×	×	×	×	×	×	×	✓	✓	✓	✓	✓	8
10.	Puja Chandela	×	×	×	×	×	×	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	9

The total number of students is 10. Therefore, the top group and the bottom group will comprise of 50% each, thus containing 5 students each. On arranging scores in descending order :

Scores	Students	
14	Mukta Gulati	
13	Usha Pande	
13	Imran	
12	Jnana Prakash	50% superior group
11	Sunil Tomar	
9	Puja Chandela	
8	Amandeep	
6	Karuna	50% inferior group
5	Rajendra	
4	Jitendra	

After this, right responses in each of superior group and inferior group will be counted for each item. The number of right responses of each item will be divided by the total number of students of each class, and the value obtained thus will be the difficulty index. It is clear from the above table that Usha Pande, Mukta Gulati, Jnana Prakash, Imran and Sunil Tomar have answered almost all questions correctly, so they have been placed in the superior group. On the other hand, the remaining students : Karuna, Jitendra, Rajendra, Amandeep and Puja Chandela have answered almost all questions wrongly, so they have been placed in the inferior group.

Difficulty Level Index

Item No.	Number of right responses	Difficulty index (R_H)		Number of right responses	Difficulty index (R_L)		Joint Difficulty Index ($R_H = R_L$)	
		Higher Group			Inferior Group		Joint Difficulty Index	
1.	5	5/5	1.00	3	3/5	.60	1.60	.80
2.	5	5/5	1.00	3	3/5	.60	1.60	.80
3.	5	5/5	1.00	3	3/5	.60	1.60	.80
4.	3	3/5	.60	2	2/5	.40	1.00	.50
5.	5	5/5	1.00	1	1/5	.20	1.20	.60
6.	5	5/5	1.00	0	0/5	.00	1.00	.50
7.	2	2/5	.40	1	1/5	.20	0.60	.30
8.	3	3/5	.60	1	1/5	.20	0.80	.40
9.	4	4/5	.80	1	1/5	.20	1.00	.50
10.	4	4/5	.80	1	1/5	.20	1.00	.50
11.	4	4/5	.80	1	1/5	.20	1.00	.50
12.	5	5/5	1.00	3	3/5	.60	1.60	.80
13.	5	5/5	1.00	3	3/5	.60	1.60	.80
14.	3	3/5	.60	3	3/5	.60	1.20	.60
15.	1	1/5	.20	3	3/5	.60	0.80	.40
16.	4	4/5	.80	3	3/5	.60	1.40	.70

$$\text{Difficulty Index (D.I.)} = \frac{R_H + R_L}{2}$$

Where, D. I. = Difficulty Index

R_H = Right responses in Higher Group

R_L = Right responses in Lower Group

In the above table, the items with .80 difficulty index is the easiest and with .30 is the most difficult.

In order to calculate discriminating value, the difference between the number of right responses in the two groups is found out. This is called the discrimination index or validity index (V.I.). This is,

V.I. = R_H – R_L (N being the same for both groups)

The following table displays the discriminating value or validity index :

Validity Index Table

Item No.	Number of right responses in superior group (R_H)	Number of right responses in inferior group(R_L)	$(V.I. = R_H - R_L)$	Degree of Value Validity
1.	5	3	+2	Ordinary
2.	5	3	+2	Ordinary
3.	5	3	+2	Ordinary
4.	3	2	+1	Bad
5.	5	1	+4	Very good
6.	5	0	+5	Best
7.	2	1	+1	Bad
8.	3	1	+2	Ordinary
9.	4	1	+3	Good
10.	4	1	+3	Good
11.	4	1	+3	Good
12.	5	3	+2	Ordinary
13.	5	3	+2	Ordinary
14.	3	3	0	Very bad
15.	1	3	– 2	Negative
16.	4	3	+1	Bad

It is clear from the above table that the V.I. of the 15th item is negative. It means that the number of right responses in the inferior group is more than the number of right responses in the superior group. Therefore, this item is not suitable and it should not be included in the final draft of the test. In the same way, the discriminating value of 14th item is zero, which means that this question cannot discriminate between good and bad students, because both types of students can attempt it correctly, and from this viewpoint, this is not suitable for inclusion in the final test. Besides, the discriminating value of 4th, 7th and 16th questions is not satisfactory, so they will have to be excluded from the final draft. The remaining questions can be selected in the final draft with regard to their difficulty index and discriminating value.

Selection of Items for Final Form of the Test

Item No.	Difficulty Index	Discriminating Value	Result
1.	.80	+ 2	Selected
2.	.80	+ 2	Selected
3.	.80	+ 2	Selected
4.	.50	+ 1	Excluded
5.	.60	+ 4	Selected
6.	.50	+ 5	Selected
7.	.30	+ 1	Excluded
8.	.40	+ 2	Selected
9.	.50	+ 3	Selected

10.	.50	+ 3	Selected
11.	.50	+ 3	Selected
12.	.80	+ 2	Selected
13.	.80	+ 2	Selected
14.	.60	0	Excluded
15.	.40	– 2	Excluded
16.	.70	+ 1	Excluded

Thus, items 1, 2, 3, 5, 6, 8, 9, 10, 11, 13 and 13 have been selected for the final form of the test.

10.5.2 Calculating Difficulty Index By Formula

According to the formula method, the following formula is used to calculate difficulty level of items :

$$\text{D.I.} = \frac{N_i}{N_t} \times 100$$

Where, D.I. = Item difficulty Index

N_i = Number of students attempting it correctly

N_t = Total number of students.

For example, in a test, 120 out 150 students have correctly attempted an item, and 30 students have attempted it incorrectly, then the difficulty index of that item will be :

$$\text{D.I.} = \frac{120}{150} \times 100$$

$$= 80\% \text{ or } .80$$

10.6 Correcting Difficulty Indices for Chance Errors

10.6.1 Guilford's Formula

There is every possibility while administering a multiple items that students will guess answers without actually knowing the right responses. In order to curb this tendency, the correction formula is applied, which is are as follows :

$$P_c = \left(R - \frac{W}{K-1} \right) \frac{1}{N}$$

Where, P_c = The percentage of those who know the right responses.

R = The percentage of those who give right response.

W = The percentage of those who give wrong response.

N = The number of candidates in the sample.

K = The number of alternatives.

Example : Supposing out of a total of 300 students, only 240 students have given the right response for an item, and this item had four alternatives. Now, the difficulty level of this item will be :

$$P_c = \left(240 - \frac{60}{4-1}\right)\frac{1}{300} \quad [W = N - R]$$

$$= \left(240 - \frac{60}{3}\right)\frac{1}{300}$$

$$= (240 - 20)\frac{1}{300}$$

$$= \frac{220}{300} = .733$$

$$= .733 \times 100 = 73.3\% \qquad \text{(Difficulty Index in Percent)}$$

Sometimes it is seen in multiple choice items that a student is not able to attempt some items of the test. It could be due to lack of time or the items being very difficult, or some other cause. Such items are called unattempted or unreached items. For explanation, supposing in a 50-item test, a candidate has attempted all questions from 1 to 25, he has not attempted items 26 and 27, and then he has solved all items from 28 to 40, and has not solved any item after that. In such a situation, items 41 to 50 will be called unattempted or unreached items, but not items 26 and 27. besides, there will be three groups of candidates : who have given right responses, who have given wrong responses and who have not responded.

In solving this type of questions, the formula is amended to the following form :

$$P_c = \left(R - \frac{W}{K-1}\right)\frac{1}{N - NR}$$

Here, NR is the number of candidates who have not been able to solve an item or have not attempted it. The remaining all symbols have been used as in the previous formula.

Example : Supposing, a 100-item test was administered to 300 candidates. Each item contained 4 alternatives. This is also assumed that 150 were right responses, 120 were wrong responses and 30 were unattempted. In such a situation, the difficulty index of that particular item will be as follows:

$$P_c = \left(150 - \frac{120}{4-1}\right)\frac{1}{300 - 30}$$

$$= \left(150 - \frac{120}{3}\right)\frac{1}{270}$$

$$= (150 - 40)\frac{1}{270}$$

$$= 110 \times \frac{1}{270}$$

$$= \frac{11}{27}$$

$$= 0.4 \times 100$$

$$= 40\%$$

10.6.2 Kelley's Formula

When the number of students is very large, in such a situation, Kelley's top 27% and bottom 27% formula proves convenient, which is as follows :

$$P_c = \frac{1}{2}\left\{\left(R_H - \frac{W_H}{K-1}\right)\frac{1}{N_H - NR_H} + \left(R_L - \frac{W_L}{K-1}\right)\frac{1}{N_L - NR_L}\right\}$$

Where, R_H = Number of candidates in top group giving right response

W_H = Number of candidates in top group giving wrong response.

N_H = Number of total candidates in top group.

NR_H = Number of candidates in top group who have not attempted the item.

R_L = Number of candidates in bottom group giving right response.

W_L = Number of candidates in bottom group giving wrong response.

N_L = Number of total candidates in bottom group.

NR_L = Number of candidates in bottom group who have not attempted the item.

Example : An achievement test was administered upon a group of 380 students approximately. The Test-administrator is interested in calculating Difficulty-Index of Item No. 75 by Kelleys, T-B 27% method for which the different values are as under :

$$N_H = 100, R_H = 70, W_H = 20 \text{ and } NR_H = 10$$

$$N_L = 100, R_L = 20, W_L = 60 \text{ and } NR_L = 20$$

Solution :

$$\therefore \qquad P_c = \frac{1}{2}\left\{\frac{70 - \dfrac{20}{5-1}}{100-10} + \frac{20 - \dfrac{60}{5-1}}{100-20}\right\}$$

$$= \frac{1}{2}\left\{\frac{70 - \dfrac{20}{4}}{90} + \frac{20 - \dfrac{60}{4}}{80}\right\}$$

$$= \frac{1}{2}\left\{\frac{70-5}{90} + \frac{20-15}{80}\right\}$$

$$= \frac{1}{2}\left\{\frac{65}{90} + \frac{5}{80}\right\}$$

$$= \frac{1}{2}\left\{\frac{5200+450}{7200}\right\}$$

$$= \frac{1}{2} \times \frac{5650}{7200}$$

$$= \frac{1}{2} \times \frac{113}{114}$$

$$= \frac{113}{228}$$

$$= .39$$

$$= .39 \times 100$$

$$= .39\% \quad \text{(Difficulty-Index)}$$

10.6.3 Calculation of D.P. Using Chi-square Test Formula

When the values calculated by the above formulae are different, it presents difficulty in their comparative study; therefore, another formula is applied which is called Chi-square test formula. this formula has two forms. The first form is applied in the situation when the value of R_H is more than R_L, and the second form is applied when the value of R_L is more than R_H.

$$\text{D.P.} = \cfrac{R_H - R_L - 1}{\sqrt{R_T 1 - \cfrac{R_L}{N_T - NR_L}}} \qquad \text{(First form, when } R_H > R_L\text{)};$$

and

$$\text{D.P.} = \cfrac{R_H - R_L + 1}{\sqrt{R_T\left(1 - \cfrac{R_T}{N_T - NR_T}\right)}} \qquad \text{(Second form when } R_L > R_H\text{)}$$

Example : 340 class-10 students were administered a test. All related values of the test are given below. Find out discriminating value using any square formula.

$$N_H = 100, R_H = 65, \text{ and } NR_H = 15$$

$$N_L = 100, R_L = 25 \text{ and } NR_L = 35$$

Solution : ∴ Discriminating Value :

$$= \cfrac{R_H - R_L - 1}{\sqrt{R_T\left(1 - \cfrac{R_T}{N_T - NR_T}\right)}}$$

$$= \cfrac{65 - 26}{\sqrt{90 - \left(1 - \cfrac{90}{200 - 50}\right)}} \qquad \begin{cases} N_T = N_H + N_L \\ R_T = R_H + R_L \\ NR_T = NR_H + NR_L \end{cases}$$

$$= \cfrac{65 - 26}{\sqrt{90 - \left(1 - \cfrac{90}{200 - 50}\right)}}$$

$$= \cfrac{39}{\sqrt{90\left(1 - \cfrac{3}{5}\right)}}$$

$$= \cfrac{39}{\sqrt{90\left(\cfrac{2}{5}\right)}}$$

$$= \cfrac{39}{\sqrt{36}}$$

$$= \cfrac{39}{6}$$

$$= 6.5$$

Therefore, D.P. = 6.5

10.6.4 Horst's Formula

Supposing the students' responses in a multiple choice item has been noted down as below :

Item Analysis :

(a)	Determining the number of actual items	12
(b)	Excluding weak and ambiguous items	8
(c)	Ensuring that selected items represent the entire test	20
(d)	All of the above	10

Here, the alternative (c) is the greatest obstacle because it has attracted most of the students. This is stronger than even alternative (d). So, Horst has determined some basic assumptions for his formula $S = D_p$, which are as follows :

(a) All obstacles do not attract all students equally, though there is an arrangement in their attraction.

(b) Here importance is given to the most potent obstacle, which is very important from the viewpoint of multiple choice items.

(c) The most potent obstacle should not be more attractive than the right alternative. If it is so, then another obstacle should replace it.

In the above formula, placing values in the formula :

$$S = R - D_p$$
$$= 10 - 20$$
$$= -10$$

Where, S = The number of students knowing the right response in fact.

R = The number of students giving right response to the item.

D_p = The most potent obstacle.

The negative (minus) mark indicates that nobody known the right alternative. Therefore, the alternatives should be reformed, which can be as follows :

Self Assessment

1. Multiple Choice Questions :

Choose the correct option :

(i) According to _____ "The procedure used to judge the quality of an item is called item analysis.

(a) Lamarck (b) Freeman (iii) Guilford (iv) Ryen

(ii) If a test is survey type, then questions with difficulty index between _____ and with discrimination index .3 or more are considered suitable.

(a) 40–80, .5 (b) 30–80, .2 (c) 20–80, .3 (d) 10–80, .1

(iii) The $P_C = \left(R - \dfrac{W}{K-1}\right)\dfrac{1}{N}$ formula is known as _____.

(a) Kelly's formula (b) Guilford's formula

(c) Chi-square formula (d) Host's formula

(iv)   In $P_C = \dfrac{1}{2}\left\{\left(R_H - \dfrac{W_H}{K-1}\right)\dfrac{1}{N_H - NR_H} + \right.$

$\left.\left(R_L - \dfrac{W_L}{K-1}\right)\dfrac{1}{N_L - NR_L}\right\}$ the R_H denotes _____

(a)   Number of candidates in top group giving right response

(b)   Number of candidates in top group giving wrong response

(c)   Number of total candidates in top group

(d)   Number of total candidates in bottom group

(v)   Difficulty percentage = $\dfrac{R_H - R_L - 1}{\sqrt{R_T 1 - \dfrac{R_L}{N_T - NR_T}}}$ is called

(a)   Guilford formula

(b)   Chi-square test formula

(c)   Kelly's formula

(d)   27% Top and 27% bottom right response method

10.7 The Discrimination Index

The 'Discrimination Index' is a basic measure which shows the extent to which a question discriminates or differentiates between students in the 'High Marks' group and 'Low Marks' group. This index can be interpreted as an indication of the extent to which overall knowledge of the content area or mastery of the skills is related to the response on an item. Most crucial for a test item is that whether a student got a question correct or not is due to their level of knowledge or ability and not due to something else such as chance or test bias.

Note in our example earlier, 11 students in the high group and 2 students in the low group selected the correct answer. This indicates positive discrimination, since he item differentiates between students in the same way that the total test score does. That is, students with high scores on the test (high group) got the item right more frequently than students with low scores on the test (low group). Although analysis by inspection may be all that is necessary for most purposes, an index of discrimination can be easily computed using the following formula :

Discrimination Index =

Number of Students in 'High Marks' group (R_H) with the Correct Answer *u*) Minus ($-$)

Number of Students in 'Low Marks' group (R_L) with the Correct Answer divided by ($\div$)

Half the Total Number of Students who Attempted the Question (½T)

Example : A test was given to a group of 43 students and 10 out of the 13 'High Marks' group got the correct answer compared to 5 out of 13 'Low Marks' group who got the correct answer. The discrimination index is computer as follows :

$$= \frac{R_H - R_L}{\frac{1}{2}T} = \frac{10-5}{\frac{1}{2}(26)} = \frac{10-5}{13} = 0.38$$

What does a Discrimination Index of 0.38 mean ? The discrimination index is a coefficient that shows the extent to which the question discriminates or differentiates between 'High Marks"

students and 'Low Marks' students. Blood and Budd (1972) provides the following guidelines on the meaning of the discrimination index as follows :

Value	Description	Suggestion
> 0.40	high discrimination	Question is retained
0.20 – 0.39	moderate discrimination	Question is checked & revised
0.10 – 0.19	low discrimination	Question is removed or rewritten
0.00 & negative	no discrimination	Question is removed

Table 10.2 Interpretation of the Discrimination Index

A question that has a high discrimination index is able to differentiate between students who know and do not know the answer. When we say that a question has a low discrimination index, it is not able to differentiate between students who know and students who do not know. A low discrimination index means that more "Low Marks' students got the correct answer because the question was too simple. It could also mean that students from both the 'High Marks' group and 'Low Marks' group got the answer wrong because the question was too difficult.

The formula for the discrimination index is such that if more students in the 'High Marks' group chose the correct answer than did students in the low scoring group, the number will be positive. At a minimum, then, one would hope for a positive value, as that would indicate that it is knowledge of the question that resulted in the correct answer.

- The greater the *positive* value (the closer it is to 1.0), the stronger the relationship is between overall test performance and performance on that item.

- It the discrimination index is *negative,* that means that for some reason students who scored low on the test were more likely to get the answer correct. This is a strange situation which suggests poor validity for an item.

Task What is Horst's formula?

10.8 Calculating Discriminating Value by Formula

To calculate discriminating value of an item by the formula method, we arrange answer books in a definite descending order after marking them. In this, the answer books with higher scores will be at the top and those with lower scores will be at the bottom. Now, the top 1/3 of the answer books are separated into one group and the bottom 1/3 of the answer books are segregated into another group. The remaining answer books will not be needed. It is clear that the top 1/3 answer books belong to good students and the bottom 1/3 answer books to weak students. Now, we have to calculate discrimination between these two groups. At first, it is found out what percent of students in the top group have solved an item correctly, and how many have not. These will be called P_1, Q_1 respectively. Thus, P_1 and Q_1 are calculated for each item in the top group. In the same way, the percentage of students in the bottom group solving each item correctly is found out and how many of them have not solved them correctly. These two will be called P_1 and Q_2 respectively. Here too, P_2 and Q_2 will be calculated for each group. Supposing, the number of students in the top group is N_1 and the number in the bottom group is N_1. Though N_1 and P_2 are equal, because we take 1/3 of the total number of students.

Supposing, we know P_1, Q_1, P_2, Q_2, N_1 and N_2 for each item. Now, these values are applied in the following formula in order to know their discriminating value for each item :

$$D = \frac{P_1 - P_2}{\sqrt{\dfrac{P_1Q_1}{N_1} + \dfrac{P_2Q_2}{N_2}}}$$

In this formula, D means discriminating value. If this value is more than 1.96, then we can say that our items are discriminating, that is, they are good enough to discriminate between good and weak students. Thus, we calculate the discriminating value for each item. The items having D higher than 1.96 or 2.00 will be considered as good.

The following formula is used to estimate the discriminating value of the test :

D.P. = PQ

Where, D.P. = Discriminating value

P = Number of students solving items correctly (in percent)

Q = The number of students solving items incorrectly (in percent)

Example : In a test, 50 students have solved an item correctly and the remaining incorrectly. Estimate discriminating value of the item when the test was administered on 80 students.

Formula, D.P. = PQ

Here, $P = \dfrac{R}{N} \times 100$

$= \dfrac{50}{80} \times 100$

= 62.5%

And, Q = 100 – P [or $Q = \dfrac{W}{N} \times 100$]

= 100 – 62.5

= 37.5 percent

Thus, D.P. = 37.5 × 62.5

= 2343.5 (about)

[Note : An item with 2500 discriminating value is considered suitable.]

4.27% Top and 27% Bottom Right Response Method : This is a very popular method and it was mentioned by Garrett in this way.

Under this method, the answer books are arranged in the descending order, that is, the answer book obtaining the highest marks is placed at the top and the one obtaining the lowest marks is placed at the bottom. Then the top 27% and bottom 27% are allowed to be the while the remaining answer books are taken out. Supposing, we had a total of 300 answer books, then the top 81 and bottom 81 answer books are taken while other 138 answer books are not considered. Now the top and bottom groups are analyzed for the right responses for each item and their percentage is found out. Then the difficulty index and bi-serial correlation, that is, discriminating value is calculated using the table given on the previous page.

The use of table can be explained by the following example :

Example : In an examination, the number of candidates in the top 27% and bottom 27% solving the first three of the eighty questions is as follows :

Item No.	27% Top Right Responses	27% Bottom Right Responses
1.	74	50
2.	62	38
3.	42	10

Now, validity index and difficulty index of each item has to be calculated.

We shall use the table for validity index. For item 1, when we move our finger along 27% column below 74 and 27% row along 50, we come to the point where both these cross each other. The number written there is 26, that is the validity index of the item. Therefore, the reliability of this item is .26. All numbers is the intervening numbers in the table have point (fraction). In the same way, the validity index for item 2 and 3 is .25 and .41.

In order to calculate difficulty index, these two percentages are added and average is calculated.

Therefore, the difficulty index of items 1 = $\dfrac{74+50}{2}$ = $\dfrac{124}{2}$ = .62

In the same way, the difficulty index for items 2 and 3 is .50 and .26.

Thus, we make a table, like the one given below, for all 300 items of the test :

Item No.	27% Top Right Responses	27% Bottom Right Responses	Validity Index.	Difficulty Index
1.	74	50	.26	.62
2.	62	38	.25	.50
3.	42	10	.41	.16
.	.	.	.	.
.	.	.	.	.
300	.	.	.	.

The items with .20 or more validity index are considered satisfactory. The items with validity index less than this are deleted. If a very good test is needed, then only those items are selected which have a validity index over .50.

10.9 Relationship Between Difficulty Index and Discrimination Index

Theoretically, the more difficult a question (or item) or easier the question (or item), the lower will be the discrimination index. Stanley and Hopkins (1972) provide a theoretical model to explain the relationship between the difficulty index and discrimination index of a particular question or item. According to the model, a difficulty index of 0.2 can result in a discrimination index of about 0.3 for a particular item (which may be described as an item of 'moderate discrimination'). Note that, as the difficulty index increase from 0.1 to 0.5, the discrimination index increases even more. When the difficulty index reaches 0.5 (described as an item of 'moderate difficulty'), the discrimination index is positive 1.00 (very high discrimination). Interestingly, a difficulty index of > 0.5 leads to the discrimination index decreasing.
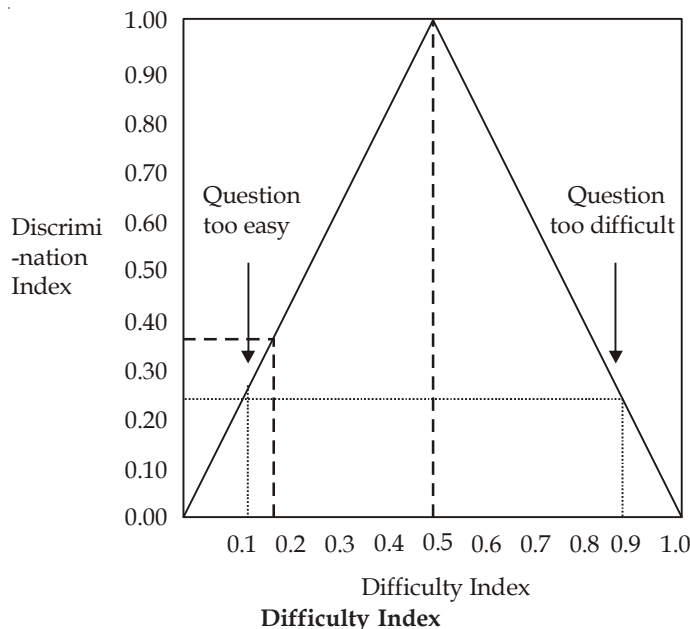
Theoretical Relationship between Index of Difficulty and Discrimination Index

[*source* : G. Stanley and D. Hopkins, 1972. Introduction to Educational Measurement and Testing. Boston : Macmillan.p.270]

- For example, a difficulty index of 0.9 results in a discrimination index of about 0.2 which is described as an item of low to moderate discrimination. What does this mean ? The more difficult a question, the harder is for that question or item to discriminate between those students who know and do not know the answer to the question.

- Similarly, when the difficulty index is about 0.1, the discrimination index drops to about 0.2. What does this mean ? The easier a question, the harder is for that question or item to discriminate between those students who know and do not know the answer to the question.

10.10 Effectiveness of Distractors

In addition to examining the performance of an entire test item, teachers are also interested in examining the performance of individual distractors (incorrect answer options) on multiple-choice items. By calculating the proportion of students who chose each answer option, teachers can identify which distractors are "working" and appear attractive to students who do not know the correct answer, and which distractors are simply taking up space and not being chosen by many students. To eliminate blind guessing which results in a correct answer purely by chance (which hurts the validity of a test item), teachers want as many plausible distractors as is feasible. Analyses of response options allow teachers to fine tune and improve items they may wish to use again with future classes. Let us examine performance on an item or question.

Which European power invaded Melaka in 1511 ?

	Portuguese	Dutch	Spanish	English	
Options	A	B	C	D	No Response
High Marks Group (*n* = 12)	5	5	0	2	0
Low Marks Group (*n* = 12)	3	4	0	5	0

Table : Effectiveness of Distractors

Generally, a good distractor is able to attract more 'Low Marks' students to select that particular response or distract 'Low Marks' students towards selecting that particular response. What determines the effectiveness of distractors ? See Figure which shows how 24 students selected the options A, B, C and D for a particular question. Option B is a less effective distractor because many 'High Marks' students ($n = 5$) selected option B. Option D is relatively a good distractor because 2 students from the 'High Marks' group and 5 students from the 'Low Marks' group selected this option. The analysis of response options shows that those who missed the item were about equally likely to choose answer B and answer D. No students chose answer C. Answer C does not act as a distractor. Students are not choosing between four answer options on this item, they are really choosing between only three options, as they are not even considering answer C. This makes guessing correctly more likely, which hurts the validity of the item. The discrimination index can be improved by modifying and improving options B and C.

Activity					
Which British resident was killed by Maharajalela in Pasir Salak ?					
H. Low Birch Brooke Gurney					
Options	A	B	C	D	No Response
High Marks ($n = 15$)	4	7	0	4	0
Low Marks ($n = 15$)	6	3	2	4	0
The answer is B.					
Analyse the effectiveness of the distractors.					

10.11 A Practical Approach in Item Analysis

Some teachers may find the techniques discussed earlier as time consuming which cannot be denied (especially when you have a test consisting of 40 items). However, there is a more practical approach which may take less time. Imagine that you have administered a 40 item test to a class of 30 students. Surely, it will take a lot of time to analyse the effectiveness of each item which may discourage teachers from analysing each item in a test. Diederich (1971) proposed a method of item analysis which can be conducted by the teacher and the students in his or her class. The following are the steps :

Step 1 : Arrange the 30 answer sheets from the highest score obtained until the lowest score obtained.

Step 2 : Select the answer sheet that obtained a middle score. Group all answer sheets above this score as 'High Marks' (mark a 'H' on these answer sheets). Group all answer sheets below this score as 'Low Marks' group (mark a 'L' on these answer sheets).

Step 3 : Divide the class into two groups (High and Low) and distribute the 'High' answer sheets to the High groups and the Low answer sheet to the Low group. Assign one student in each group to be the counter.

Step 4 : The teacher than asks the class.

Teacher : "The answer for Question # 1 is 'C' and those who got it correct, raise your hand'.

Counter from 'H' group : "Fourteen for group H"

Counter from 'L' group : "Eight from group L"

Step 5 : The teacher records the responses on the whiteboard as follows :

	High	Low	Total of Correct Answers
Question # 1	14	8	22
Question # 2	12	6	18
Question # 3	16	7	23
Question # n	n	n	n

Step 6 : Compute the *Difficulty Index* for Question # 1 as follows :

$$\text{Difficulty Index} = \frac{R_H + R_L}{30} = \frac{14 + 8}{30} = 0.73$$

Step 7 : Compute the *Discrimination Index* for Question # 1 as follows :

$$\text{Discrimination Index} = \frac{R_H - R_L}{\frac{1}{2}30} = \frac{14 - 8}{15} = \frac{6}{15} = 0.40$$

Note that earlier, we took 27% of answer sheets in the 'High Marks' group and 27% of answer sheets in the 'Low Marks' group from the total answer sheets. However, in this approach we divided the total answer sheets into two groups. There is no middle group. The important thing is to use a large enough fraction of the group to provide useful information. Selecting the top and bottom 27 per cent of the group is recommended for more refined analysis. This method may be less accurate but it is a 'quick and dirty' method.

Self Assessment

2. Fill in the blanks :

(*i*) _____ is a basic measure which shows the extent to which a question discriminates between students in the High marks groups and Low marks group.

(*ii*) For calculating discriminating value of an item by the formula method, the answer books are arranged in a definite _____ order.

(*iii*) 27% Top and 27% bottom right response method was mentioned by _____.

(*iv*) _____ provide a theoretical model to explain the relationship between difficulty index and discrimination index.

10.12 Summary

- According to Lamark, "The procedure used to judge the quality of an item is called item-analysis."

- The chief objective of item-analysis of a test is to select those items which are suitable, reliable and valid for that class of students or individuals for whom the test has been constructed. A good test requires.

- In brief, item-analysis is needed for the following purposes :

  (*i*) Rejecting the items which are weak and have double meaning.

  (*ii*) Each selected item should represent the whole test.

  (*iii*) Determining the number of items to be included in the test.

- Criteria for Item-analysis : The purpose of item-analysis is to select suitable and good test items from the many items the test-maker has made.

- This decision is taken on two bases :

  (*i*) Difficulty index, and

  (*ii*) Discrimination index or validity index.

- Process of Item-analysis

- To undertake item-analysis, at first the test is administered to a sample of those individuals or students for whom it has been constructed. After that, all test papers are marked and the following processes are sequentially followed :

  (*i*) Test-papers or answer-sheets are arranged in the ascending order of scores.

  (*ii*) To prepare item-analysis chart. In this chart, at first, the serial number of items is written in the columns.

  (*iii*) After this, the answer-sheet of each student is taken and it is seen whether he has solved it correctly or not, and a right tick mark is given for the right response and a cross mark is given for the wrong response.

  (*iv*) In each column of the chart from top to bottom, the right tick marks are counted and are changed into percentage in order to write them in the form of difficulty index.

  (*v*) To calculate discrimination index, we divide the analysis chart into almost three equal parts from top to bottom.

  (*vi*) Now, each item of the test in the chart is studied to see whether its difficulty index is between 20-80 or 10-50 as per the objective of the test and whether its discrimination index is above .3 or not.

- Using the information provided in you can compute the '*Difficulty Index*' which is a quantitative indicator with regards to the difficulty level of an individual item or question. It can be calculated using the following formula :

- Difficulty Index

$$= \frac{\text{Number of Students with the Correct Answer (R)}}{\text{Total number of Students who Attempted the Question (T)}}$$

$$= \frac{R}{T} = \frac{13}{24} = 0.54$$

- The ratio of students' group that solves a question correctly or knows it actually is called its difficulty index. Difficulty index assists to arrange items in sequence, and indicates what item will be placed first and what item will follow it or will come in the end.

- 50% Top and 50% Bottom Group Right Response Method

- For example, a subject teacher wants to prepare a weekly objective test containing 10 items. He has Calculating Difficulty Index By Formula

- According to the formula method, the following formula is used to calculate difficulty level of items :

- The 'Discrimination Index' is a basic measure which shows the extent to which a question discriminates or differentiates between students in the 'High Marks' group and 'Low Marks' group. This index can be interpreted as an indication of the extent to which overall knowledge of the content area or mastery of the skills is related to the response on an item.

- Correcting Difficulty Indices for Chance Errors

- Guilford's Formula : There is every possibility while administering a multiple items that students will guess answers without actually knowing the right responses. In order to curb this tendency, the correction formula is applied, which is are as follows :

  - When the number of students is very large, in such a situation, Kelley's top 27% and bottom 27% formula proves convenient, which is as follows :

  - $P_c = \dfrac{1}{2}\left\{\left(R_H - \dfrac{W_H}{K-1}\right)\dfrac{1}{N_H - NR_H} + \left(R_L - \dfrac{W_L}{K-1}\right)\dfrac{1}{N_L - NR_L}\right\}$

- Calculation of D.P. Using Chi-square Test Formula : When the values calculated by the above formulae are different, it presents difficulty in their comparative study; therefore, another formula is applied which is called Chi-square test formula. this formula has two forms. The first form is applied in the situation when the value of R_H is more than R_L, and the second form is applied when the value of R_L is more than R_H.

- Supposing the students' responses in a multiple choice item has been noted down as below:

- Item Analysis :

  (a) Determining the number of actual items   12

  (b) Excluding weak and ambiguous items   8

  (c) Ensuring that selected items represent the entire test   20

  (d) All of the above   10

- Here, the alternative (c) is the greatest obstacle because it has attracted most of the students. This is stronger than even alternative (d). So, Horst has determined some basic assumptions for his formula $S = D_p$, which are as follows :

  (a) All obstacles do not attract all students equally, though there is an arrangement in their attraction.

  (b) Here importance is given to the most potent obstacle, which is very important from the viewpoint of multiple choice items.

  (c) The most potent obstacle should not be more attractive than the right alternative. If it is so, then another obstacle should replace it.

- Calculating Discriminating Value of Validity Index by Formula

- To calculate discriminating value of an item by the formula method, we arrange answer books in a definite descending order after marking them. In this, the answer books with higher scores will be at the top and those with lower scores will be at the bottom.

- 4.27% Top and 27% Bottom Right Response Method : This is a very popular method and it was mentioned by Garrett in this way.

- Under this method, the answer books are arranged in the descending order, that is, the answer book obtaining the highest marks is placed at the top and the one obtaining the lowest marks is placed at the bottom. Then the top 27% and bottom 27% are allowed to be the while the remaining answer books are taken out.

- The more difficult a question (or item) or easier the question (or item), the lower will be the discrimination index. Stanley and Hopkins (1972) provide a theoretical model to explain the relationship between the difficulty index and discrimination index of a particular question or item. According to the model, a difficulty index of 0.2 can result in a discrimination index of about 0.3 for a particular item (which may be described as an item of 'moderate discrimination').

• In addition to examining the performance of an entire test item, teachers are also interested in examining the performance of individual distractors (incorrect answer options) on multiple-choice items. By calculating the proportion of students who chose each answer option, teachers can identify which distractors are "working" and appear attractive to students who do not know the correct answer, and which distractors are simply taking up space and not being chosen by many students.

10.13 Keywords

Item   :   A single article or object

analysis   :   A detailed study or examination of something in order to understand more about it.

discrimination   :   The practice of treating somebody or a particular group in society less fairly particular group in society less fairly than others.

10.14 Review Questions

1. What is item analysis ? Why is the need of item analysis ?

2. Explain the stepwise process of item analysis

3. How can of Calculate the difficulty value of items ? Explain different methods with examples

4. Explain the Kelly's formula.

5. Describe the relationship between Difficulty index and discrimination index

6. What is effectiveness of distractors ?

Answer : Self Assessment

1. (*i*)   a   (*ii*)   c   (*iii*) b   (*iv*) a   (*v*) b

2. (*i*)   Discrimination index   (*ii*)   decending

    (*iii*) Garret   (*iv*) Stanley and Hopkins

10.15 Further Readings

Books

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 11 : Development of Norms of a Test

Objectives

The objectives of this unit can be summarised as below :

- To discuss about the meaning of norms
- To describe the need for establishing norms
- To explain the types of norms
- To discuss about the characteristics of norms

Introduction

Test norms consist of data that make it possible to determine the relative standing of an individual who has taken a test. By itself, a subject's raw score (e.g., the number of answers that agree with the scoring key) has little meaning. Almost always, a test score must be interpreted as indicating the subject's position relative to others in some group. Norms provide a basis for comparing the individual with a group. We will discuss about norms of a test and its various aspects.

11.1 Meaning of Norms

In a test, norm is that score which has been obtained by a group. In other words, "By norms we mean specimen of work which represent the commonest type of work for the whole group in question. In the field of research, when different tests are administered and scoring is given. It is a difficult task to deduce some inferences on the basis of these scores, unless we have a basis for them. On the basis of norm, we can compare two candidates in the test, and can find out the situation or place of a candidate in the group. We apply norm in order to eradicate interpretive errors. Norms have been defined as the standard, performance of a group of pupils in a test. It is essential to keep in mind that there is difference between norm and standard. Norms indicate the actual achievement of students at standardized level, while standard indicate the desired level of performance. Norms are such standards with which we can compare any scores for analysis, or deduce our conclusions from them. In other words, norms are the average scores of actual measurement in a test administered on a specific group of students. Norms are averages or values determined by actual measurement of a group of persons who are the representatives of

specific population. While standard possesses a desirable objective, which can be less or more than the obtained norm.

To prepare norms, we administer a test on a large population in order to change the scores into percentiles or standard scores. Norms are used as the central tendency of scores of a certain group.

Some common definitions of norm are as follows :

1. Norms are quantitative criterion for different levels.

2. Norms are averages determined under prevailing conditions.

3. Psychological test norms represent the test standardization.

11.2 Need for Establishing Norms

Norms have to be established for the following causes :

1. To formulate raw scores such that they can be added with other scores.

2. To make raw scores capable of being compared.

3. To make raw scores more meaningful.

The group of individuals on whom these norms are prepared, are called reference group, that is, norms indicate standard performance of reference group. Generally, norms help us to answer the following queries :

1. The obtained scores of an individual are whether less or more than the scores he has scored in another test measuring the same trait. On the basis of this type of comparison, an individual's strength or weakness can be found out.

2. How can future success be predicted for an individual ?

3. What is the place of an individual in a test as compared to other individuals in the test.

4. How can the individual's success be studied conveniently and successfully.

11.3 Types of Norms

Norms are of two types depending on the group of students or individuals for which they have been established. Chiefly, there are four types of norms :

1. Age norms.

2. Grade norms.

    In view of the scale or unit of expression, norms can be of two types :

3. Percentile norms.

4. Standard score norms.

l. **Age Norms :** This type of norms were used in Binet's intelligence tests. The basic assumption in this type of norms is that variable should increase with the age. The variable must increase with the age.

    By age norms is meant the average scores of students of different ages. Age norms find out only those variables which increase with age, such as height, weight, intelligence, reading ability, vocabulary, mental age etc. Supposing, we have to find out height age norm. for this, a test is administered on a group of students of different age groups, and the scores are written down on the test. The average of scores obtained by students of different age groups is called the norm for that age group. For example, if we want to know the height norm for students of 12-year age group. For this, we shall choose some students of 12 years of age

who should represent the entire population. Then their height will be measured and average will be calculated. This average will be called the height age norm for all 12-year students. Now we can measure a student's height and analyze it in accordance with this norm. If his height is more than this norm, then he will be called a taller than average boy. For example, if the average height of 15-year old age group is 135 *cm*, and Nishant's height is 145 *cm*, then we can say that Nishant is taller than his average age group, or that he is taller as relative to his age. On the other hand, if the average height of 16-year boys is 145 *cm*, then we can say that his height is equal to 16-year age group. In the field of intelligence measurement, the concept of mental age is a form of age norms. If a student is able to solve all questions meant for an 18-year age group, then he will be said to have 18-year mental age, even if his physical age is merely 15 years. In all intelligence tests, norms are presented in the form of mental age and norms in achievement tests are presented in the form of educational age.

Age norms can be easily established. This is the reason that they are widely used in educational field.

"Age or grade norms locate the pupil in terms of age or grade groups, but not necessarily with pupils of his own age and grade."

– D. Baron and H.W. Bernard

Limitations

(a) Chief limitation of age norms is that the development and growth of each child is not uniform in all age groups. A child grows three times more in 5-7 age group than 11.14 age group. Consequently, the difference in norms is not uniform for different age groups.

(b) These norms cannot be used in personality tests, interest tests, aptitude tests, attitude tests etc.

(c) The unit of age is unsuitable to express the level of ability during adolescence and youth.

(d) These norms should be used only when all variables increase along with age; such as weight, intelligence, educational achievement etc.

(e) The selection of representative sample is a difficult task.

(f) Some traits do not show mental development with age, such as vocabulary increases with age, but maze tracing is obstructed after adolescence. Therefore, this type of traits cannot be shown by age norms.

(g) These norms are appropriate for the students up to the age group of 20 years.

2.   **Grade Norms :** Grade norms are similar to age norms, with the only difference that while age norms are related with age, grade norms are related with class. They are also called class norms. By grade norms in a test is meant the average scores of students of different classes. This is administered on a classified students in the school. Like age norms, the other variable should be such which increases with the age.

To ascertain grade norms, a test is administered on students of different classes. The students selected from a specific class represent the entire population of that class. Then the scores of students for each class are found out. The average scores for each class are called grade norms for that class. Supposing, we want to establish grade norms in science. Representative groups are taken from different classes, as VII, VIII, IX and X, and their average scores are found out. Now, the students whose present achievement has to be measured is administered the test, and his scores are explained on the basis of grade norms. If a VII grade student is able to achieve the average score of IX grade, then he will be considered a strong student. On the contrary, if a IX grade student attains the average score meant for a VII grade, he will be called a weak student. Grade norms are mostly established for achievement tests. Thus, grade norms are related with the performance of average students of all classes.

Grade norms are very important for teachers. They can analyze the performance and ability of students on their basis, and can ascertain the position of a certain child in the class.

Notes Standard score norms are also expressed by other norms in addition to Sigma score norms, such as Z score, t-score, Sten-score, Stanine-score etc.

Limitations

(a) The variance in grade norms is not very explicit.

(b) The rate of educational achievement, intelligence development and other variables is not uniform as per the grade.

(c) Grade norms, like age norms, are not uniform.

(d) These norms can be used only in formal educational institutions.

(e) If students of the same age group have to compared, then grade norms do not assist us. For example, we can compare 7-year-old Punit with other boys aged 8, 9 or 10 years, but not with many other as old as Punit. In such a situation, students are compared with one another on the basis of percentile norms.

3. **Percentile Norms :** By percentile norms in a test is meant the different percentiles obtained by a large group of students. In other words, percentile norms are those scores, the number of students obtaining scores below than that is equal to the percentage of such students. For example, 75th percentile norm tells that 75% students have scored below this score, and only 25% students have obtained scores above it. In calculating percentile norm, a candidate is compared with the group of which he is a member. By percentile scores is meant the grade of a candidate in percentiles. Supposing 100 individuals are taking part in a race. One of them runs the fastest and stands first. He is better than 99 individuals, so his percentile value is 99. The individual standing second in the race is better than 98 individuals, so his percentile position is 98th. The distance between the first and second individuals does not influence their percentile positions. The individual running last is followed by no other individual, so his percentile position will be 0. In the same way, under educational situations, when several students of the same or different schools are studied, it is quite convenient and useful to transform their sequences into percentile ranks. In ordinary words, percentile is the point on the scale below which a fixed percentage of the distribution falls.

In order to know percentile value, a test is administered on a large group, and different percentile values are calculated on the basis of scores obtained by students. These percentile values are percentile norms. Because, it is possible to use them on all individuals of the common group under all circumstances, so it can be said about them that percentile norms provide a basis for interpreting the score of an individual in terms of his standing in some particular group.

In calculating percentile and percentile rank, suppose a test was administered on a large group, and M = 100, S.D. = 1, then different percentiles will be as follows :

On the above basis, if a candidate obtains 100 marks in the test, then his percentile rank will be 50 and his performance level will be considered as average.

Merits

(a) They can be analyzed easily.

(b) It is not necessary to administer the test on a sample representative group, as is done in other tests. Therefore, no hypothesis has to be formulated for these norms. So, these are used widely.

(c) These norms are useful in all types of circumstances, such as educational, industrial, military fields etc.

(d) Percentile norms can be easily developed.

(e) They can be used to meaningfully express the scores with different units and numerical standards.

(f) Generally only percentile norms are ascertained for personality tests, IQ tests, attitude tests, aptitude tests etc.

Limitations

(a) It is not possible to carry out statistical analysis of these norms.

(b) The percentile scores of different tests cannot be compared unless the groups on which they were administered are not comparable; for example, if in a personality test, percentile norms have been developed for adolescent girls taken from a large group, then the scores of all adolescent girls can be compared with these.

(c) In normal situations, percentile norms tell the relative position of each individual, but it does not make out the difference in scores between two individuals.

(d) Percentile norms are often confused with percent scores.

(e) The relative position of an individual is ascertained on the basis of these norms. It is not possible to analyze actual ability or capability of an individual objectively.

(f) The units of percentile scores are not uniform. If the details of actual scores are almost common, then there is much difference in changing proximate scores into percentile values, while there is not much difference in changing scores at extreme ends.

4. **Standard Score Norms :** The greatest shortcoming of percentile norms is that the units of scores is not equal in this, that is, the two consecutive percentiles are not equally or uniformly distanced. For example, the difference between 30th and 40th percentiles is not equal to the difference between 60th and 70th percentiles. Due to this shortcoming, these norms cannot be used to compare the differences among different candidates. Therefore, test-makers look for such units which are meaningful throughout the entire expanse. From this standpoint, the standard score norms are widely used. These norms are also called Z score norms.

By standard score norms is meant to change the raw scores of candidates into standard scores. This type of norms are found out with the help of standard deviation (S.D. or σ). This standard deviation is a measurement of the expanse of scores of a group. Standard norms are based on normative group. These norms analyze the achievement of an individual on the basis of his scores in the context of the particular group. Because these express uniform units, so they are different from percentile norms. Their basic unit is the standard deviation of the reference group, and the scores of an individual are expressed below median of the group and above standard deviation unit.

Basically, all these are the same thing, and are the modified forms of the same scale.

The mean and standard deviation of Z-score are 0 and 1 respectively. The mean and standard deviation of *t*-score are 50 and 10 respectively. The mean and standard deviation of Sten-score are 5.5 and 4 respectively, and those of Stanine-score are 5 and 2 respectively. T-score was first used by **McCall** and **Sten-score** was first used by **R.V. Cattell.** If in a test, the mean of a large group is M and standard deviation is σ, then the scores can be calculated by the following scores :

$$Z - \text{Score} = X - M/\sigma$$

$$t - \text{Score} = 50 + (X - M/\sigma) \times 10$$

$$\text{Sten} - \text{Score} = 5.5 = (X - M/\sigma) \times 2$$

$$\text{Stanine} - \text{Score} = 5 = (X - M/\sigma) \times 2$$

T-scores are also like *t*-scores, with the only difference that in T-scores, the raw scores are normalized. The basis of all standard scores is normalized distribution curve. If distribution is not equal, then equivalent standard scores will not be correct too. Therefore, the distributor is made equal, and then raw scores are changed into standard scores. All other characteristics of T-scores are similar to *t*-scores. By Stanines is meant 'standard nine', that is, the scale which has nine units. Stanine 9 is the largest unit, and Stanine 1 is the smallest, and at the middle is Stanine 5.

To calculate standard norm, the test is administered on a large group and standard deviation and mean of scores are found out. Now the score to be analyzed is changed into some standard score (Z or *t* etc.), and the thus obtained standard score is analyzed. The value of standard score tells how less or more a student is from the mean. A test-maker makes a table for changing raw scores into standard scores, with the help of which the raw scores of the test can be changed into standard scores. For example, Usha Pande has scored 60 marks in a mathematics achievement test, and on that test, the value of Z-score is 2 for 60 score; it would mean that Usha Pande obtains +2 more marks than an average girl in mathematics achievement. 2Z = 0Z + 2Z). In the same way, if the value of *t*-score of Nishi's English scores is 75, it would mean that she obtains 25 *t*-score than average girls. (75*t* = 50 *t* + 25 *t* or 75 = 50 + 25). In the same way, Sten and Stanine scores can also be analyzed.

Generally, standard scores and norms appear to be similar, but this is only an illusion. In fact, there is basic difference between the two. According to Ross and Stanley, the term 'standard' has objective inherent in it, that is, it tells about our goal. While 'norm' indicates present achievement. We want to proceed towards the standard which remains stable, while the degree of stability is deficient in norms.

Did u know? Z-score is generally plus/minus 3 and *t*-score between 20-80, while Sten-score and Stanine-score are between 0-9. Relative Z, t, Sten and Stanine value to a score tell that how more or less that score is from the mean (0, 50, 5.5 or 5 respectively.)

11.4 Characteristics of Norms

It is essential for norms of a test to have the following essential qualities :

1. Novelty.

2. Representation.

3. Meaningfulness.

4. Comparability.

1. **Novelty :** By being novel or up-to-date is meant that the norms should not be outdated, that is, norms should not be constructed on the basis of test scores which were administered a long way back because time interval can effect change in students' abilities. For this reason, if an intelligence test constructed in 1990 will not be suitable to analyze scores of students on a test administered in 2004, as it would not be proper. Therefore, the norms should be changed from time to time so that they remain novel and up-to-date.

2. **Representation :** By representation of norms is meant that norms should be developed from the scores obtained from the representative group, whose scores have to be analyzed. Therefore, if some skill of class 9 students has to be analyzed, then the norms too should be developed on the basis of scores obtained from class 9 students. Besides, these students should be equal to other students in other traits. Test norms should be constructed on the basis of scores obtained from a large group. A small group cannot represent the whole

population adequately, due to which a norm developed on a small group can give incorrect interpretation.

3. **Meaningfulness :** By meaningfulness is meant the type of norms. The evolved norms should be dependent on the test objectives and measurable traits. Where traits increase with an increase in age, it would be proper to develop age norms or grade norms. However, if personality has to be measured, then percentile norms or standard score norms should be applied. In the same way, if the aim of a test is to ascertain the desirability of a student's physical or educational achievement, then it would be more adequate to use age or grade norms, but if the aim of a test is to find out the position of the student in a large group, then the percentile and standard norms can be used.

4. **Comparability :** Comparability is an important characteristic of norms. Test norms should be mutually comparable, only then these norms can be used to compare different students. Besides, norms should be sufficiently described, that is, the different reference points should be clearly explained so that the students' ability can be clearly explained in words.

In the context of above characteristics of norms, the following precautions should be exercised while determining norms :

(*i*) Norms should be carefully calculated. For this, scores of a student in different tests should be referred, and not his scores in only one test.

(*ii*) The sample used for ascertaining norms should always represent the entire population.

Task What are Z-score norms ?

Self Assessment

1. Multiple choice questions :

(*i*) Norms indicate the actual achievemnt of _____ at standardized level.

(a) Students   (b) Teachers

(c) Principals   (d) Supervisors

(*ii*) The group of individuals on whom, norms are prepared are called _____.

(a) Age group   (b) Reference group

(c) Percentage group   (d) grade group.

(*iii*) Grade norms are also called as _____.

(a) age norms   (b) class norms

(c) Percentile norms   (d) score norms

(*iv*) _____ are those scores, the number of students obtaining scores below than that is equal to the percentage of such students.

(a) Percentile norms   (b) age norms

(c) grade norms   (d) standard norms

(*v*) _____ are based on normative group

(a) age norm   (b) grade norm

(c) Standard norm   (d) Percentile norms

(*vi*) The ____ is a measurement of the expanse of scores of a group.

(a) Mean   (b) standard deviation

(c) Median   (d) mode

(*vii*) T-score was first used by _____.

        (a)   Mc Call   (b)   R.V. Cattell

        (c)   Kelly   (d)   Lamark

11.5 Summary

- In a test, norm is that score which has been obtained by a group. In other words, "By norms we mean specimen of work which represent the commonest type of work for the whole group in question. In the field of research, when different tests are administered and scoring is given.

- Norms have been defined as the standard, performance of a group of pupils in a test. It is essential to keep in mind that there is difference between norm and standard. Norms indicate the actual achievement of students at standardized level, while standard indicate the desired level of performance. Norms are such standards with which we can compare any scores for analysis.

- Norms have to be established for the following causes :

  (*i*) To formulate raw scores such that they can be added with other scores.

  (*ii*) To make raw scores capable of being compared.

  (*iii*) To make raw scores more meaningful.

- Norms are of two types depending on the group of students or individuals for which they have been established. Chiefly, there are four types of norms :

  (*i*) Age norms.

  (*ii*) Grade norms.

  (*iii*) Percentile norms.

- Standard score norms.

- Age Norms : This type of norms were used in Binet's intelligence tests. The basic assumption in this type of norms is that variable should increase with the age. The variable must increase with the age.

- Grade Norms : Grade norms are similar to age norms, with the only difference that while age norms are related with age, grade norms are related with class. They are also called class norms. By grade norms in a test is meant the average scores of students of different classes.

- Percentile Norms : By percentile norms in a test is meant the different percentiles obtained by a large group of students. In other words, percentile norms are those scores, the number of students obtaining scores below than that is equal to the percentage of such students.

- Standard Score Norms : The greatest shortcoming of percentile norms is that the units of scores is not equal in this, that is, the two consecutive percentiles are not equally or uniformly distanced. These norms are also called Z score norms.

- By standard score norms is meant to change the raw scores of candidates into standard scores. This type of norms are found out with the help of standard deviation (S.D. or σ). This standard deviation is a measurement of the expanse of scores of a group. Standard norms are based on normative group.

- Characteristics of Norms

- It is essential for norms of a test to have the following essential qualities :

  (*i*) Novelty.

(*ii*) Representation.

(*iii*) Meaningfulness.

(*iv*) Comparability.

- Novelty : By being novel or up-to-date is meant that the norms should not be outdated, that is, norms should not be constructed on the basis of test scores which were administered a long way back because time interval can effect change in students' abilities.

- Representation : By representation of norms is meant that norms should be developed from the scores obtained from the representative group, whose scores have to be analyzed.

- Meaningfulness : By meaningfulness is meant the type of norms. The evolved norms should be dependent on the test objectives and measurable traits.

- Comparability : Comparability is an important characteristic of norms. Test norms should be mutually comparable, only then these norms can be used to compare different students.

11.6 Keywords

Norms   :   A situation or a pattern of behaviour that is usual or expected.

Score   :   The number of points somebody gets for correct aswers in a test.

Novelty   :   The quality of being new different and interesting.

11.7 Review Questions

1.   What is the meaning of norms ?

2.   Explain the need for norms

3.   Describe the different types of norms.

4.   Explain the characteristics of norms.

Answers : Self Assessment

1.   (*i*)   a   (*ii*)   b   (*iii*)   b   (*iv*)   a
     (*v*)   c   (*vi*)   b   (*vii*)   a

11.8 Further Readings

Books

1.   Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2.   Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3.   Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4.   Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 12 : Conversion of Raw Scores into Standard Scores, T-scores, C-scores, Z-scores, Stanine Scores, Percentiles

CONTENTS

Objectives

Introduction

12.1 Meaning of Raw Scores

12.2 Standard Scores

12.3 Z-Scores

12.4 T-Scores (Transformed Scores)

12.5 C-Scores

12.6 Stanine Scores

12.7 Percentile Scores

12.8 Summary

12.9 Keywords

12.10 Review Questions

12.11 Further Readings

Objectives

The objectives of this unit can be summarized as below :

- To explain the Meaning of Raw Score
- To discuss about Standard Score
- To describe about Z-Score
- To explain about T-Score
- To discuss about C-Score
- To describe about Stanine Score
- To discuss about Percentile Score

Introduction

By score is meant quantitative measure of a quality of an object or person. We know that the physical qualities of a person can be measured in standard units; for example, the weight of a person in *kg*, and his height in *cm*; so the scores obtained in these units can be easily compared. For example, if the weight of Ram is 60 *kg* and that of Shyam is 30 *kg* then it can be said that the weight of Ram is double than that of Shyam. On the contrary.

In absence of a standard unit of these measures, they cannot be analysed and compared. For example, if Ram gets 60 marks in a mathematics test and Shyain 30; then it cannot be said that Ram's ability in mathematics is double than that of Shyam. It can be more or less than double. It

is evident that the scores obtained from educational measurements cannot be analysed on their own basis. So, in statistics, these raw scores are converted into several types of new scores and by these new scores, the raw scores are analysed. These converted scores are called derived scores in statistics.

Standard scores indicate where your score lies in comparison to a norm group. For example, if the average or mean score for the norm group is 25, then your own score can be compared to this to see if you are above or below this average.

12.1 Meaning of Raw Scores

Raw score is an original datum that has not been transformed. This may include, for example, the original result obtained by a student on a test (i.e., the number of correctly answered items) as opposed to that score after transformation to a standard score or percentile rank or the like.

Often the conversion must be made to a standard score before the data can be used. For example, an open ended survey question will yield raw data that cannot be used for statistical purposes as it is; however a multiple choice question will yield raw data that is either easy to convert to a standard score, or even can be used as it is.

12.2 Standard Scores

Standard scores are those scores which have a definite reference point or norm, and this norm is the Mean (M) of the scale. Besides, it has a definite standard deviation (σ). Several standard scores can be obtained using mean (M) and standard deviation (σ). The chief among them are Z-scores, T-scores, C-scores and Stanine scores.

Suppose that you have just completed your midterm for this class and you were told that you had a score of 55. How would you feel ? The first question you might ask is how many points were possible ? If you were told that 85 points were possible you might not feel too well since you correctly answered about 64% of the questions.

Now let's assume that the test was very difficult and the instructor didn't expect anyone to have mastered all of the content on it. What other information might you wish to know ? Perhaps you would like to know the average (mean) grade on the test. If you were to discover that it was 50, you might feel better about your performance because you were above average.

You might be interested to know how the scores were spread above and below the average (mean). In particular, you are probably wondering how far above the mean you were compared to others in the class. Were most of the grades close to the mean, or were students' grades far above or below the mean. One way to measure the dispersion or spread of scores is with the range (subtract the low score from the high score). Suppose the range were 30 points with the high being 75 and the low being 45. You might not feel to well about your grade even though it was above the average. The problem with the range is that one extreme score can influence it very much. In this case, maybe only one person earned 75 and the next high was 56 with everyone else falling between 46 and 56, a range of 10.

Instead of using the range, we use the standard deviation when we talk about the spread of scores. In the midterm example, suppose you were told that 68% of the people who took the test has a score from 48 to 52. In other words, 68% of the people fell 2 points above or 2 points below the mean. In that case, we would say that the test scores had a standard deviation of 2. Assuming that the scores fell into a normal distribution, we would also know that 95% of the students would have scores within two standard deviations above or below the mean. In our case that would 4 points above (54) or 4 points below (46) the mean (50). You would feel rather good about your score of 55. Knowing the mean and standard deviation makes it possible to interpret raw scores and compare different individuals' performances with each other or an individual's performance on one test with his or her performance on another test.

> **?**
>
> *Did u know?* A common way to make comparisons is to calculate z-scores. A z-score tells how many standard deviations someone is above or below the mean. A z-score of -1.4 indicates that someone is 1.4 standard deviations below the mean. Someone who is in that position would have done as well or better than 8% of the students who took the test.

12.3 Z-Scores

Meaning of Z-Score

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

Calculation of Z-Scores

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

Z-score, $\qquad Z = \dfrac{X - M}{\sigma}$

In which, $\qquad$ Z = Z-score

$\qquad$ X = Raw score

$\qquad$ M = Mean

$\qquad \sigma$ = Standard deviation

Example 2

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

Calculation

Z scores of Student A, $\qquad Z = \dfrac{X - M}{\sigma}$

$$= \dfrac{90 - 65}{10}$$

$$= \dfrac{25}{10}$$

$$= 2.5$$

Z scores of Student B, $\qquad Z = \dfrac{X - M}{\sigma}$

$$= \dfrac{35 - 65}{10}$$

$$= \dfrac{-25}{10}$$

$$= -3$$

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

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

Example 3

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

Calculation

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

$$= 65 + (2.5 \times 10)$$

$$= 65 + 25$$

$$= 90$$

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

$$= 65 - 30$$

$$= 35$$

> *Notes* There are no standard units to measure non-physical qualities of a person such as intelligence, aptitude, personality, and ability, etc. The measures of these are unit less, for example, obtained score of a student in an examination as 40 and his IQ as 110.

Utility of Z-Scores

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

Table—1

Meaning of Z-scores

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

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

Example 4

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

Calculation

Z-score of mathematics marks of Student A, (Z) = $\dfrac{X - M}{\sigma}$

$$= \dfrac{80 - 60}{8}$$

$$= \dfrac{20}{8}$$

$$= 2.5$$

Z-score of language marks of Student A, (Z) $= \dfrac{X - M}{\sigma}$

$$= \dfrac{70 - 40}{10}$$

$$= \dfrac{30}{10}$$

$$= 3$$

Analysis

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

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

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

12.4 T-Scores (Transformed Scores)

Meaning of T-scores

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

Calculation of T-Scores

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

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

In which, $\qquad\qquad$ T $=$ T-score

X = Raw score

M = Mean

σ = Standard deviation

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

T-score, T = 50 + 10Z

Example 5

Look at example 4. The score in mathematics = 80 and Z-score = 2.5; and the score in language = 70 and Z-score = 3. Convert these scores into T-scores.

Calculation

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

= 50 + 10 x 2.5

= 50 + 25

= 75

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

= $50 + 10 \times 3$

= 50 + 30

= 80

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

Utility of T-Scores

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

Table—2

Meaning of Various T-Scores

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

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

12.5 C-Scores

Meaning of C-Scores

C-scores are those converted scores of raw scores of which the mean (M) is 5 and standard deviation (σ) is 2. For it, the range of raw scores is distributed into 11 equal unit-groups which are displayed on a 11-point scale from 0 to 10. These groups are sequential, that is, the scores of the second group are more than the first one, and the scores of third are more than that of the second; in the same way, the scores of eleventh group are more than that of the tenth.

Calculation of C-Scores

There are two methods of converting a raw score into C-score— by Z-score and by percentile rank. The following formula is used to convert a raw scores into C-score by Z-score :

C-score, C = 5 + 2Z

To convert a raw score into C-score by percentile rank (PR), the raw score is first changed into percentile rank and then C-score is found out using the following Table—3.

Table—3

11-Point C-Score Scale

C-scores	Z-scores limits	Range of PRs	% of Raw scores	% in whole Number
10	More than + 2.25	More than 98.78	1.22	1
9	More than + 1.75 to +2.25	95.99 to 98.78	2.79	3
8	More than + 1.25 to + 1.75	89.44 to 95.99	6.55	7
7	More than + 0.75 to + 1.75	77.34 to 89.44	12.10	12
6	More than + 0.25 to 0.75	59.87 to 77.34	17.47	17
5	Less than - 0.25 to + 0.25	40.13 to 59.87	19.74	20
4	Less than - 0.75 to - 0.25	22.66 to 40.13	17.47	17
3	Less than - 1.25 to - 0.75	10.56 to 22.66	12.10	12
2	Less than - 1.75 to - 1.25	4.01 to 10.56	6.55	7
1	Less than - 2.25 to - 1.75	1.22 to 4.01	2.79	3
0	Less than - 2.25	Less than -2.22		

Utility of C-Scores

In some situations in the field of education, the use of C-scores is more useful than the use of Z-scores and T-scores, however such situations are very rare.

The first demerit of its use is awarding zero score. The second demerit is that there comes much difference in percentage (%) if C-score goes up or down by even 1 point.

12.6 Stanine Scores

Meaning of Stanine Scores

Stanine scores are those converted scores of raw scores of which the mean (M) is 5 and standard deviation (σ) it 1.96. For it, the raw scores are divided into 9 unit-groups and they are displayed on a 9-point scale from 1 to 9. The groups of a Stanine scale are not equal, because this scale is

made by joining the two classes on either side of C-scale. So, the range of score of points 1 and 9 is 1.002, and for remaining 7 classes (2 to 8) is 0.052. These groups are sequential, that is, the second group is better than the first one, and the third is better than the second, and in the same way, the ninth group is better than the eighth.

Calculation of Stanine Scores

Stanine scores can be obtained using any of C-scores, percentile rank (PR) and T-scores. We have already discussed the methods of calculating C-scores, percentile ranks (PR) and T-scores. Having known the varue of anyone of them, the table 4 is used to find out the value of Stanine score.

Table—4

Stanine Score, C-Score, Percentile Rank, T-Score Table

Stanine Scores	C-Scores	Percentile Rank	T-Score
9	9 -10	96 - 100	67.5 +
8	8	89 - 96	62.5 - 67.5
7	7	77 - 89	57.5 - 62.5
6	6	60 - 77	52.5 - 57.5
5	5	40 - 60	47.5 - 52.5
4	4	23 - 40	42.5 - 47.5
3	3	11 - 23	37.5 - 42.5
2	2	4 -11	32.5 - 37.5
1	0 - 1	1 - 4	less than 32.5

Utility of Stanine Scores

Only 9 numbers are used in Stanine, while 11 in C-score. So the analysis of these scores is easier than those of C-scores. The meaning of Stanine scores is generally taken as follows :

Table—5

Meaning of Stanine Scores

Stanine Score	Interpretation
9	very good
7 – 8	good
4 – 6	average
2 – 3	low
1	very low

In the classes on edges (classes 1 and 9) of Stanine scores, the score fall 4 – 4% which are not according to rules. There are strong possibilities of committing error on classifying on its basis. Another shortcoming in it is that the best 1% students are placed in the lower 3% students, so these cannot be used for giving guidance.

12.7 Percentile Scores

A percentile score is another type of converted score. Your raw score is converted to a number indicating the percentage of the norm group who scored below you. For example, a score at the

60th percentile" means that the individual's score is the same as or higher than the scores of 60% of those who took the test. The 50th percentile is known as the median and represents the middle score of the distribution.

> **Notes** Percentiles have the disadvantage that they are not equal units of measurement. For instance, a difference of 5 percentile points between two individual's scores will have a different meaning depending on its position on the percentile scale, as the scale tends to exaggerate differences near e mean and collapse differences at the extremes.

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

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

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

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

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

> **Task** What are transformed scores ?

Self Assessment

(*i*)   A raw score is on original datum that has not been transformed.

(*ii*)   The formula for Converting raw scores into Z-scores is $Z = \dfrac{X - M}{\sigma}$

(*iii*)   The formula for Converting raw scores into T-scores is $T = 5 + 10Z$

(*iv*)   The formula for Converting raw scores into C-scores is $C = 5 + 2Z$

(*v*)   The Stanine scores are better them C-scores from analysis viewpoint .

(*vi*)   The meaning of more than $+ 2Z$ score is Good.

(*vii*)   A percentile score is not the other type of Converted score.

12.8 Summary

- Raw score is an original datum that has not been transformed.

- Standard scores are those scores which have a definite reference point or norm, and this norm is the Mean (M) of the scale.

- Z-Scores

  - Z-scores are those converted scores of raw scores of which the mean (M) is zero (0) and standard deviation (σ) is one (1). These scores are obtained by linear transformation of raw scores, so they fall in the category of linear standard scores. The unit of Z-scores is similar to the standard deviation (σ).

- Calculation of Z-Scores

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

    Z-score, $\quad Z = \dfrac{X - M}{\sigma}$

    In which, Z = Z-score

    $\qquad$ X = Raw score

    $\qquad$ M = Mean

    $\qquad \sigma$ = Standard deviation

- In the field of education, the Z-scores are used in meaningful analysis of scores of a student, comparison of scores of two students and for comparison of marks of a student in-two subjects.

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

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

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

  In which, T = T-score

  $\qquad$ X = Raw score

  $\qquad$ M = Mean

  $\qquad \sigma$ = Standard deviation

- The mean (M) of T-scores is 50 and standard deviation (σ) is 10. so, if the range of general distribution is taken to be 100, the range of T-scale becomes from 0 to 100.

- C-scores are those converted scores of raw scores of which the mean (M) is 5 and standard deviation (σ) is 2. For it, the range of raw scores is distributed into 11 equal unit-groups which are displayed on a 11-point scale from 0 to 10.

- There are two methods of converting a raw score into C-score— by Z-score and by percentile rank. The following formula is used to convert a raw scores into C-score by Z-score :

  - C-score, C = 5 + 2Z

- In some situations in the field of education, the use of C-scores is more useful than the use of Z-scores and T-scores, however such situations are very rare.

- Stanine scores are those converted scores of raw scores of which the mean (M) is 5 and standard deviation (σ) it 1.96. For it, the raw scores are divided into 9 unit-groups and they are displayed on a 9-point scale from 1 to 9.

- Stanine scores can be obtained using any of C-scores, percentile rank (PR) and T-scores. We have already discussed the methods of calculating C-scores, percentile ranks (PR) and T-scores.

- Only 9 numbers are used in Stanine, while 11 in C-score. So the analysis of these scores is easier than those of C-scores. The meaning of Stanine scores is generally taken as follows.

- A percentile score is another type of converted score. Your raw score is converted to a number indicating the percentage of the norm group who scored below you.

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

12.9 Keywords

Score   :   The number of points, goals etc.

Standardized   :   To make objects or activities of the same type have the same features or qualitites.

Calculation   :   The act or process of using numbers to find out an amount

12.10 Review Questions

1. What do you understand by Z-scores ? How are raw scores Converted into Z-scores. Discuss their utility and short comings in the field of education.

2. What do you understand by raw scores ?

3. What do you understand by standard scores ?

4. How are raw scores Converted into T score and C-Score ?

5. Of Z, T, C and stanine score, which of then is most used in the field of Education.

6. What is percentile score ?

Answers : Self Assessment

1. (*i*)   T   (*ii*)   T   (*iii*)   F   (*iv*)   T
    (*v*)   T   (*vi*)   F   (*vii*)   T

12.11 Further Readings

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Books

Unit 13 : Interpretation of Test Scores : Qualitative and Quantitative

---

CONTENTS

Objectives

The objectives of this unit can be summarised as below :

- To explain about relative rank
- To discuss about Qualitative Interpretation of test scores
- To discuss about Quantitative Interpretation of test scores

Introduction

Scores, whether raw or converted, do not suffice for the complete interpretation of an individual's performances on psychological tests. The several, aspects of test standardization thus far presented are concerned with the performance of *groups* of persons and with *average relationships* revealed by statistical treatment of results. It happens, however that although certain types of test items meet some or most of the statistical requirements of validity, they are unsatisfactory as indicators of intelligence when used for clinical purposes. For example, on the Stanford-Binet scale, the percentage of adults able to repeat eight digits forward (*digit-span* test) is approximately the same as the percentage who can solve one of the more difficult reasoning problems. Yet, in clinical examinations, psychologists find some adult mental defectives who can pass the former test, although a mental defective can never succeed with the latter. What this means is that statistical validation of a test item is not always sufficient; it must be supplemented by the pragmatic criterion of use with a wide variety of individuals in a variety of situations in order to show whether or not it has discriminative value among individuals at the several levels of ability.

13.1 Relative Rank

The raw score (that is, the actual number of units or points) obtained by an individual on a test does not in itself have much, if any, significance. One test may yield a maximum score of 150, another 200, and a third 300. Obviously, then, any point score on one of these tests is not directly comparable with the same number of points on either of the others; a score of 43 on one test cannot be directly compared with a score of 43 on another. Furthermore, the average scores of each of these will in all probability be different, as will the degree of variation of scores (called the *deviation*) both above and below the average.

13.1.1 Percentile Rank

An individual's percentile rank on a test designates the percentage of cases or scores lying below it. Thus a person having a percentile rank of 20 (P_{20}) is situated above twenty percent of the group of which he is member; or, otherwise stated, twenty percent of the group fall below this person's rank. A percentile rank of 70 (P_{70}) means that seventy percent fall below—and so on, for any percentile rank in the scale. In effect, this statistical device makes it possible to determine at which one-hundredth part of the distribution of scores or cases any particular individual is located. By this means a person's relative status, or position in the group, can be established with respect to the traits or functions being tested. And, as will be seen, psychological measurement, unlike physical measurement, derives its significance principally from relative ranks ascribed to individuals rather than from quantitative units of measurement.

A table of norms and frequency distribution often provides percentile ranks. Or, if the percentile ranks themselves are not given in a table, it is possible to calculate them easily from the frequency distribution.

The percentile method is a technique whereby scores on two or more tests, given in units that are different, may be transformed into uniform and comparable values. This method has the advantage of not depending upon any assumptions regarding the characteristics of the distribution with which it is used. The distribution might be normal, skewed, or rectangular. When a percentile rank is given for a particular individual, it refers to his rank in the specified group of scores from which it has been derived. On a test of reading comprehension at the fourth-grade level, for example, a percentile rank of 60 for a particular pupil is relevant to the group of pupils for whom the distribution of scores was found. Whether or not this same pupil would be rated at percentile 60 as a member of another fourth-grade population will depend on the comparability of the two groups. His rating might be the same, or higher, or lower. (In this connection, consult, giving norms for college freshmen.)
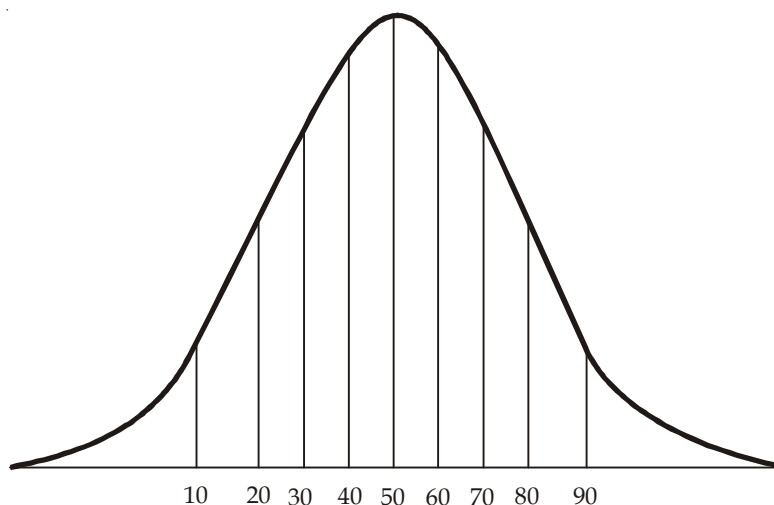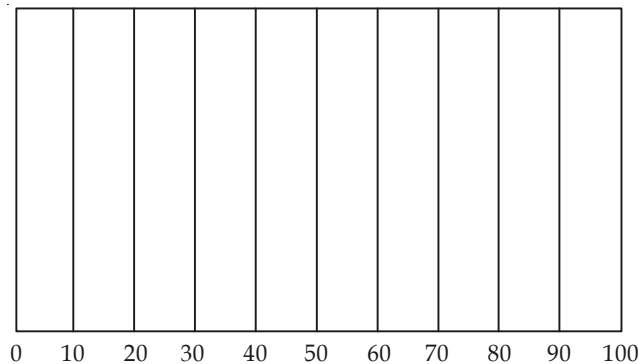

Figure 1 : Unequal distances between points on the base line of a normal curve by successive 10-percent divisions (deciles) of its area

Percentile points are based upon the number of scores (cases) falling within a certain range; hence the distance between any two percentiles represents a certain area under the curve; that is, a certain number of cases (N/100). Reference to Figure 1 shows that if percentages of the total area (total number of cases) are equal, the distances on the base line (range of scores) must be unequal, unless the distribution is rectangular (Figure 2). It is obvious from Figure 1 that differences in scores between any two percentile points become greater as we move from the median (P_{50})

toward the extremes. Inspection of the curve shows, for instance, that the distance on the base line (representing scores) between percentiles 50 and 6o and the distance between 50 and 40 (these being at the center and equal) are smaller than that between any other intervals of ten percentile points. What this means in the practical interpretation of test results is that at, and close to, the median, differences in scores between percentile ratings are smaller in the measured characteristic than they are between the same percentile differences elsewhere on the curve. See, for example, the spread of the base line between 50-60, and that between 80-90, or 90-100. Yet each of these represents ten percentile points.

0	10	20	30	40	50	60	70	80	90	100

Figure 2 : Equal distances between points on the base line of a rectangle by successive 10-percent divisions (deciles) of its area

The percentile technique has the advantage of being easily calculated, easily understood, and of making no assumptions with regard to the characteristics of the tolal distribution. It answers the question : "Where does an individual's score rank him in his group ?" Or : "Where does an individual's score rank him in another group whose members have taken the same test ?"

13.1.2 Decile Rank

The decile rank is the same in principle as the percentile; but instead of designating the one-hundredth part of a distribution, it designates the one-tenth part of the group (N/10) in which any tested person is placed by his score. The term "decile" is used to mean a dividing point. "Decile rank" signifies a range of scores between two dividing points. Thus a testee who has a decile rank of 10 (D_{10}) is located in the highest 10 percent of the group; one whose decile rank is 9 (D_9) is in the second highest 10 percent; one whose decile rank is 1 (D_1), is in the lowest 10 percent of the group.

When the number of scores in a distribution is small, percentiles are not used, because there is little or no significance in making fine distinctions in rank. The decile-ranking method may be used instead.

13.1.3 Mental Age

This concept was introduced by Alfred Binet in 1908 in conjunction with the first revision of his scale. In this scale and in its later revision, items are grouped according to age levels. For example, selected items, passed by a specified percentage of five-year-old children in the standardization sample are placed at the five-year level; items passed by a specified percentage of six-year-old children are placed at the six-year level.

To determine mental age, in the 1908 scale, Binet adopted the following rule : the child was credited with the mental age of the highest year level in which he passed all test items, or all but one. To this basic level an additional year in mental age was added for every five items he passed in higher levels; but no fractional years were added for fewer than five items passed. The defect of this method was recognized, so that in his 1911 scale, Binet modified his procedure in order to permit the addition of a fraction of a year for items passed. In the Stanford-Binet revisions, the

testee is credited with all items up through the age level at which he passes all. This is called the *basal year*. He is also credited with all items passed above the basal year. The sum of his basal plus the other credits, in terms of months, is his mental age.

Mental age norms are also used with scales that are not arranged according to age levels. These are *point scales* that yield a score usually based on the number of items correctly answered. By means of a table of norms provided for the particular test being used, it is possible to assign an individual an age rating. Thus, on a point scale, an individual who, regardless of chronological age (CA), earns a score equal to the norm of the ten-year-old population sample, will have a mental age (MA) of ten, as determined by that test.

In determining mental age, whether by using an age scale or a point scale, an individual's performance on a standardized series of test items is being compared with the performance of the average group of a representative sample at successive age levels. Hence, we define mental age as the level of development in mental ability expressed as equivalent to the chronological age at which the average group of individuals reach that level. For example, a child having an MA of eight, has reached the level of the average group of eight-year-olds in the standardization group.

At this point, our concern is only to define and clarify the mental age concept. There are several important psychological and measurement problems connected with this concept that are explained at several appropriate points in later sections.

Notes In giving the test, the examiner continues upward in the scale until an age level is reached at which the individual fails all items. This is called the *terminal* year.

13.1.4 Intelligence Quotient

The intelligence quotient, the ratio of an individual's mental age to his chronological age, is found by the formula :

$$IQ = \frac{MA}{CA}(100)$$

The ratio is multiplied by 100 to remove the decimal.

An individual's IQ indicates *rate of mental development* or *degree of brightness.* If mental development keeps pace with one's life age, the quotient is 100. If mental development lags, or is accelerated, the quotient will be less than or greater than 100, depending upon the degree of retardation or acceleration.

It is clear that mental age alone does not adequately represent an individual's mental capacity; for persons of different, at times widely different, chronological ages may and do reach the same mental age at a given time. One of the values of the IQ, therefore, is to reflect these age differences; hence it is defined in terms of rate of mental development and, as an attribute, degree of brightness.

In his volume accompanying the 1916 Stanford-Binet scale, Terman included a table showing the percentage of children at each of a number of IQ levels, each of which he gave a name; for instance, dull, normal, very superior. An individual whose test performance is normal for his chronological age earns an IQ rating of 100.

IQ range	Terman's categories
80-89	dullness
90-109	average intelligence
110-119	superior intelligence
120-140	very superior intelligence

As in any such table, the limits of each category were arbitrarily determined. These and similar categories are intended to serve only as guides in the interpretation of intelligence quotients and for purposes of statistical classification and analysis.

There are a number of problems associated with the interpretation and use of the IQ that are explained in later sections of this and other chapters. Our purpose at this stage is primarily to define and explain the meaning of the concept.

Did u know? The use of this index was first suggested by Stern (17) and Kuhlmann (12) in 1912; but it was not actually employed as part of test findings and reports until 1916 when the first edition of the Stanford-Binet scale was made available.

13.1.5 Deviation IQ

The "deviation IQ" is an adaptation of the standard score (z) technique. The method of determining the deviation IQ can be shown by using the Wechsler test's procedure as an illustration. For each individual, the raw score is converted into a weighted score by using a conversion table. The mean weighted score of the group is given a deviation IQ value of 100; the *standard deviation* of the scores is equated with a deviation IQ value of 15. Thus, a person whose point score places him at − 1 SD will have a deviation IQ of 85. One whose score is at −2 SD will have a rating of 70. Similarly, positive SD values will give ranks above 100 : +1 SD equals 115 deviation IQ; +2 SD equals 130; and so forth.

The 1960 revision of the Stanford-Binet scale also uses the deviation IQ, calculated by a different method; but the *basic principle* is the same. The principle is that an individual's intelligence quotient should be determined by the *relative* extent to which his score on the test deviates from the mean of his age group, and that an intelligence quotient of a given value should have the same *relative* significance throughout the age range. These ends are now achieved by using units of standard deviation as the basis; hence the name of the new index.

Making the mean score equal to a deviation IQ of 100 is readily understandable, since this value has long been conventional and is accepted as representing the average or normal. It also appears that the most probable standard deviation of intelligence quotients is 15-16, as found with the Stanford-Binet (which in many ways is regarded as a criterion); hence 15 has been taken to represent the standard deviation of the newer index. The choice of this value, therefore, was not an arbitrary one. Furthermore, the distribution yielded by a standard deviation of 15 points is very similar to the one to which psychologists and educators have become accustomed, and in which values at each of the several levels have acquired qualitative significance in regard to mental ability and educational promise.

The deviation IQ, furthermore, is especially useful at age levels above 16 or 18 years. For these and older persons, the use of mental age and the formula for the ratio IQ (MA/CA) have been regarded as inappropriate and questionable by many psychologists. (This aspect of mental age is discussed in Chapter 10.)

It should be clear, from the materials thus far presented, that percentiles, standard deviations, standard scores, and intelligence quotients are intimately related. Whatever index is used, its principal significance is found in the relative rank it represents and in its psychological, educational, and vocational connotations.

Although the primary purpose of this section is to define and explain these concepts used in psychological testing, it is relevant here to emphasize the *qualitative* aspects of these indexes.

Self Assessment

1. *Fill in the blanks :*

   (*i*) Relative significance of any given score is known as _____ .

  (*ii*) An individual's _____ on a test designates the percentage of cases or scores lying below it.

 (*iii*) _____ signifies a range of scores between two dividing points.

 (*iv*) An individual's _____ indicates rate of mental development or degree of brighness.

        (Intelligence quotient)

  (*v*) The _____ is especially useful at age levels above 16 or 18 years.

13.2 Qualitative Interpretation of Test Scores

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

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

The boy with an IQ of 50 probably will not be able to complete more than the second grade; the boy having the IQ of 100 should be able to complete twelve grades; the boy with an IQ of 150 will be able to progress in education as far as his interests and motives indicate. Obviously, too, the kinds of occupations that will be open to the first boy are very limited; those open to the second will be numerous; those open to the third will be practically unrestricted, so far as mental capacity is concerned. And the same may be said of the range of interests in general that will be within the scope of each. These facts are of educational and clinical significance, but at present there are no psychological or statistical means whereby one can calculate how many *times* more or less capable one person is than another.

Caution is necessary at this point. The inferences drawn in the preceding paragraph cannot be based solely upon the numerical IQ value without reference to the clinical features in the test performances or other factors not shown by the numerical index. We have assumed that there are no complicating factors and that the IQs are valid measures of the capacities and performances of the three boys. The boy with 150 IQ, however, *might be* an unstable personality who is failing in many or all of his school subjects. The boy with 100 IQ *might have been* penalized on the test by a language handicap. And the boy with an IQ of 50 *might* show a "scatter" (inconsistency and variation) of performance indicating emotional disturbance rather than intellectual impoverishment. Occasionally, also, it will be found that a high test rating *may be* attributable to an inconsistently high level of performance on one or a few types of subtests (for example,

memory span or word knowledge), just as, conversely, it occasionally happens that a person's IQ is depressed by an inconsistently low performance on one or a few subtests. "Inconsistent" means that the individual's levels of performance on these few subtests differ markedly, in one direction or another, from the general and more consistent levels of his scores on the other subtests.

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

13.2.1 Indexes Used with Educational Achievement Tests

Educational Age : The educational age index (EA) represents a pupil's average level of achievement in a group of school subjects, measured by means of standardized tests, and in terms of the average for various chronological ages in school. If, for example, a pupil's performance on the tests is at the average 12-year level, his EA is 12.

There is no fixed or uniform list of school subjects for which educational ages are obtained; therefore they are not all necessarily comparable. Furthermore, even if achievement in the same school subjects is measured, but with different tests, the EAs might still not be comparable because of possible differences in the standardization process.

Educational age is used, at times, to estimate the probable grade level at which a pupil's test performance places him, since the average age for each grade is known. This practice, however, is of doubtful merit. If a grade estimate or comparison is wanted, then grade norms should be used. These are provided by all the sound educational achievement tests.

Educational Quotient : As was to be expected, an "age" would be accompanied by a quotient. The educational quotient indicates, presumably, whether a pupil's knowledge of a group of school subjects is commensurate with his chronological age, or whether it is above or below the level to be expected of him for his age. The simple formula, therefore, is :

$$EQ = \frac{EA}{CA}(100)$$

the ratio being multipled by 100 to remove the decimal. It is hardly necessary to mention that the value of obtained educational quotients will depend upon the achievement test's reliability and validity.

13.2.2 Achievement Quotient

This index, AQ, the use of which was suggested in 1920 (7), is now rarely used. It is found by dividing educational age by mental age (EA/MA). The reason for using MA as the divisor, instead of CA, is that the former is regarded as the more valid index of a pupil's learning capacity. Hence, it was believed, dividing EA by MA yields a quotient that indicates whether or not the individual is working up to his mental capacity, as found by the intelligence test. Although it is true that mental age is a more valid index of learning capacity and educational promise than chronological age alone, the AQ has some serious defects, which account for its virtual abandonment. First, it is erroneous, to expect all mentally superior pupils to be at a level of educational achievement equal to their mental age, as the following example shows : assume that a superior child has completed the first grade at the age of 7; he has a mental age of 10, and, thus, an IQ of 143. To get an AQ of 100, he would have to earn an EA of 10. To obtain this EA, he must acquire, in one year of schoolwork, as much as the ordinary child is expected to acquire in four. It is not probable that this will happen. To generalize this point : the superior child, especially in the lower school grades, has not had, and frequently does not have, the time and length of schooling necessary to learn the amount of subject matter necessary to equate EA with MA.

A second defect in using the AQ is that frequently the population samples upon which the educational achievement tests have been standardized are not comparable with those upon which

the norms of the intelligence tests have been based. Generally, the former are less representative of the population and are dependent, of course, upon the quality of the schools in which the standardization process was carried out.

A third defect is the fact that many achievement tests do not differentiate as well among pupils as does a sound test of general intelligence. This fact tends to reduce the variability of the former and its correlation with the latter.

Currently, for the purpose of indicating a pupil's school achievement, the EA and EQ have value; but they should be supplemented by each individual's percentile rank within the distribution of scores for his grade. Since all sound tests cover a range of several grades, it is possible, if necessary, to compare any individual's score with the norms in grades above or below his own, for the purpose of finding his percentile rank within those, other levels.

13.3 Quantitative Interpretation of Test Scores

Psychological tests are standardized on the basis of the performance of a representative population; and an individual's rating is determined by the relationship of his performance to that of a group as a whole. Thus we have the several "ages" (for example, mental age) and "quotients" (for example, intelligence quotient), percentile and decile ranks, and standard scores. Any useful test should yield one or more of these. In more recent years, however, without denying the usefulness and value of these indexes of relative status, increasing emphasis has been placed upon "patterns" of performance as clinical aids to psychological diagnosis and counseling.

A person's responses to tests are now frequently analyzed for the purpose of discovering whether he shows any special abilities or disabilities, whether there are marked discrepancies between responses on some types of materials as against responses on others, or whether certain psychological processes seem to be impaired or are markedly superior to others within the individual. A general contrast, for example, might be found between tests involving verbal materials and those which are nonverbal in character; the associative processes might be disturbed; memory or spatial perception might be found to deviate markedly in one direction or another from an individual's general level of capacity. Recent investigations have indicated that patterns of response may be useful in differentiating and diagnosing the several categories of maladjusted and abnormal personalities, as well as for discerning more clearly the mental defectives.

Also, it has been found that persons of equivalent general mental status may have different patterns of performance, or abilities, which in sum, nevertheless, give them much the same over-all and general ratings in terms of a single index (mental age, percentile rank). That is to say, it is possible for two persons to have test ratings that are *numerically* similar and yet have dissimilar "mental organizations," since the components of each total rating differ to a greater or lesser degree from those of the other.

If, therefore, the psychologist's concern is not primarily with group trends or averages, but rather with a particular individual, of course he will want to know the age level of performance and the consequent quotient; but he will also analyze the details of the individual's performance for the purpose of discovering that person's particular pattern or idiom, in order to discern his particular form of mental organization, specific evidences of retardation or disability, if any, and details of his development.

In more recent years there has been a partial shift in emphasis from almost exclusive concern with the analysis of abilities and methods of psychological measurement, as such, to an examination of individual performance and individual idiom, and to the individual as a functioning and dynamic unit. After all, any given test measures only a segment of a total personality; that segment is an integral part of the totality and is influenced by the whole. Hence, the psychologist who is concerned with insight into the nature of an individual's abilities must be able to *evaluate*

a person's performance as well as *measure* it. The data and indexes derived from psychological tests are, for the most part, objectively determined; but their clinical use involves judgment, subjective assessment, and interpretation, based upon a variety of data from several sources. The experienced clinical examiner will supplement the test's numerical results with his observations of the testee's attitudes during the examination and the manner in which he attacks the problems of the test : his degree of confidence or dependence, his cooperativeness or apathy, his negativism or resentment, the richness or paucity of his responses. The individual test situation thus can be, in effect, an occasion lor general psychological observations—really a penetrating psychological interview.

Ability not only to score a test but also to assess and interpret responses and to evaluate the individual's behavior during the examination is a clinical skill the psychologist develops from working with persons rather than with tests alone. However, for the practice of his skill he must, of course, thoroughly understand the psychological and statistical foundations and hypotheses upon which the tests are based.

A few specific instances of the *qualitative* analysis and interpretation of test responses will illustrate the kinds of observations that constitute the clinical aspects that supplement numerical scoring.

Word definitions are generally acceptable at a fairly elementary level; but they vary in level and quality from purely concrete, to functional, to conceptual or abstract. Differences in quality level are indicative of differences in modes of thinking. It also happens, at times, that some words are emotionally charged for the examinee, in which case his definition and behavioral response may be revealing.

Some test items permit the exercise of considerable freedom in response. These responses may reveal the examinee's attitudes, values, and modes of meeting life situations. In this category are test items that ask, "What is the thing to do when . . . ?" Or, "Why should we   . . ?" The subject's reactions to such items, the qualities of his verbalizations in making the responses, and the presence or absence of strong feelings reveal some of the nonintellective aspects of his personality.

The subject's specific comments while performing a task are of possible significance in regard to his attitude toward himself, or toward an authority figure (the examiner), or toward other individuals and institutions in his environment.

Responses to items or random comments may reveal hostilities and anxieties, or wholesome cooperativeness and security.

The manner of speech—the use of expletives, halting and fumbling, restless movements, blushing, or, on the other hand, a relaxed attitude, mild criticism of one's own performance — provides valuable clues to the testee's personality.

Character disorders may be indicated by impetuous and uncritical responses that are incorrect but are given with assurance and pretentiousness.

The subject's ability to direct his attention toward, to concentrate upon, and to organize a task are often revealed by his mode of approach to a test problem.

The selective character, if any, of a person's vocabulary and information (two subtests widely used) will shed light upon his experiences, interests, cultural background.

A personality trait such as compulsiveness (as opposed to desirable thoroughness and self-criticism) may be revealed by excessively detailed responses and by numerous and unnecessary alternative responses.

Some types of responses indicate pathological or psychotic states : erroneously bizarre responses by an otherwise intelligent person (for example, London is in Africa; the population of the United States is 1,500,000); disjointed and irrelevant responses; and distorted interpretations of the task or problem.

Organic damage may be detected through selected kinds of subtests; for example, disturbance of the visual-motor function as indicated by the diamond copying test (Stanford-Binet) and the object assembly test (Bellevue), among others.

Scatter analysis (discussed in detail in Chapter 14) is essential to the discernment of superior, inferior, and impaired psychological functions.

Sensitized observations on the part of the examiner will enable him, in general, to evaluate *how* the subject proceeded in both success and failure.

Task   What is Educational Quotient ?

Self Assessment

2. State whether the following statements are 'true' or 'false'.

  (i) The educational age index represents a pupil's average of achievement in a group of school subjects, measured by means of standardized tests.

 (ii) Achievement tests are standardized on the basis of the performance of a representative population.

(iii) The educational age index (EA) represents a pupil's average level of achievement in a group of school subjects.

13.4 Summary

• Relative Rank : The raw score (that is, the actual number of units or points) obtained by an individual on a test does not in itself have much, if any, significance.

• Percentile Ranks : An individual's percentile rank on a test designates the percentage of cases or scores lying below it.

• The percentile method is a technique whereby scores on two or more tests, given in units that are different, may be transformed into uniform and comparable values. This method has the advantage of not depending upon any assumptions regarding the characteristics of the distribution with which it is used. The distribution might be normal, skewed, or rectangular.

• The percentile technique has the advantage of being easily calculated, easily understood, and of making no assumptions with regard to the characteristics of the tolal distribution.

• Decile Rank : The decile rank is the same in principle as the percentile; but instead of designating the one-hundredth part of a distribution, it designates the one-tenth part of the group (N/10) in which any tested person is placed by his score. The term "decile" is used to mean a dividing point. "Decile rank" signifies a range of scores between two dividing points.

• Mental Age : This concept was introduced by Alfred Binet in 1908 in conjunction with the first revision of his scale. In this scale and in its later revision, items are grouped according to age levels. For example, selected items, passed by a specified percentage of five-year-old children in the standardization sample are placed at the five-year level.

• To determine mental age, in the 1908 scale, Binet adopted the following rule : the child was credited with the mental age of the highest year level in which he passed all test items, or all but one.

- Mental age norms are also used with scales that are not arranged according to age levels. These are *point scales* that yield a score usually based on the number of items correctly answered.

- Intelligence Quotient : The intelligence quotient, the ratio of an individual's mental age to his chronological age, is found by the formula :

$$IQ = \frac{MA}{CA}(100)$$

- An individual's IQ indicates *rate of mental development* or *degree of brightness*. If mental development keeps pace with one's life age, the quotient is 100. If mental development lags, or is accelerated, the quotient will be less than or greater than 100, depending upon the degree of retardation or acceleration.

- Deviation IQ : The "deviation IQ" is an adaptation of the standard score (*z*) technique. The method of determining the deviation IQ can be shown by using the Wechsler test's procedure as an illustration.

- Qualitative significance of the intelligence quotient can be illustrated further by asking this question :

- At present there are no psychological or statistical means whereby one can calculate how many *times* more or less capable one person is than another.

- Educational Age : The educational age index (EA) represents a pupil's average level of achievement in a group of school subjects, measured by means of standardized tests, and in terms of the average for various chronological ages in school.

- Educational Quotient : As was to be expected, an "age" would be accompanied by a quotient. The educational quotient indicates, presumably, whether a pupil's knowledge of a group of school subjects is commensurate with his chronological age.

- Achievement Quotient : This index, AQ, the use of which was suggested in 1920 (7), is now rarely used. It is found by dividing educational age by mental age (EA/MA). The reason for using MA as the divisor, instead of CA, is that the former is regarded as the more valid index of a pupil's learning capacity.

- Scores, whether raw or converted, do not suffice for the complete interpretation of an individual's performances on psychological tests. The several, aspects of test standardization thus far presented are concerned with the performance of *groups* of persons and with *average relationships* revealed by statistical treatment of results.

- Psychological tests, as already noted, are standardized on the basis of the performance of a representative population; and an individual's rating is determined by the relationship of his performance to that of a group as a whole.

- A person's responses to tests are now frequently analyzed for the purpose of discovering whether he shows any special abilities or disabilities, whether there are marked discrepancies between responses on some types of materials as against responses on others, or whether certain psychological processes seem to be impaired or are markedly superior to others within the individual.

- The subject's specific comments while performing a task are of possible significance in regard to his attitude toward himself, or toward an authority figure (the examiner), or toward other individuals and institutions in his environment.

- The manner of speech—the use of expletives, halting and fumbling, restless movements, blushing, or, on the other hand, a relaxed attitude, mild criticism of one's own performance
  - provides valuable clues to the testee's personality.

Notes

- Character disorders may be indicated by impetuous and uncritical responses that are incorrect but are given with assurance and pretentiousness.

- The selective character, if any, of a person's vocabulary and information (two subtests widely used) will shed light upon his experiences, interests, cultural background.

- A personality trait such as compulsiveness (as opposed to desirable thoroughness and self-criticism) may be revealed by excessively detailed responses and by numerous and unnecessary alternative responses.

- Organic damage may be detected through selected kinds of subtests; for example, disturbance of the visual-motor function as indicated by the diamond copying test (Stanford-Binet) and the object assembly test (Bellevue), among others.

- Scatter analysis (discussed in detail in Chapter 14) is essential to the discernment of superior, inferior, and impaired psychological functions.

- Sensitized observations on the part of the examiner will enable him, in general, to evaluate *how* the subject proceeded in both success and failure.

13.5 Keywords

Relative   : Considered and judged by being compared with some else

Quantitative : Connected with the amount or number of something rather than with how good it is

13.6 Review Questions

1. Explain the terms "Percentile and Decile" ranks.

2. What is intelligent Quotient ?

3. Discuss about qualitative interpretation of test scores.

4. Describe the quantitative interpretation of test scores.

Answer : Self Assessment

1. (*i*)   relative rank   (*ii*)   percentile rank

    (*iii*) Decile rank   (*iv*) Intelligence quotient

    (*v*)   deviation IQ

2. (*i*)   T   (*ii*)   F   (*iii*)   T

13.7 Further Readings

Books

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 14 : Criterion Referenced Test

CONTENTS

Objectives

Introduction

14.1 Meaning of Criterion Referenced Test (CRT)

14.2 Features of Criterion Referenced Test (CRT)

14.3 Advantages of Criterion Referenced Test (CRT)

14.4 Disadvantages of Criterion Referenced Test (CRT)

14.5 Applications of Criterion Referenced Test (CRT)

14.6 Summary

14.7 Keywords

14.8 Review Questions

14.9 Further Readings

Objectives

The objectives of this unit can be summarized as below :

- To explain the meaning of criterion referenced test
- To discuss about the features of criterion referenced test
- To describe about the advantages and disadvantages of criterion referenced test
- To describe the applications of criterion referenced test

Introduction

A *criterion-referenced test* is a test that provides a basis for determining a candidate's level of knowledge and skills in relation to a well-defined domain of content. Often one or more performance standards are set on the test score scale to aid in test score interpretation. Criterion-referenced tests, a type of test introduced by Glaser (1962) and Popham and Husek (1969), are also known as domain-referenced tests, competency tests, basic skills tests, mastery tests, performance tests or assessments, authentic assessments, objective-referenced tests, standards-based tests, credentialing exams, and more. What all of these tests have in common is that they attempt to determine a candidate's level of performance in relation to a well-defined domain of content. This can be contrasted *with norm-referenced tests*, which determine a candidate's level of the construct measured by a test in relation to a well-defined reference group of candidates, referred to as the norm group. So it might be said that criterion-referenced tests permit a candidate's score to be interpreted in relation to a domain of content, and norm-referenced tests permit a candidate's score to be interpreted in relation to a group of examinees. The first interpretation is content-centered, and the second interpretation is examinee-centered.

14.1 Meaning of Criterion Referenced Test (CRT)

Definition : A test in which questions are written according to specific predetermined criteria. A student knows what your standards are for passing and only competes against him or herself while completing.

A criterion-referenced test is designed to measure how well test takers have mastered a particular body of knowledge.

These tests generally have an established "passing" score. Students know what the passing score is and an individual's test score is determined by knowledge of the course material.

He runs a second tryout, having established that 10 seconds in the 100 yard dash is competitive in the event. He now picks those who run the dash in 10 seconds or less. This is criterion-referenced testing. He knows the runners he selected can compete. He gets the funds.

Notes   The term "criterion-referenced test" is not part of the everyday vocabulary in schools, and yet, nearly all students take criterion-referenced tests on a routine basis.

14.2 Features of Criterion Referenced Test (CRT)

Features of Criterion reference tests are as follows:-

(*i*)   Criterion-referenced test place a primary focus on the content and what is being measured. Norm-referenced tests are also concerned about what is being measured but the degree of concern is less since the domain of content is not the primary focus for score interpretation. In norm-referenced test development, item selection, beyond the requirement that items meet the content specifications, is driven by item statistics. Items are needed that are not too difficult or too easy, and that are highly discriminating. These are the types of items that contribute most to score spread, and enhance test score reliability and validity.

(*ii*)   With criterion-referenced test development, extensive efforts go into insuring content validity. Item statistics play less a role in item selection though highly discriminating items are still greatly valued, and sometimes item statistics are used to select items that maximize the discriminating power of a test at the performance standards of interest on the test score scale.

A good norm-referenced test is one that will result in a wide distribution of scores on the construct being measured by the test. Without score variability, reliable and valid comparisons of candidates cannot be made. A good criterion-referenced test will permit content-referenced interpretations and this means that the content domains to which scores are referenced must be very clearly defined. Each type of test can serve the other main purpose (norm-referenced versus criterion-referenced interpretations), but this secondary use will never be optimal. For example, since criterion-referenced tests are not constructed to maximize score variability, their use in comparing candidates may be far from optimal if the test scores that are produced from the test administration are relatively similar.

Because the purpose of a criterion-referenced test is quite different from that of a norm-referenced test, it should not be surprising to find that the approaches used for reliability and validity assessment are different too.

(*iii*)   With criterion-referenced tests, scores are often used to sort candidates into performance categories. Consistency of scores over parallel administrations becomes less central than consistency of classifications of candidates to performance categories over parallel administrations. Variation in candidate scores is not so important if candidates are still assigned to the same performance category.

Therefore, it has been common to define reliability for a criterion-referenced test as the extent to which performance classifications are consistent over parallel-form administrations.

For example, it might be determined that 80% of the candidates are classified in the same way by parallel forms of a criterion-referenced test administered with little or no instruction in between test administrations. This is similar to parallel form reliability for a norm-referenced test except the focus with criterion-referenced tests is on the decisions rather than the scores. Because parallel form administrations of criterion-referenced tests are rarely practical, over the years methods have been developed to obtain single administration estimates of decision consistency that are analogous to the use of the corrected split-half reliability estimates with norm-referenced tests.

(*iv*) With criterion-referenced tests, the focus of validity investigations is on (1) the match between the content of the test items and the knowledge or skills that they are intended to measure, and (2) the match between the collection of test items and what they measure and the domain of content that the tests are expected to measure. The "alignment" of the content of the test to the domain of content that is to be assessed is called *content validity evidence*. This term is well known in testing practices.

Many criterion-referenced tests are constructed to assess higher-level thinking and writing skills, such as problem solving and critical reasoning. Demonstrating that the tasks in a test are actually assessing the intended higher-level skills is important, and this involves judgments and the collection of empirical evidence. So, construct validity evidence too becomes crucial in the process of evaluating a criterion referenced test.

(*v*) Probably the most difficult and controversial part of criterion-referenced testing is setting the performance standards, i.e., determining the points on the score scale for separating candidates into performance categories such as "passers" and "failers." The challenges are great because with criterion-referenced tests in education, it is common on state and national assessments to separate candidates into not just two performance categories, but more commonly, three, four, or even five performance categories. With four performance categories, these categories are often called *failing, basic, proficient, and advanced.*

Did u know? Criterion-referenced tests, a type of test introduced by Glaser (1962) and Popham and Husek (1969), are also known as domain-referenced tests, competency tests, basic skills tests, mastery tests, performance tests or assessments, authentic assessments, objective-referenced tests, standards-based tests, credentialing exams, and more.

14.3 Advantages of Criterion Referenced Test (CRT)

1. **Mastery of Subject Matter.**

    • Criterion-referenced tests are more suitable than norm-referenced tests for tracking the progress of students within a curriculum. Test items can be designed to match specific program objectives. The scores on a criterion referenced tests indicate how well the individual can correctly answer questions on the material being studied, while the scores on a norm-referenced test report how the student scored relative to other students in the group.

    Criterion-Referenced Tests can be Managed Locally.

    • Assessing student progress is something that every teacher must do. Criterion-referenced tests can be developed at the classroom level. If the standards are not met, teachers can specifically diagnose the deficiencies. Scores for an individual student are independent of

how other students perform. In addition, test results can be quickly obtained to give students effective feedback on their performance. Although norm-referenced tests are most suitable for developing normative data across large groups, criterion-referenced tests can produce some local norms.

14.4 Disadvantages of Criterion Referenced Test (CRT)

• Criterion-referenced tests have some built-in disadvantages. Creating tests that are both valid and reliable requires fairly extensive and expensive time and effort. In addition, results cannot be generalized beyond the specific course or program. Such tests may also be compromised by students gaining access to test questions prior to exams. Criterion-referenced tests are specific to a program and cannot be used to measure the performance of large groups.

Analyzing Test Items

• Item analysis is used to measure the effectiveness of individual test items. The main purpose is to improve tests, to identify questions that are too easy, too difficult or too susceptible to guessing. While test items can be analyzed on both criterion-referenced and norm-referenced tests, the analysis is somewhat different because the purpose of the two types of tests is different.

14.5 Applications of Criterion Referenced Test (CRT)

Criterion-referenced tests are used in many ways. Classroom teachers use them to monitor student performance in their day-to-day activities. States find them useful for evaluating student performance and generating educational accountability information at the classroom, school, district, and state levels. The tests are based on the curricula, and the result provide a basis for determining how much is being learned by students and how well the educational system is producing desired results. Criterion-referenced tests are also used in training programs to assess learning. Typically pretest-posttest designs with parallel forms of criterion-referenced tests are used. Finally, criterion-referenced tests are used in the credentialing field to determine persons qualified to receive a license or certificate.

Task   What is the use of criterion reference tests are used in credentialing field.

Self Assessment

1. Fill in the blanks :

  (*i*) _____ test provides basis for determining a candidate's level of knowledge and skills in relation to a well defined domain of Content.

 (*ii*) Criterion reference tests have an established _____ scores.

(*iii*) _____ use criterion reference test to monitor student performance in their day to day activities.

(*iv*) Criterion reference tests are also used in the training programs to assess _____.

14.6 Summary

• A test in which questions are written according to specific predetermined criteria. A student knows what your standards are for passing and only competes against him or herself while completing.

- A criterion-referenced test is designed to measure how well test takers have mastered a particular body of knowledge.

- The term "criterion-referenced test" is not part of the everyday vocabulary in schools, and yet, nearly all students take criterion-referenced tests on a routine basis.

- Criterion-referenced tests place a primary focus on the content and what is being measured.

- With criterion-referenced test development, extensive efforts go into insuring content validity. Item statistics play less a role in item selection though highly discriminating items are still greatly valued, and sometimes item statistics are used to select items that maximize the discriminating power of a test at the performance standards of interest on the test score scale.

- With criterion-referenced tests, scores are often used to sort candidates into performance categories. Consistency of scores over parallel administrations becomes less central than consistency of classifications of candidates to performance categories over parallel administrations.

- It has been common to define reliability for a criterion-referenced test as the extent to which performance classifications are consistent over parallel-form administrations.

- With criterion-referenced tests, the focus of validity investigations is on (1) the match between the content of the test items and the knowledge or skills that they are intended to measure, and (2) the match between the collection of test items and what they measure and the domain of content that the tests are expected to measure.

- Many criterion-referenced tests are constructed to assess higher-level thinking and writing skills, such as problem solving and critical reasoning. Demonstrating that the tasks in a test are actually assessing the intended higher-level skills is important, and this involves judgments and the collection of empirical evidence.

- Most difficult and controversial part of criterion-referenced testing is setting the performance standards, i.e., determining the points on the score scale for separating candidates into performance categories such as "passers" and "failers." The challenges are great because with criterion-referenced tests in education, it is common on state and national assessments to separate candidates into not just two performance categories, but more commonly, three, four, or even five performance categories.

- Criterion-referenced tests are more suitable than norm-referenced tests for tracking the progress of students within a curriculum. Test items can be designed to match specific program objectives.

- Assessing student progress is something that every teacher must do. Criterion-referenced tests can be developed at the classroom level. If the standards are not met, teachers can specifically diagnose the deficiencies.

- Criterion-referenced tests have some built-in disadvantages. Creating tests that are both valid and reliable requires fairly extensive and expensive time and effort. In addition, results cannot be generalized beyond the specific course or program.

- Item analysis is used to measure the effectiveness of individual test items. The main purpose is to improve tests, to identify questions that are too easy, too difficult or too susceptible to guessing.

- Criterion-referenced tests are used in many ways. Classroom teachers use them to monitor student performance in their day-to-day activities. States find them useful for evaluating student performance and generating educational accountability information at the classroom, school, district, and state levels.

14.7 Keywords

Test   :   An examination of somebody's knowledge or ability, consisting of questions for then to answer or activities for them to carry out.

criterion   :   A standard or principal by which something is judged

Predetermined   :   to decide something in advance so that it does not happen by chance.

14.8 Review Questions

1. What do you mean by criterion referenced test ?

2. Give the advantages and disadvantages of criterion referenced test.

3. Explain the feature of criterion referenced test.

4. What are the applications of criterion referenced test

Answers : Self Assessment

(*i*)   Criterion referenced test   (*ii*)   Passing

(*iii*)   Classroom teachers   (*iv*)   learning

14.9 Further Readings

Books

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 15 : Norm Referenced Test

Objectives

The objectives of this unit can be summarised as below :

- To explain the meaning of norm referenced test
- To discuss about the need for norm referenced test
- To describe the features of norm referenced test
- To discuss about the accuracy of test score
- To explain about the advantages and disadvantages of norm referenced test

Introduction

A **norm-referenced test** (NRT) is a type of test, assessment, or evaluation which yields an estimate of the position of the tested individual in a predefined population, with respect to the trait being measured. This estimate is derived from the analysis of test scores and possibly other relevant data from a sample drawn from the population.[1] That is, this type of test identifies whether the test taker performed better or worse than other test takers, but not whether the test taker knows either more or less material than is necessary for a given purpose.

15.1 Meaning of Norm Referenced Test (NRT)

Definition : This type of test determines a student's placement on a normal distribution curve. Students compete against each other on this type of assessment. This is what is being referred to with the phrase, 'grading on a curve'. Norm-referenced tests allow us to compare a student's skills to others in his age group. Norm-referenced tests are developed by creating the test items and then administering the test to a group of students that will be used as the basis of comparison. Statistical methods are used to determine how raw scores will be interpreted and what performance levels are assigned to each score. Many tests yield standard scores, which allow

comparison of the student's scores to other tests. They answer questions such as, "does the student's achievement score appear consistent with his cognitive score ?" The degree of difference between those two scores might suggest or rule out a learning disability.

After the norming process, the tests are used to assess groups of students or individuals using standardized, or highly structured, administration procedures. These students' performance is rated using scales developed during the norming process.

Educators use norm-reference tests to evaluate the effectiveness of teaching programs, to help determine students' preparedness for programs, and to determine diagnosis of disabilities for eligibility for IDEA special education programs or adaptations and accommodations under Section 504.

15.2 Need for Norm Referenced Test (NRT)

The major reason for using a norm-referenced tests (NRT) is to classify students. NRTs are designed to highlight achievement differences between and among students to produce a dependable rank order of students across a continuum of achievement from high achievers to low achievers. School systems might want to classify students in this way so that they can be properly placed in remedial or gifted programs. These types of tests are also used to help teachers select students for different ability level reading or mathematics instructional groups.

With norm-referenced tests, a representative group of students is given the test prior to its availability to the public. The scores of the students who take the test after publication are then compared to those of the norm group. Tests such as the California Achievement Test the Iowa Test of Basic Skills (Riverside), and the Metropolitan Achievement Test (Psychological Corporation) are normed using a national sample of students. Because norming a test is such an elaborate and expensive process, the norms are typically used by test publishers for years. All students who take the test during that seven year period have their scores compared to the original norm group.

While norm-referenced tests ascertains the rank of students, criterion-referenced tests (CRTs) determine "...what test takers can do and what they know, not how they compare to others. CRTs report how well students are doing relative to a pre-determined performance level on a specified set of educational goals or outcomes included in the school, district, or state curriculum.

Educators or policy makers may choose to use a CRT when they wish to see how well students have learned the knowledge and skills which they are expected to have mastered. This information may be used as one piece of information to determine how well the student is learning the desired curriculum and how well the school is teaching that curriculum.

Both NRTs and CRTs can be standardized. The U.S. Congress, Office of Technology Assessment (1992) defines a standardized test as one that uses uniform procedures for administration and scoring in order to assure that the results from different people are comparable. Any kind of test-from multiple choice to essays to oral examinations-can be standardized if uniform scoring and administration are used. This means that the comparison of student scores is possible. Thus, it can be assumed that two students who receive the identical scores on the same standardized test demonstrate corresponding levels of performance. Most national, state and district tests are standardized so that every score can be interpreted in a uniform manner for all students and schools.

Human beings make tests. They decide what topics to include on the test, what kinds of questions to ask, and what the correct answers are, as well as how to use test scores. Tests can be made to compare students to each other (norm-referenced tests) or to see whether students have mastered a body of knowledge (criterion or standards-referenced tests). This fact sheet explains what NRTs are, their limitations and flaws, and how they affect schools.

15.3 Features of Norm Referenced Test (NRT)

Norm-referenced tests (NRTs) compare a person's score against the scores of a group of people who have already taken the same exam, called the "norming group." When you see scores in the paper which report a school's scores as a percentage -- "the Lincoln school ranked at the 49th percentile" -- or when you see your child's score reported that way -- "Jamal scored at the 63rd percentile" -- the test is usually an NRT.

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

Creating the bell curve.

NRTs are designed to "rank-order" test takers -- that is, to compare students' scores : A commercial norm-referenced test does not compare all the students who take the test in a given year. Instead, test-makers select a sample from the target student population (say, ninth graders). The test is "normed" on this sample, which is supposed to fairly represent the entire target population (all ninth graders in the nation). Students' scores are then reported in relation to the scores of this "norming" group.

To make comparing easier, testmakers create exams in which the results end up looking at least somewhat like a bell-shaped curve (the "normal" curve, shown in the diagram). Testmakers make the test so that most students will score near the middle, and only a few will score low (the left side of the curve) or high (the right side of the curve).

Scores are usually reported as percentile ranks : The scores range from 1st percentile to 99th percentile, with the average student score set at the 50th percentile. If Jamal scored at the 63rd percentile, it means he scored higher than 63% of the test takers in the norming group. Scores also can be reported as "grade equivalents," "stanines," and "normal curve equivalents."

One more question right or wrong can cause a big change in the student's score : In some cases, having one more correct answer can cause a student's reported percentile score to jump more than ten points. It is very important to know how much difference in the percentile rank would be caused by getting one or two more questions right.

In making an NRT, it is often more important to choose questions that sort people along the curve than it is to make sure that the content covered by the test is adequate : The tests sometimes emphasize small and meaningless differences among testtakers. Since the tests are made to sort students, most of the things everyone knows are not tested. Questions may be obscure or tricky, in order to help rank order the testtakers.

Tests can be biased : Some questions may favor one kind of student or another for reasons that have nothing to do with the subject area being tested. Non-school knowledge that is more commonly learned by middle or upper class children is often included in tests. To help make the bell curve, testmakers usually eliminate questions that students with low overall scores might get right but those with high overall scores get wrong. Thus, most questions which favor minority groups are eliminated.

NRTs usually have to be completed in a time limit : Some students do not finish, even if they know the material. This can be particularly unfair to students whose first language is not English or who have learning disabilities. This "speededness" is one way testmakers sort people out.

Did u know? **Commercial, national, norm-referenced "achievement" tests include** the California Achievement Test (CAT); Comprehensive Test of Basic Skills (CTBS), which includes the "Terra Nova"; Iowa Test of Basic Skills (ITBS) and Tests of Academic Proficiency (TAP); Metropolitan Achievement Test (MAT); and Stanford Achievement Test (SAT, not to be confused with the college admissions SAT). "IQ," "cognitive ability," "school readiness," and developmental screening tests are also NRTs.

15.4 Accuracy of Test Score

The items on the test are only a sample of the whole subject area : There are often thousands of questions that could be asked, but tests may have just a few dozen questions. A test score is therefore an estimate of how well the student would do if she could be asked all the possible questions.

All **tests have "measurement error."** No test is perfectly reliable. A score that appears as an absolute number -- say, Jamal's 63 -- really is an estimate. For example, Jamal's "true score" is probably between 56 and 70, but it could be even further off. Sometimes results are reported in "score bands," which show the range within which a test-takers' "true score" probably lies.

There are many other possible causes of measurement error : A student can be having a bad day. Test-taking conditions often are not the same from place to place (they are not adequately "standardized"). Different versions of the same test are in fact not quite exactly the same.

Sub-scores on tests are even less precise : This is mostly because there are often very few items on the sub-test. A score band for a Juanita's math sub-test might show that her score is between the 33rd and 99th percentile because only a handful of questions were asked.

Scores for young children are much less reliable than for older students : This is because young children's moods and attention are more variable. Also, young children develop quickly and unevenly, so even an accurate score today could be wrong next month.

What do score increases mean ? **If your child's or your school's score goes up on a norm-referenced test, does that mean she knows more or the school is better ?** Maybe yes, maybe not. Schools cannot teach everything. They teach some facts, some procedures, some concepts, some skills -- but not others. Often, schools focus most on what is tested and stop teaching many things that are not tested. When scores go up, it does not mean the students know more, it means they know more of what is on that test.

For example, history achievement test "A" could have a question on Bacon's Rebellion (a rebellion by Black slaves and White indentured servants against the plantation owners in colonial Virginia). Once teachers know Bacon's Rebellion is covered on the exam, they are more likely to teach about it. But if those same students are given history test "B," which does not ask about Bacon's Rebellion but does ask about Shay's Rebellion, which the teacher has not taught, the students will not score as well.

Teaching to the test explains why scores usually go down when a new test is used : A district or state usually uses an NRT for five to ten years. Each year, the score goes up as teachers become familiar with what is on the test. When a new test is used, the scores suddenly drop. The students don't know less, it is just that different things are now being tested.

Notes Multiple-choice and short-answer questions do not measure most knowledge that students need to do well in college, qualify for good jobs, or be active and informed citizens.

15.5 Advantages of Norm Referenced Test (NRT)

To compare students, it is often easiest to use a Norm-Referenced Test because they were created to rank test-takers : It there are limited places (such as in a "Gifted and Talented" program) and choices have to be made, it is tempting to use a test constructed to rank students, even if the ranking is not very meaningful and keeps out some qualified children.

NRT's are a quick snapshot of some of the things most people expect students to learn : They are relatively cheap and easy to administer. If they were only used as one additional piece of information and not much importance was put on them, they would not be much of a problem.

15.6 Disadvantages of Norm Referenced Test (NRT)

The damage caused by using NRTs is far greater than any possible benefits the tests provide : The main purpose of NRTs is to rank and sort students, not to determine whether students have learned the material they have been taught. They do not measure anywhere near enough of what students should learn. They have very harmful effects on curriculum and instruction. In the end, they provide a distorted view of learning that then causes damage to **Norm-Referenced Measures (NRM)**

Most appropriate when one wishes to make comparisons across large numbers of students or important decisions regarding student placement and advancement. Norm-referenced measures are designed to compare students (i.e., disperse average student scores along a importance placed upon high scores, the content of a standardized test can be very influential in the development of a school's curriculum and standards of excellence.

The testing profession, in its *Standards for Educational and Psychological Measurement,* states, "In elementary or secondary education, a decision or characterization that will have a major impact on a test taker should not automatically be made on the basis of a single test score."

Any one test can only measure a limited part of a subject area or a limited range of important human abilities : A "reading" test may measure only some particular reading "skills," not a full range of the ability to understand and use texts. Multiple-choice math tests can measure skill in computation or solving routine problems, but they are not good for assessing whether students can reason mathematically and apply their knowledge to new, real-world problems.

Most NRTs focus too heavily on memorization and routine procedures : Tests like these cannot show whether a student can write a research paper, use history to help understand current events, understand the impact of science on society, or debate important issues. They don't test problem-solving, decision-making, judgement, or social skills.

Tests often cause teachers to overemphasize memorization and de-emphasize thinking and application of knowledge : Since the tests are very limited, teaching to them narrows instruction and weakens curriculum. Making test score gains the definition of "improvement" often guarantees that schooling becomes test coaching. As a result, students are deprived of the quality education they deserve.

Norm-referenced tests also can lower academic expectations : NRTs support the idea that learning or intelligence fits a bell curve. If educators believe it, they are more likely to have low expectations of students who score below average.

Task   What is TAP ?

Self Assessment

1. *Fill in the blanks :*

 (*i*) Norm Referenced Test determines a student's placement on a _____.

 (*ii*) The major reason for using a Norm Referenced Test is to Classify _____.

 (*iii*) Most achievement NRTs are _____ test.

 (*iv*) Norm Referenced Test can lower _____.

15.7 Summary

• This type of test determines a student's placement on a normal distribution curve. Students compete against each other on this type of assessment.

• The major reason for using a norm-referenced tests (NRT) is to classify students. NRTs are designed to highlight achievement differences between and among students to produce a dependable rank order of students across a continuum of achievement from high achievers to low achievers.

• California Achievement Test the Iowa Test of Basic Skills (Riverside), and the Metropolitan Achievement Test (Psychological Corporation) are normed using a national sample of students. Because norming a test is such an elaborate and expensive process, the norms are typically used by test publishers for years.

• Norm-referenced tests (NRTs) compare a person's score against the scores of a group of people who have already taken the same exam, called the "norming group."

• Most achievement NRTs are multiple-choice tests.

• NRTs are designed to "rank-order" test takers -- that is, to compare students' scores.

• To make comparing easier, testmakers create exams in which the results end up looking at least somewhat like a bell-shaped curve.

• Scores are usually reported as percentile ranks.

• One more question right or wrong can cause a big change in the student's score.

• In making an NRT, it is often more important to choose questions that sort people along the curve than it is to make sure that the content covered by the test is adequate.

• Tests can be biased.

• Commercial, national, norm-referenced "achievement" tests include.

• The items on the test are only a sample of the whole subject area.

• All tests have "measurement error."

• There are many other possible causes of measurement error.

• Sub-scores on tests are even less precise.

• Scores for young children are much less reliable than for older students.

• Teaching to the test explains why scores usually go down when a new test is used.

• To compare students, it is often easiest to use a norm-referenced test because they were created to rank test-takers.

• NRT's are a quick snapshot of some of the things most people expect students to learn.

• Any one test can only measure a limited part of a subject area or a limited range of important human abilities.

- Most NRTs focus too heavily on memorization and routine procedures.

- Tests often cause teachers to overemphasize memorization and de-emphasize thinking and application of knowledge.
- The damage caused by using NRTs is far greater than any possible benefits the tests provide.

15.8 Keywords

Norm   : A situation or a pattern of behaviour that is usual or expected

Achievement : The act or process of achieving something

Reference   : The act of asking somebody for help or advice

15.9 Review Questions

1. What do you understand by norm referenced test ?
2. What are the features of norm referenced test ?
3. Give the advantages and disadvantages of norm referenced test ?
4. Why is the need of norm referenced test ?

Answers : Self Assessment

 (*i*) normal distribution curve   (*ii*) Students

 (*iii*) Multiple choice   (*iv*) academic expectations

15.10 Further Readings

Books

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*
2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*
3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*
4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 16 : Factors Influencing Test Scores : Nature of Test, Psychological Factors and Environmental Factors

---

CONTENTS

Objectives

The objectives of this unit can be summarised as below :

- To explain about standardized test score

- To discuss about the nature of test

- To describe about the psychological factors

- To discuss about the environmental factors

Introduction

There are numerous factors that can affect test scores. In fact, it's unrealistic to attempt to create an all-encompassing list. Rather, there are more common factors that affect test scores, for one. Secondly, there are similarities between various influences on testing results. More specifically, some factors relate to the student, some to the parents, teachers, etc. People are highly influenced by their environment, and this is often more true of children.

Obviously, the ability of the child is going to influence their test scores. Working hard can result in success in a variety of areas, and it's not beneficial to make excuses for failing. Nonetheless, there are situations where children are unable to do well in certain subjects. People aren't inclined to accept the idea that people have biological limitations preventing them from becoming the next Einstein. In many ways, deny one's limitations is beneficial as people regularly give up before what might have been a critical stage in their mathematical or artistic growth, for example. It can be difficult to determine what causes low test scores, in some cases, but sometimes individual ability is a factor. Here we will discuss about various factrs influencing test scores. Here we will discuss about various factors influencing test scores.

16.1 Standaridized Test Scores

A standardized test score is an indication of how much knowledge a child has accumulated and

can put down on paper on a certain day within a limited time frame. That is all it is : an indication. There are many additional factors, some of which are more important, that a teacher must consider before assigning a mark for a report card, or deciding on the appropriate placement for a student for the upcoming term.

The quality of a student's daily work is important as well as his marks on unit tests and term tests. Is his Contribution to projects as well done as that of his peers ? Is he responsible about completing Homework assignments ? Does he have a positive attitude toward school and education?

There are many factors which can influence the results of standardized tests.

Notes   Most people are capable of doing something, but they don't study properly. They equate studying during commercials with a lengthy study session, or they believe that "they know it" simply because they can answer the questions with unlimited time and access to the material.

16.2 Nature of Test

Marks awarded by the teachers or examiners on a given test are the "test scores' on whose basis varied decisions are taken about the students performance, teaching effectiveness and performance standards. In fact the students' test score on an examination depends on many factors of course one of these is student's knoweldge, understanding or ability in the subject-matter of examination, which determines the student's final mark. This factor influence the student's own Score (not of others) but there are other factors that influence score of all students, more or less equally. These are given below :

(*i*) **Difficulty level of Examination :** It is not possible to judge the exact difficulty level of an examination question at the time of its Construction. It may turn out to be very difficult when hardly a few students pass. Should we fail the whole class ? No if it turns out to be too easy that more than 75% of students get distinction marks. Should then every one get high or very high scores ? No. some adjustment has to be made while assigning the final marks.

(*ii*) **Nature of group tested :** In a particular school the level of students may be high compared with that of students from another school. This would affect the difficulty level of a question for a particular group of students.

(*iii*) **Instructional effectiveness :** Level of teaching may very from one school to another. Accordingly the teaching impact on students learning would also vary, which in turn may affect the difficulty level of an examination from school to school.

(*iv*) **Miscalculation of time limit :** If the teacher or the paper setter makes a wrong guess about the total time required to attempt the questions, he may give more questions than required for students in a give span of time. the students may not be able to attempt all questions within the time limit (miscalculated or estimated)

In that case should the marks of all students be lowered just because the teacher or paper setter made a wrong guess ? cortainly not.

(*v*) **Sampling inadequacies :** To what extent the questions set Cover adequately the prescribed Course Content by the teacher or paper setter, effecting the marks especially in essay-type questions, which usually are loss in number and do not ensure good coverae ? Thus chance of inadequate Converage ? Thus chance of inadequate coverage in testing or even in teaching does affect the row score of students, which ultimately effects many decisions.

16.3 Psychological Factors

The various psychological factors are as follows:-

- **Anxiety :** Some children "freeze" in a testing situation. They are unable to focus or concentrate long enough to give a true picture of their actual knowledge in a subject area. This is especially true when parents are stressed about the test, and have pressured him to do well.

- **Physical Health :** A child may be coming down with the flu, have a headache, a sore throat, or be overtired. He may have slept in, missed breakfast and be hungry. All of these can affect test performance.

- **Emotional Health :** The student may be upset because parents are in the midst of separation or divorce proceedings. A close relative may be seriously ill. He may be witness to an ongoing abusive relationship at home. He may be excited and distracted because of an upcoming holiday, a birthday or a sleepover.

- **Cultural Differences :** Children who come from a different culture may have a problem understanding the content of questions or situations appearing in the tests. For example. Jewish or Muslim children might have trouble understanding questions centering on Christmas or Easter.

- **Economic Differences :** Children coming from homes where money is scarce will be disadvantaged in may ways. As well as lacking proper nourishment, they may never have been to a farm, a zoo or a circus. Any reference to these or other sites familiar to most of the students, will puzzle them and negatively affect their scores.

- **Parental Attitudes :** Children to whom parents have read since infancy, whose homes are scattered with books and papers, and whose parents enjoy reading themselves, have a huge advantage over those whose homes which contain no reading materia. Children whose parents are supportive and interested in their children's education will do better than the children of those who show little interest in their progress at school.

- **Classroom Situation :** If the classroom is crowded, if it contains several children with learning or behavioral disabilities, with no aid to help, if the teacher has changed one or more times during the year, the students will not score as well on a standardized test. Learning must be incremental, discipline must be consistent, and those having difficulty should have a little one-on-one attention from the teacher.

There are many factors which can influence the scores of standardized tests. Competent teachers realize this, and they also know their children and their strengths and weaknesses well. They will take all factors into account before deciding on a report card mark. or a placement for the following term. The standardized test is only one tool in teacher's stockpile of evaluation.

16.4 Environmental Factors

Neurobiological theory tells us that the development of a humans various forms of intelligence usually ends by the age of 16, when a criticial point is reaches. After this point of intellectual maturity, IQ is thought to be relatively stable. The environment has its biggest affect in determining someone's fluid intelligence up and till the critical point. Fluid Intelligence is someone's ability to think logically in verbal and spatial terms and detect patterns too.

For a person to develop certain intellectual abilities, they need to be provided with the appropriate environmental stimuli during childhood, before the critical period for adapting their neuronal connections ends. It should be mentioned that some researchers believe that the critical period effect is a result of the manner by which intellectual abilities are acquired—that changes in neuronal connections inhibit or prevent possible future changes which may explain differences in Intelligence types between people of different cultures.

Environmental Factors that affect test include : modern media, education, breast feeding, womb conditions, butrition, pollution, nurture and parenting, prejudices and self belief, national culture, head injuries, sleep problems, drug and alcohol abuse, mental illnesses, stress and diseases.

Modern Media

Modern Media forms including TV, film, computer games, electronic gadgets and devices, the Internet and computer software are giving people a cognitive work out each time they participate in these forms of modern media. Computer games, software and gadgets require great analytical and logical thought and probing in order to be used to their entirety.

Television shows are becoming more complex with up to 18 threads happening at once often spreading over multiple episodes are re-appearing at different times, compared to just 1 50 years ago. The viewer is required to keep track of many threads at once. It is also becoming harder to make sense of what is actually happening on the screen which as peoples analytical skills and pattern recognition skills are being pushed.

Education

Increasing the time in education and quality of education before the critical limit and to a lesser extent afterwards, has been shown to improve a person's performance on an intelligence test.

Womb Conditions

High intakes of alchohol and drug use in mothers have been shown to significantly lower the IQ of their children. Extent of prenatal exposure to marijuana affects development of intelligence later in childhood.

Prolonged period of time without access to oxygen during the delivery can lead to brain damage and mental retardation. Also, low birth weights have been linked to lower intelligence scores later in lives of the children.

Did u know? In a comprehensive study involving 11 studies and ouver 7000 children, children who were breast fed were found to do 5.2 points better on IQ tests than children who were not breastfed.

Nutrition

Poor nutrition, characterized by zinc, iron, vitamin B and protein deficiencies, leads to low IQ, which leads to later antisocial behavior.

Pollution

Lead exposure has been proven to have significant effects on the develpment of a child's intelligence. In a long-term study done by Baghurst et al. 1992, children who grew up next to a lead-smelting plant had significantly lower intelligence test scores, negatively correlated with their blood-lead level exposure.

Task How Parental attitudes affect the child performance is test ?

Self Assessment

1. *Fill in the blanks :*

   (*i*) Marks awarded by the teachers on a given test are test scores.

        (*ii*) Difficulty level of test affects on test scores.

(*iii*) Children whose parents are suppostive and interested in their children's education will do better than the children of those who show little interest in their interest in their progres at school.

(*iv*) Under anxiety, children can conconcantrate on their studies and exams.

(*v*) Children who come from a different culture may have not any kind of problem understanding the content of questions.

(*vi*) High intake of alcohol and drug use in matters have been shows to significantly lower the IQ of their children.

(*vii*) Lead exposure increase the intelligence test score.

16.5 Summary

- A standardized test score is an indication of how much knowledge a child has accumulated and can put down on paper on a certain day within a limited time frame. That is all it is : an indication. There are many additional factors, some of which are more important, that a teacher must consider before assigning a mark for a report card, or deciding on the appropriate placement for a student for the upcoming term.

- In fact the students' test score on an examination depends on many factors of course one of these is student's knoweldge, understanding or ability in the subject-matter of examination, which determines the student's final mark. This factor influence the student's own Score (not of others) but there are other factors that influence score of all students, more or less equally. These are given below :

  (*i*) Difficulty level of Examination: It is not possible to judge the exact difficulty level of an examination question at the time of its Construction.

  (*ii*) Nature of group tested

  (*iii*) Instructional effectiveness

  (*iv*) Miscalculation of time limit

  (*v*) Sampling inadequacies

- Psychological Factors

- Anxiety : Some children "freeze" in a testing situation. They are unable to focus or concentrate long enough to give a true picture of their actual knowledge in a subject area.

- Physical Health : A child may be coming down with the flu, have a headache, a sore throat, or be overtired.

- Emotional Health : The student may be upset because parents are in the midst of separation or divorce proceedings. A close relative may be seriously ill. He may be witness to an ongoing abusive relationship at home.

- Cultural Differences : Children who come from a different culture may have a problem understanding the content of questions or situations appearing in the tests.

- Economic Differences : Children coming from homes where money is scarce will be disadvantaged in may ways. As well as lacking proper nourishment, they may never have been to a farm, a zoo or a circus.

- Parental Attitudes : Children to whom parents have read since infancy, whose homes are scattered with books and papers, and whose parents enjoy reading themselves, have a huge advantage over those whose homes which contain no reading material.

- Classroom Situation : If the classroom is crowded, if it contains several children with learning

or behavioral disabilities, with no aid to help, if the teacher has changed one or more times during the year, the students will not score as well on a standardized test.

- Neurobiological theory tells us that the development of a humans various forms of intelligence usually ends by the age of 16, when a criticial point is reaches. Afrer this point of intellectual maturity, IQ is thought to be relatively stable. The environment has its biggest affect in determining someone's fluid intelligence up and till the critical point.

- Environmental Factors that affect test include : modern media, education, breast feeding, womb conditions, butrition, pollution, nurture and parenting, prejudices and self belief, national culture, head injuries, sleep problems, drug and alcohol abuse, mental illnesses, stress and diseases.

16.6 Keywords

Psychological : Connected with a person's mind

Environmental : Connected with the conditions that affect the behaviour and development of somebody/something

Neurobiological : relating to never orto thescience of neurology

16.7 Review Questions

1. What do you mean by standardized test scores ?

2. What are the psychological factors affect the test scores ?

3. Explain the nature of test influence the scores.

4. Discuss various environmental factors which affect test scores.

Answers : Self Assement

1. (*i*) T   (*ii*) T   (*iii*) T   (*iv*) F   (*v*) F

   (*vi*) T   (*vii*) F

16.8 Further Readings

Books

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 17 : Integrated Approach of Evaluation

CONTENTS

Objectives

Introduction

17.1 Integrated Approach to Evaluation

17.2 Integrating Rational, Formative and Summative Evaluation

17.3 Integrated Evaluation at Different Stages

17.4 Summary

17.5 Keywords

17.6 Review Questions

17.7 Further Readings

Objectives

The objectives of this unit can be summarized as below :

*   To discuss about the Integrated approach to Evaluation

*   To explain about the integrating rational, formative and summative evaluation

Introduction

It is well known that the quality of curriculum determines the standards of education laid in the form of curriculum objectives, as enunciated in N.C.F. for various stages of school education. Evaluation theorists and practising evaluators emphasise three key components of evaluation. viz. : (a) the process of gathering information, (b) Forming judgments about the worth of the phenomena, and (c) using judgments for taking educational decisions. Therefore, for evaluating a curriculum, a programme or the learners, we have to use these three steps. Since pupils' evaluation is dependent on the curriculum intentions or restrictions, both pupils' evaluation and curriculum evaluation are closely related. We will discuss about Evaluation of Education

17.1. Integrated Approach to Evaluation

 Pupil evaluation as also curriculum evaluation are dependent on programmes undertaken to transact the curriculum, which in turn have impact on pupils' achievement vis-a-vis pupil evaluation. On the other hand pupils' learning vis-a-vis their evaluation is also conditioned by learning milieu of the school, which provides the needed school plant facilities, services, personnel, money, material and social environment in which pupils learn. This curriculum transactions, programme effects and pupils' learning are all commensurate with the institutional or school effectiveness. Therefore emerging demand on today's evaluators is to integrate pupils' evaluation with curriculum evaluation, programme evaluation and institutional self-evaluation. Each of these is described below, indicating the trend in evaluation.

17.1.1 Curriculum Evaluation

Four major components of the curriculum are : (a) curriculum objectives, (b) curriculum content, (c) curriculum process, methodology of teaching and learning, and (d) outcomes of learning. Each of these components has to be evaluated. What is worth appreciating is how pupils' evaluation is related to these components.

Curriculum Objectives

Evaluation of objectives here refers to finding out :

(*i*)   to what extent the instructional objectives are relevant to learners' needs, i.e. pupils' evaluation should be undertaken in terms of competencies intended to be developed;

(*ii*) to what extent these objectives are specified in terms of competencies to be developed in pupils; and

(*iii*) to what extent the performance standards are reflected in pupil' evaluation to validate those objectives, i.e. stated intended competencies ?

Curriculum Content

Evaluation of content or syllabus refers to :

(*i*)   to what extent the content is relevant in attainment of instructional objectives;

(*ii*) to what extent the content selected is accurate, adequate and up-to-date; and

(*iii*) to what extent it reflects the substantive structure of the subject at a given maturity level of pupils ?

Curriculum Processes or Methodology of Teachings and Learning

Focus of evaluation is on :

(*i*)   reflecting the mode of inquiry (syntactical structure) of the discipline;

(*ii*) involving the students maximally, in instructional process (questioning etc.); and

(*iii*) using readiness and diagnostic unit testing at various instructional phases to provide and get regular feedback about students' learning.

Curriculum Outcomes

Here evaluation of curriculum outcomes refers to :

(*i*)   pupils' performance in terms of instructional objectives;

(*ii*) efficacy of instructional effectiveness or instructional impact; and

(*iii*) collateral effects and concomitant learning.

We can observe from the needed emphasis the integral relationship of pupils' evaluation and curriculum evaluation in terms of all the four curriculum components.

17.1.2 Programme Evaluation

Literature and research on programme evaluation in India are quite scanty. There , are hardly any evidences about programmes that are evaluated for judging programme effectiveness. Most of the programmes that are conducted are judged on logical, basis or on the reflective opinion of subject experts. It is usually taken for granted that a programme like a seminar on evaluation, development of curriculum, writing of a text-book, training of paper-setters etc. when conducted, is good and successful. It is taken for granted. At the most, at the end of the programme a few open-ended questions are given for participants to reflect their opinion. In fact what is required is to evaluate every programme in its context, the inputs given, the processes undertaken, outputs accrued and the impact it made. For good evidences and forming judgments about a programme, the following components need to be addressed before taking decisions about the programme effectiveness.

Context Evaluation

This refers to :

(*i*)   social setting or context ofthe programme in terms of purpose and scope;

(*ii*) antecedents of the programme that necessitated it, i.e. previous experiences; and

(*iii*) existing resources and support services available etc.

Input Evaluation

This refers to :

(*i*)   financial inputs given in the programme;

(*ii*)   personnel involved, their qualifications and expertise;

(*iii*) material and other support services provided; and

(*iv*)   time spent on various tasks or activities.

Process Evaluation

This refers to :

(*i*)   efforts put in by teachers in terms of materials, teaching testing etc. ;

(*ii*)   efficiency with which the tasks were undertaken; and

(*iii*) effectiveness of processes undertaken, in terms of involvement of persons and expertise availed of.

Product or Outcome Evaluation

This refers to :

(*i*)   attainment of programme objectives, students' performance or results etc. ;

(*ii*)   quality or materials produced or reports etc.; and

(*iii*) growth and development of staff, participants and other clientele.

Impact Evaluation

This refers to programme effect in terms of :

(*i*)   community support, parents' concern, students' involvement etc.;

(*ii*)   improved instructional and evaluation practices; and

(*iii*) better learning environments in school.

Institutional Self-Evaluation

Institutional self-evaluation (I.S.E.) is a planned and organised inquiry carried out by the staff, students and principals of the institution into the total effectiveness of institutional operations to find out the adequacies and inadequacies on the basis of which steps can be taken to improve teaching, learning and assessment practices. In a way it is also school-based evaluation, usually called 'continuous comprehensive evaluation' (C.C.E.) in Indian context. However, it cannot be equated with C.C.E., which refers to the concept, model, assessment procedure, tools of measurement, marking, grading and reporting of subject-wise results to pupils. Whereas focus of C.C.E. is on pupils' assessment and their improvement, focus of I.S.E. is on school effectiveness in terms of assessing the institution (not pupils alone) as a whole.

Therefore I.S.E. aims at determining the efforts, efficiency and effectiveness of the institution. It covers pupils' attainments, teachers' instructional impact, learning milieu of school, utilisation of resources and all such factors and conditions that contribute to or hinder various school processes. Focus is on judging school effectiveness and taking suitable measures to improve the existing conditions.

It involves comprehensive and systematic evaluation effort carried out by the staff of the institution. It aims primarily at identifying the problems and producing positive and desirable changes in the institution. Thus it fits into the framework of formative-summative model of Scriven (1972). Its aims are : (a) to achieve an understanding of the institution in terms of strengths and weaknesses relative to aims and social responsibility of the institution; and (b) to serve for the vitalising and

updating goals and operations to improve both performance and quality of institutional outputs. Such an institutional self-study can be undertaken annually or at an internal of every 2-3 years to get at the issues, problems, possibilities, solutions and review of the policies.

The following are essential characteristics of institutional self-evaluation :

Pluralistic Approach to Institutional Analysis and Review

For this the following steps are necessary :

(*i*) We have to identify institutional clientele (students, teachers, administrators, parents etc.) who are interested and ready to participate in study and analysis of problems.

(*ii*) Participants have to identify specific institutional problems and suggest solutions.

Using Psychometric and Descriptive Approaches for Data Collection and Processing

For this the advantage can be taken of :

(*i*) open discussion among the participant observers;

(*ii*) emotional interaction of parents, community and supervisors;

(*iii*) extensive data collection, analysis and interpretation; and

(*iv*) evaluation and provision of research support.

Using Humanistic Illuminative Evaluation

This involves views, opinions and concerns of all stake-holders of institutional community like:

(*i*) principal, staff, students and auxiliary staff;

(*ii*) alumni, employers and colleges where the students seek admissions;

(*iii*) managing committee, communit representatives and P.T.A. representative; and

(*iv*) curriculum and evaluation experts for providing technical knowhow (not for judgmenting).

Using Reports and Documentary Evidences

This refers to analysis of :

(*i*) inspectors' or supervisors' reports of visitation during the years;

(*ii*) inter-institutional cooperative testing results and boards results;

(*iii*) governing bodies or managing committee reports; and

(*iv*) student-intake data, i.e. background of students studying in the school.

Institutional self-study is an in-house evaluation, which is formative in nature and should not pose any threat to staff of the institution. Its suggestion a recommendations are more likely to be accepted and implemented. There is need for seminars and workshops execute institutional self-studies on regular basis which would make institutions more aware of their usefulness.

Did u know? If evaluation becomes an integral part of all projects, it would go a long way to ensure attainment of objectives and improve educational practices.

17.2 Integrating Rational, Formative and Summative Evaluation

Whether it is pupil, curriculum or programme evaluation it is important, to use three phases of evaluation at different stages of pupil evaluation, curriculum evaluation and programme evaluation.

Pupils' Evaluation

(*i*)   At the planning stage it is rational or logical evaluation which helps plan, devise or structure a programme, a scheme, a curriculum or any other project. In pupils' evaluation we devise evaluation tool to measure their achievements on different objectives on the basis of logical thinking, documented reports, experience gained earlier and consensus of experts.

(*ii*)   During development or trialling of the test material (formative stage), formative evaluation planned or unplanned goes on, providing the needed facts on the material and the difficulties faced, leading to adaptation of test material side by side to suit the test situation.

(*iii*) It is only after empirical evidence through actual use in unit tests, term tests or annual examination that we really know the quality of the test or test items. For that item analysis is undertaken to find out the facility and discrimination indices on the basis of which summative judgments are made. Thus constructing items on logical basis (rational), reviewing, improving and adapting them to fit various objectives (formative) and then using them in testing to get empirical evidences about their quality in terms of F.I., D.I. etc. (summative) reflect integration of the three modes of evaluation.

Curriculum Evaluation

(*i*)   On the basis of rational evaluation, the curriculum objectives, curriculum context, methodology of curriculum transactions and intended outcomes are delineated by curriculum developers.

(*ii*)   During formative evaluation, when developmental process and process of implementation are going on, certain changes, deletions, adaptations, improvement etc. are made to suit the situation, thereby adjusting the means and ends.

(*iii*) Summative evaluation is undertaken when the curriculum is implemented for a year or two. It helps to validate all the four components of curriculum. Accordingly, during renewal of the curriculum, the needed changes are made to make curriculum objectives more realistic, curriculum content more relevant, methodology of curriculum transaction more appropriate and the intended outcomes of curriculum more tangible and measurable.

Programme Evaluation

(*i*)   On the basis of rational evaluation, programme is planned and developed with respect to the context in which it is being planned, the inputs visualised, processes envisaged, product or output expected and the likely effects on learning environment and social milieu of the school.

(*ii*)   Formative evaluation goes on during the implementation stage, to get and provide feedback on the efficiency and effectiveness of various components of the programme, leading to en-route changes, adaptation, elimination and improvements.

(*iii*) Summative evaluation of a programme is undertaken when it is over. It helps throw light on congruence or discrepancy between the intended and the observed with respect to all the components. Accordingly, context evaluation, input evaluation, process evaluation, product evaluation and impact evaluation have to be undertaken.

Notes   Evaluation is no longer limited to cognitive learning outcomes but assessment of holistic development of learners, teaching-learning strategies and school environment as a whole.

17.3 Integrated Evaluation at Different Stages

1. At early childhood or pre-primary stage, there need not be any formal evaluation and every teaching-learning activity should be without reflecting any judgment.

2. At primary stage evaluation will be formative, continual and comprehensive, using observation and orals for classes I and II as a part of teaching-learning activities. In classes III to V, besides orals and observation techniques, use of Criterion Referenced (C.R.) tests for assessing competencies for mastery, evaluation of co-scholastic attributes through observation and rating scale to be reported once in 3 months on 3-point absolute grading on scholastic and 3-point direct grading on co-scholastic attributes.

3. At upper primary stage, apart from oral and written tests, the use of assignments and project work will be made as also the C.R. tests. Observation, rating scales should be reported quarterly. Self-evaluation by students and peer evaluation along with use of 5-point absolute grading on scholastic and 3-point direct grading on co-scholastic aspects should be followed.

4. At secondary stage 9-point scale absolute grading on scholastic and 5-point direct grading on scholastic aspects, use of cumulative record card, self-appraisal and peer appraisal will continue.

5. At higher secondary stage courses may be organised in four semester and based on credit system, first three by school and 4th semester examination by the board. Introduction of tutorial, continuation of C.R. testing, 9-point scale using absolute grading and grading by converting marks directly into grades should be   enforced. Boards may use 9-point grading on the curve for classifying students. Direct grading on assessment of co-scholastic aspects to be recorded semester-wise on 5-points and providing opportunity for students to improve their grades.

6. For maintaining standards, achievement surveys are recommended, using standardised tests. The results of these surveys will be used for developing institutional, regional, state and national profiles as also for planning and designing appropriate interventions for improving the standard.

7. Reference is made to the establishment of National Evaluation Organisation (N.E.O.) to evolve national standards and common standards of achievement and bring about uniformity of standards among various boards. This would also offset the now prevailing multiplicity of entrance tests conducted by different institutions. Suggestion is also made to have N.E.O. for general academic courses to project true educational profile of each state at macro level and of individual institutions at the micro level. N.P.E. and P.O.A. have visualised this N.E.O. as a quality-control mechanism to organise nation-wide tests on voluntary basis to enable norms for comparability of performance and conducting independent tests, besides developmental activities and research studies in the areas having bearing on students' evaluations.

Task   What is input evaluation ?

Self Assessment

1. *Fill in the blanks :*

   (*i*) The school based evalution is also called _____ .

   (*ii*) _____ helps of Validate all the components of curriculum.

(*iii*) A secondary stage _____ absolute grading on scholastic and _____ direct grading.

(*iv*) At the _____ it is _____ evaluation which helps plan, devise or structure a programme, a scheme, a curriculum or any other project.

17.4 Summary

- Quality of curriculum determines the standards of education laid in the form of curriculum objectives, as enunciated in N.C.F. W. for various stages of school education. Evaluation theorists and practising evaluators emphasise three key components of evaluation. viz. : (a) the process of gathering information, (b) Forming judgments about the worth of the phenomena, and (c) using judgments for taking educational decisions.

- Curriculum Evaluation

- Four major components of the curriculum are : (a) curriculum objectives, (b) curriculum content, (c) curriculum process, methodology of teaching and learning, and (d) outcomes of learning.

- Curriculum Objectives

- Evaluation of objectives here refers to finding out :

 (*i*) to what extent the instructional objectives are relevant to learners' needs, i.e. pupils' evaluation should be undertaken in terms of competencies intended to be developed;

- Curriculum Content

- Evaluation of content or syllabus refers to :

 (*i*) to what extent the content is relevant in attainment of instructional objectives;

 (*ii*) to what extent the content selected is accurate, adequate and up-to-date; and

- Curriculum Processes or Methodology of Teachins and Learning

- Focus of evaluation is on :

 (*i*) reflecting the mode of inquiry (syntactical structure) of the discipline;

 (*ii*) involving the students maximally, in instructional process (questioning etc.); and

- Curriculum Outcomes

- Programme Evaluation

- Literature and research on programme evaluation in India are quite scanty. There , are hardly any evidences about programmes that are evaluated for judging programme effectiveness. Most of the programmes that are conducted are judged on logical, basis or on the reflective opinion of subject experts. It is usually taken for granted that a programme like a seminar on evaluation, development of curriculum, writing of a text-book, training of paper-setters etc.

- Context Evaluation

- This refers to :

 (*i*) social setting or context ofthe programme in terms of purpose and scope;

 (*ii*) antecedents of the programme that necessitated it, i.e. previous experiences; and

- Input Evaluation

- This refers to :

 (*i*) financial inputs given in the programme;

 (*ii*) personnel involved, their qualifications and expertise;

- Process Evaluation

- This refers to :

  (*i*) efforts put in by teachers in terms of materials, teaching testing etc. ;

  (*ii*) efficiency with which the tasks were undertaken; and

- Product or Outcome Evaluation

- This refers to :

  (i) attainment of programme objectives, students' performance or results etc. ;

  (ii) quality or materials produced or reports etc.; and

- Impact Evaluation

- This refers to programme effect in terms of :

  (*i*) community support, parents' concern, students' involvement etc.;

- Institutional self-evaluation (I.S.E.) is a planned and organised inquiry carried out by the staff, students and principals of the institution into the total effectiveness of institutional operations to find out the adequacies and inadequacies on the basis of which steps can be taken to improve teaching, learning and assessment practices.

- Pluralistic Approach to Institutional Analysis and Review

- For this the following steps are necessary :

  (*i*) We have to identify institutional clientele (students, teachers, administrators, parents etc.) who are interested and ready to participate in study and analysis of problems.

- Integrating Rational, Formative and Summative Evaluation

- Pupils' Evaluation

  (*i*) At the planning stage it is rational or logical evaluation which helps plan, devise or structure a programme, a scheme, a curriculum or any other project. In pupils' evaluation we devise evaluation tool to measure their achievements on different objectives on the basis of logical thinking, documented reports, experience gained earlier and consensus of experts.

  (*ii*) During development or trialling of the test material (formative stage), formative evaluation planned or unplanned goes on, providing the needed facts on the material and the difficulties faced, leading to adaptation of test material side by side to suit the test situation.

- Curriculum Evaluation

  (*i*) On the basis of rational evaluation, the curriculum objectives, curriculum context, methodology of curriculum transactions and intended outcomes are delineated by curriculum developers.

  (*ii*) During formative evaluation, when developmental process and process of implementation are going on, certain changes, deletions, adaptations, improvement etc. are made to suit the situation, thereby adjusting the means and ends.

  (*iii*) On the basis of rational evaluation, programme is planned and developed with respect to the context in which it is being planned, the inputs visualised, processes envisaged, product or output expected and the likely effects on learning environment and social milieu of the school.

- Integrated Evaluation at Different Stages

  (*i*) At early childhood or pre-primary stage, there need not be any formal evaluation and every teaching-learning activity should be without reflecting any judgment.

  (*ii*) At primary stage evaluation will be formative, continual and comprehensive, using observation and orals for classes I and II as a part of teaching-learning activities. In classes

Notes

III to V, besides orals and observation techniques, use of Criterion Referenced (C.R.) tests for assessing competencies for mastery, evaluation of co-scholastic attributes through observation and rating scale to be reported once in 3 months on 3-point absolute grading on scholastic and 3-point direct grading on co-scholastic attributes.

(*iii*) At upper primary stage, apart from oral and written tests, the use of assignments and project work will be made as also the C.R. tests. Observation, rating scales should be reported quarterly.

(*iv*) At secondary stage 9-point scale absolute grading on scholastic and 5-point direct grading on scholastic aspects, use of cumulative record card, self-appraisal and peer appraisal will continue.

(*v*) At higher secondary stage courses may be organised in four semester and based on credit system.

17.5 Keywords

Integrated : The act or process of combining two or more things that they work together

Approach : A way of dealing with somebody /something

17.6 Review Questions

1. What are different types of evaluation under integrated approach to evaluation.

2. What is pupil's evaluation ?

3. Explain the term curriculum evaluation ?

4. Discuss the integrated evaluation at different stages.

Answers : Self Assessment

1. (*i*) Continuous and Comprehensive evaluation (CCE)

(*ii*) Summative evaluation

(*iii*) 9 point scale, 5 point scale

(*iv*) Planning stage, logical

17.7 Further Readings

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 18 : Marking System : Need, Problems, Components

CONTENTS

Objectives

Introduction

18.1 Meaning of Marking System

18.2 Need of Marking System

18.3 Technical Questions in Marking

18.4 Derived Scores

18.5 Components of Marking System

18.6 Techniques of Marking

18.7 Advantages of Marking System

18.8 Problems of Marking System

18.9 Summary

18.10 Keywords

18.11 Review Questions

18.12 Further Readings

Objectives

The objectives of this question can be summarised as below :

- To explain the Meaning of Marking
- To discuss about the Meaning of Marking System
- To describe about the Technical Questions in Marking
- To explain about Derived Scores
- To discuss about the Components of Marking System
- To explain about Techniques of Marking
- To describe about the Advantages of Marking System
- To explain about the Problems of Marking System

Introduction

Marks and marking systems are deeply embedded in the educational culture (marks are performance indicators). They have become the basis of a number of actions and decisions within a given educational institution, as also in the whole system of education, for promotion, selection, employment, teacher's evaluation etc.

18.1 Meaning of Marking System

What the marks should mean, need to be clearly defined, developed, adopted and enforced

through a system of marking in an institution. A marking system is basically a system of communication. If teachers do not assign ranks in conformity with the policies reflecting the system, the marks tend to loose their significance and marking system fails to serve its needed function. The degree to which the marks have the same meaning for all those who use them serves the purposes of communication meaningfully. Marking system of an institution is of concern to teachers of other schools, departments and other institutions, and hence the meaning of each mark must be clearly defined.

Two major types of marking system are in vogue. The numerical marks in terms of percentage varying from zero to 100 marks. Since a student's mark is independent of other students' marks, the system to percentage marking is called 'Absolute marking'. Second type of marking system is based on the use of a small number of letter marks, often five letter A, B, C, D and F, to express various levels of performance. We will discuss only Marking system.

18.2 Need of Making System

18.2.1 Needs of students

(*i*)   To know their present performance and rate of progress.

(*ii*) To know their adequacies and inadequacies in learning.

(*iii*) To compare their performance with their class-mates.

(*iv*) To know their potentials and limitations in various subject areas.

18.2.2 Needs of parents

(i)   To know whether their ward is progressing satisfactorily in the class.

(*ii*) To alert them if some problems are developing.

(*iii*) To help them plan for the future of their child.

(*iv*) To know how to use marks on standardised test scores for decision making for further education of the child.

18.2.3 Needs of teachers

(*i*)   To know in general the performance of students in terms of syllabus tested.

(*ii*) To identify the areas of weaknesses for remedial teaching.

(*iii*) To judge the effectiveness of their instruction on the basis of students' performance.

(*iv*) To find out if certain topics or concepts in the prescribed curriculum are beyond the students' level.

18.2.4 Needs of administrators or principals

(*i*)   To compare inter-school performance of students.

(*ii*) To judge whether tests from different schools compare well in terms of instructional objectives tested.

(*iii*) To ensure that the intended learning outcomes are being tested properly.

(*iv*) To maintain inter subject-intra school parity of performance standard.

18.2.5 Needs of curriculum specialists

(*i*)   To find out if certain curriculum objectives need to be reviewed.

(*ii*) To identify whether curriculum content (syllabus) of certain subjects needs to be revised.

(*iii*) To know whether instructional effectiveness is satisfactory for achieving instructional objectives.

(*iv*) To judge whether the tests used were reasonably valid and reliable.

They provide the basis on which inferences and conclusions are drawn about students' abilities, parents' decisions, teachers' instructional impact, administrators' views of performance standards, and curriculum specialists; concern about relevance of curriculum and quality of measuring instruments.

Notes   Thus we find that the marks or scores we award to students reflect the outcomes with which the students, parents, teachers, administrators and curriculum specialists are all interested.

18.3 Technical Questions in Marking

Some technical problems relating to assignments of marks are given below :

1. What should a mark be based on ?

   What characteristics of a student be represented in a mark ?

   (*i*) Should it be appraisal of competence in a curriculum portion?

   (*ii*) Should it be modified by other factors like amount of work completed; mechanical aspects of work completed such as neatness, legibility, oral expression etc.; and scholastic aptitude indicated ?

   (*iii*) Should it be judged by common standard or in relation to the student's potential to achieve ?

   (*iv*) What type of tests, exercises or behaviours would provide evidence of achievement that the mark is to represent, e.g. specified objectives or expected outcomes of learning.

2. How data or evidence on different components be weighted?

   Achievement indicators of evidences on tests, orals, exercise, projects, assignments etc. are reported on students' performance. How weight for each component be decided ? Some of these measures could be :

   (*i*) Validity of the information is what an indicator provides (knowledge, skills, competency) and the faithfulness with which it is reflected.

   (*ii*) Reliability of the data or evidence, i.e. the more the errors of measurement the less the weight, e.g. the highest reliability in objective tests, less in essay type tests and lowest in orals.

   (*iii*) Effective weight of each component is determined by its standard deviation. For example, to determine the weight of term paper vis-a-vis written examination we note the S.D. of each as also the mean. If the teacher wants to give double the weightage to the test than the term paper, we have to see that the effect of test be twice that of the term paper in combined score, as illustrated below.

   Test   Mean = 60   S.D. = 10

   Term paper   Mean = 3.0   S.D. = 0.5 (on a 5-point scale)

   Here if marks are combined as such, test will have 20 times effect on paper in determining the individual's standing in the combined score. Since teacher wants only twice as great, we

multiply score on term test by 10, making the mean 30 (3.0 × 10) and standard deviation 5.0 (0.5 × 10). Now the two S.D. are in the ratio 2:1 [60 and 30 mean] and [10 and 5 S.D.] for the test and the term paper.

3. How many categories should be used

Reporting can be done in terms of 2 (pass-fail or mastery and non-mastery), 3 (A, B, F/C) or 4 (A, B, C, F/D) or 5 (A, B, C, D, F/E), 7, 9, ... or 100 or 101 categories. As the number of categories increase, the number of borderline decisions increases correspondingly, but each decision becomes less crucial for total academic record of students. There is a trade off of increased frequency of potential error or unfairness in grading for decreased size and importance of that error. The number of categories for grading is related to reliability of evidence on which decisions are taken. If standard error of measurement is large in a component that enters into a final grade or grade itself relative to the unit, discriminations become meaningless (Table 1).

Table 1 : Reliability categories at 4 levels

Grading categories	0.60	0.70	0.80	0.90
15 (A+toF-)	27%	31%	37%	50%
5 (A to F)	70%	77%	85%	96%
3 (Honours-Pass-Fail)	91%	95%	98%	99.9%

It is evident that making discrimination into more than 15 is futile.

4. How many should get what marks (A-B-C-D-E) ?

 (i) Symbols that represent marks are basically ordinal, not cardinal systems,

 (ii) A is better than B, B better than C and so on down to F.

 (iii) Lack of universal consensus on that each step represents equal step of quality, i.e. A is as better than B as C is better than D or D is better than F.

 (iv) Symbols are embedded in an educational culture, and therefore cultural role of symbols is as important as is their psychometric properties. Decision to decide the grade for advance courses is socio-cultural or psychometric.

 (v) Successive symbols should represent equal steps along an interval scale.

 (*vi*) Competence is normally distributed in accordance with normal curve. Range of normal curve can then be divided into equal segments. For example, in case of 5-point scale A to F percentages of cases are given in Table 2.

Table 2 : Percentages of cases in 5-point scale of normal curve

Symbol	Standard score range on curve	Percentage of cases
A	+1.5 to +2.5 or higher	7
B	+0.5 to+1.5	24
C	–0.5 to +0.5	38
D	–1.5 to-0.5	24
F	–2.5 or below to–1.5	7

5. To what frame of reference should marks be assigned ?
Marks can be assigned with reference to the following frames of references :

   (a) **Marks in relation to perfection :** This relates to mastery of the concepts to be learnt, as we have in criterion-referenced testing, where specified or intended learning

outcomes are clearly listed in term of competencies implied in various instructional objectives. However, its utility is limited to a small fraction of school learning.

Moreover, the task is to be specified to such an extent that it leads to trial outcomes that become unwieldy.

(b) **Marks in terms of students 'potential to perform :** It raises the issue of whether what one can do, does he really do ? How best can we measure the potential that is not contaminated by the present achievement ? Can we really make the needed judgment of achievement in relation to potential, segregating it from simple achievement ? If so, how valid and reliable such a judgment can be made by teachers. Moreover, how marks can then be used in different contexts, e.g. for admissions to higher courses, employments, selections etc., if assigned in terms of potential of the performer. What is important is that what John can in fact do or not, whether he does the best he can.

(c) **Performance in relation to peer groups :** It means that an individual's performance can be judged in relation to some reference group. This is the type of referent that can be used in relation to a class, school, district, state or the national group (as used in standardised test) norm. This is indeed norm-referenced judgment. Problem here is the nature of the group in relation to which a student's performance is judged, whether random or stratified. Government schools, central schools, public schools etc all make a difference in judging performance of a student, with reference to a brighter, poorer or similar group performance and the size of the group.

(d) **How standards to be equated :** How best to anchor any specific group to that frame of reference ? For example, common syllabus and course objectives taught by teachers of all sections of a school or teachers of different schools following the same course provide reasonable anchoring device. In groups that are more diverse, one may think of the previous grade-point average or scholastic aptitude test as a rough anchor (Thorndike). If common test is given to three sections and frequency is worked out, we can calculate the number getting As, Bs, Cs and so on, depending upon the policy of grading in an institution or an examining agency, e.g. 15%As, 25%Bs, 40%Cs, 15%Ds, and 5%Fs etc. Sometimes a student getting B grade may finally get A grade if his previous grade-point average is taken into account by the institution or another indirect indicator like S.A.T. is considered besides the anchoring test. Percentage of each grade awarded corresponds, step by step, with the percentage in that category on the common test. This may lead to more than 15% As in a section and less than 25% Bs or less than 15% As in another section with more than 25% Bs etc.

18.4 Derived Scores

Test scores can be expressed in a variety of different units in relation to different scales of measurement. A score may represent very good performance on a test, but the same score on another test may represent very poor score. Therefore some common basis has to be established to make comparison of scores on those widely different tests possible. This may reflect relation of derived scores to norms. These are of three types.

(a) **Derived Scores Based on Median Performance**

These are the grade scores or grade equivalents and the age scores Both depend on norm tables with raw scores that grade equivalent or age scores can be determined.

(*i*) **Grade-equivalent scores :** Indicate the position on a grade scale at which a students' test performance places him. A child scores on a test that is equal to the median or average score of pupils, of say 5th grade students. His grade equivalent on the subject-matter of the test is 4.0 regardless of whether he is in class V or III. Sometimes these

are called G scores. Grades and months are commonly listed as 4.3 or 5.2, i.e. 3 months into grade 4 or 2 months into grade 5.

(*ii*) **Age equivalents :** Indicate the position on an age scale at a point on which a pupil's performance places him. For example, a child with 4.3 grade equivalent may be found on a table of norms equivalent to 8.5 grade equivalent which means equivalent to median score of 8 years and 5 months olds on the same test. He may otherwise be about a year younger or older in his assignment on the subject-matter of the test. Age equivalents are represented by Educational Quotient (E.Q.) for achievement, say in Reading. Likewise Mental Age (M. A.) is considered on general intelligence and Reading Age (R.A.) for achievement in Reading. E.A. of 9-8 indicates the same level as average children of 9 years and 8 months.

(*iii*) **Quotients as derived scores :** Such derived scores show relationship between two characteristics of the child or indicate the manner in which growth of various types is related, e.g. Intelligence Quotient (I.Q.), Educational Quotient (E.Q.) and Reading Age (R.A.). Suppose a pupil has Chronological Age (C.A.) of 8.4, Educational Age (E.A.) of 9-2, a Mental Age (M. A.) of 9-7 and Reading Age (R.A.) of 9-4. For such a child.

$$\text{His E.Q.} = 100\,\frac{EA}{CA} = 100\,\frac{9-2}{8-4} = 100\,\frac{110\,(\text{months})}{100\,(\text{months})} = 110$$

$$\text{His I.Q.} = 100\,\frac{MA}{CA} = 100\,\frac{9-7}{8-4} = 100\,\frac{115\,(\text{months})}{100\,(\text{months})} = 115$$

$$\text{His R.Q.} = 100\,\frac{RA}{CA} = 100\,\frac{9-4}{8-4} = 100\,\frac{112\,(\text{months})}{100\,(\text{months})} = 112$$

(b) Derived scores Based on Variability of Performance

These are of two types :

(*i*) Percentile ranks, and

(*ii*) Scores that express position on a scale in units of standard deviation or quartile deviation.

(c) Derived scores Based on Standard Deviation

These may be of 4 types :

(*i*) **Z scores :** lt is a simple method of showing deviation of a score from arithmetic mean of distribution. Its formula is:

$$Z = \frac{X - M}{S.D}$$

where X = raw score; M = Arithmetic mean; and S.D. = standard deviation of the raw scores.

(*ii*) **T score :** It is similar to Z score except that it eliminates the use of negative values and decimals. A.M. is taken as 50 and S.D. as 10 to calculate the T score, using the formula :

$$T = \frac{10\,(X - M)}{S.D} + 50$$

where, X, M and S.D represent raw score, arithmatic mean and the standard deviation respectively.

(*iii*) **Standard scores, scaled scores , equated scores and converted scores :** These are derived scores that help compare the scores on different parts of the same test or even on different tests. These are based on mean and standard deviations for some standard group.

(*iv*) **Informal derived scores :** Relative ranks and letter marks on a test (A, B, C, D, E, F) are other types of derived scores.

18.5 Components of Marking System

Numerical basis for assigning marks should include different aspects like home task, project, test scores and class-room contribution. Weightage to each component may be worked out on the basis of Mean and S.D. of component score, for getting the combined score. However, precise weighthing of components on numerical basis is not crucial to the quality of scores assigned.

18.6 Techniques of Marking

Scaling and Equating

Scaling is a technique that standardises raw scores or marks from one scale to another. Two raw scores may be described equivalent and assigned the same degree of excellence in relation to some relevant group. Equivalence means that when raw scores are normally distributed, the two definitions produce identical results (Harper). When two distributions of raw scores are not normal, the calibrated scores will be different for these two approaches. Two raw scores are defined as equivalent and all therefore are translated into the same scaled score if they are at the same distance from the means of their distribution, in terms ot standard deviations of their

distribution, in terms of standard deviations of their distribution, i.e. $\frac{x-m}{\text{S.D.}}$ (x = score, m = mean, and S.D. = standard deviation). Two raw scores are considered equivalent and are therefore translated into the same scale score if they are both exceeded oy the same proportion of examinees. For example, if 10% of the students receive a mark of 55 or higher on an examination P, and a mark of 63 or higher on examination Q, these two raw marks (55 and 63) should be awarded the same scaled marks or grade. Calibration of any marks will control enough of the factors to make the scaled marks or grades much more valid than the raw marks. Since in many cases we deal with population not a sample (any students who sat for examination), scores need not necessarily be distributed according to normal curve. The following methods can take care of non-normally distributed groups.

Use of Linear or Normalised Method

Linear-scale transformation produces a set of scores whose shape distribution is identical with that of the raw scores, whereas normalised scale transformations force a non-normal distribution of raw scores into a normal distribution of scaled scores.

Which of the two should be used ? Draw a graph of the raw scores awarded by several examiners and smoothen the curves. If distribution is approximately normal, any of the two methods as far convenience can be used. But if they are skewed or very irregular, find out why they are not normal. If it is because of sampling, use the linear scale method. If it is due to examiner who skewed the marks or due to the peculiarity of the test that has produced non-normal distribution, use the normalised scaling. Situation second is more common than the first one. Advantage of normalising is that it makes the scaled scores strictly comparable at all levels, though linearly-scaled scores may not have equivalent distribution, the linearly scaled marks supposedly cannot legitimately be added or compared directly.

Equating

Sometimes when we suspect that a particular group is not really a representative group, we may prefer to have a population (not sample), e.g. "all students of this years Biology class that I have taught are to be judged on the same standard as those of last year". This is called equating because the final grades or marks are based on this year's statistics and on past statistics also.

(a) Statistical equating : In case of objective examinations, a mini test of 20 items or 20% whichever is greater (ETS) are included from a previously used test, for the purpose of calibration. These items may be put together or scattered throughout. Compare the performance of students or of these items with those of the previous year to know whether the group is better or worse than before. Calibration is done accordingly. All of GRE Aptitude Tests Scaled Scores are statistically equated with the standard of the 1952 test (Harper, 1990) and therefore it is possible to compare the abilities of the 2003 batch with any previous group of the with past 50 years. These items of 1952 may now be out of date, but every year items are scaled to 1952 items and thus are ultimately connected statistically with the 1952 scale.

(b) Equating grades : Some subjects are tougher than others. To adjust grades that remain standard across the subjects that are widely different, the following method (adapted from Ebel, 1972) is recommended (Harper, 1990) :

(*i*) Some common tests like entrance examination, aggregate marks at previous level, or it can be a special 'scaling' or a Link Test (e.g. Aptitude test) have to be used.

(*ii*) For all entering examinees or students on the entrance test or scaling test, frequency distribution should be made.

(*iii*) Translate these marks into grades using linear or normalised method.

(*iv*) Make a separate table for each group that is to be graded separately (class, section, course or subjects). Table would indicate the number of students in that group who had each grade on the entrance test 5 As, 15 Bs, etc.

(*v*) This table acts as a guide for that group. Teachers are expected to assign approximately the same number of As, Bs, Cs etc. at the end of the course as these were common on the entrance test. Teacher may, however, use his judgment to assign higher or lower grade depending on brighter with lazy group (compared to grades on entrance test), but he will have to justify wide deviations when questioned by academic committee.

Did u know? ICSCE has been using grades since long. CBSE tried to introduce grades but had revert back to marking in the wake of public opinion and resistance from university system. With the new Curriculum Framework of NCERT it is now the time to switch over again. There is need for creating awareness about usefulness of grading systems, among students, teachers, parents and public at large.

18.7 Advantages of Marking System

There are other systems of marking having different purposes to serve, with their own advantages and limitations, as expressed below.

(*i*) Multiple marks on various aspects of achievement can improve marking but cost more in extra efforts and may not be worth the improvement expected.

(*ii*) The more the marks available in the system indicating different levels of achievement, the more reliable the marks will be, but the usability of this system is limited.

(*iii*) System of pass-fail marks does not serve the functions that marks are supposed to do, especially in finely graded marks.

(*iv*) Numerical marks lay emphasis as measurement and simplify calculations of grade-point averages.

(*v*) For quality control of marking system, it necessary publicise the distribution of marks subject by subject.

(*vi*) Alternative to strict marking on the curve is relative marking, which divides the score scale into equal intervals. It may require calculation of some average measure like the median and measures of variability like standard deviation.

18.8 Problems of Marking System

At present we assess students on their learning for three different purposes: (a) with reference to his own progress, i.e. to know his progress (self-referenced); (b) (With reference to his peer group (norm-referenced); and (c) with reference to criteria set up in terms of Intended Learning Outcomes (criterion-referenced). In all these cases purpose may be different but the students are always awarded numerical marks on an interval scale of 0-100 marks. However, artificially created zero represents 'nothingness' and 100 represents 'perfection' in achievement. Both are fallacious, since zero does not mean that the student does not know anything, and 100 does not mean that he knows everything about the subject. This makes comparison of scores difficult from test to test as well as from subject to subject.

Moreover, many types of measurement errors creep in due to intra-examiner's and inter-examiner's subjectivity in marking, inadequacy in sampling of content and competencies implied in instructional objectives, poor quality of evaluation tools (question papers), arbitrary time limit and variations in testing situations. Some of these errors are identifiable but others are not. The magnitude of such errors according to research reports varies from 7 to 15%. Can our 101 (0-100) point numerical scale serve to discriminate individuals on the basis of a single score, i.e. 50, 51, 52 etc. If standard error of measurement is 10, then true score of 50 is likely to be between 40 and 60 in 2 out of 3 cases. It is therefore difficult to measure precisely an individual on 101 point scale, when measurement errors are known to exist from 7 to 15%. This limitation can be overcome if we use ability bands that represent different ranges of scores. These ability bands vary according to the number of categories one wishes to use for classification.

Task   What is Scaling ?

Self Assessment

(*i*) Marks are the _____ .

(*ii*) The system to percentage marking is called _____ .

(*iii*) _____ use marking to compare inter school performance of students.

(*iv*) _____ indicates the position on a grade scale at which a students test performance.

(*v*) _____ lay emphasis as measurement and simply Calculations of _____ .

18.9 Summary

- A marking system is basically a system of communication. If teachers do not assign ranks in conformity with the policies reflecting the system, the marks tend to loose their significance and marking system fails to serve its needed function. The degree to which the marks have the same meaning for all those who use them serves the purposes of communication meaningfully. Marking system of an institution is of concern to teachers of other schools, departments and other institutions, and hence the meaning of each mark must be clearly defined.

- The numerical marks in terms of percentage varying from zero to 100 marks. Since a student's mark is independent of other students' marks, the system to percentage marking is called

'Absolute marking'. Second type of marking system is based on the use of a small number of letter marks, often five letter A, B, C, D and F, to express various levels of performance.

- Needs of students

  (*i*) To know their present performance and rate of progress.

  (*ii*) To know their adequacies and inadequacies in learning.

- Needs of parents

  (*i*) To know whether their ward is progressing satisfactorily in the class.

  (*ii*) To alert them if some problems are developing.

- Needs of teachers

  (*i*) To know in general the performance of students in terms of syllabus tested.

  (*ii*) To identify the areas of weaknesses for remedial teaching.

- Needs of administrators or principals

  (*i*) To compare inter-school performance of students.

  (*ii*) To judge whether tests from different schools compare well in terms of instructional objectives tested.

- Needs of curriculum specialists

  (*i*) To find out if certain curriculum objectives need to be reviewed.

  (*ii*) To identify whether curriculum content (syllabus) of certain subjects needs to be revised.

- Some technical problems relating to assignments of marks are given below :

  (*i*) What should a mark be based on ?

  (*ii*) How data or evidence on different components be weighted ?

  (*iii*) How many categories should be used

  (*iv*) How many should get what marks (A-B-C-D-E) ?

  (*v*) To what frame of reference should marks be assigned ?

      (a) Marks in relation to perfection

      (b) Marks in terms of students 'potential to perform

      (c) Performance in relation to peer groups

      (d) How standards to be equated

- Test scores can be expressed in a variety of different units in relation to different scales of measurement. A score may represent very good performance on a test, but the same score on another test may represent very poor score. Therefore some common basis has to be established to make comparison of scores on those widely different tests possible. This may reflect relation of derived scores to norms. These are of three types.

(a) Derived Scores Based on Median Performance

  - These are the grade scores or grade equivalents and the age scores Both depend on norm tables with raw scores that grade equivalent or age scores can be determined.

  (*i*) Grade-equivalent scores : Indicate the position on a grade scale at which a students' test performance places him.

  (*ii*) Age equivalents : Indicate the position on an age scale at a point on which a pupil's performance places him.

(b)   Derived scores Based on Variability of Performance   **Notes**

These are of two types :

(*i*) Percentile ranks, and

(*ii*) Scores that express position on a scale in units of standard deviation or quartile deviation.

(c)   Derived scores Based on Standard Deviation

These may be of 4 types :

(*i*) Z scores

(*ii*) T score

(*iii*) Standard scores, scaled , equated scores and converted scores

(*iv*) Informal derived scores

*   Calibration, Scaling and Equating

*   Scaling is a technique that standardises raw scores or marks from one scale to another. Two raw scores may be described equivalent and assigned the same degree of excellence in relation to some relevant group.

*   Two raw scores are defined as equivalent and all therefore are translated into the same scaled score if they are at the same distance from the means of their distribution, in terms ot standard deviations of their distribution, in terms of standard deviations of their distribution, i.e. $\dfrac{x-m}{S.D.}$ (x = score, m = mean, and S.D. = standard deviation).

*   Use of Linear or Normalised Method

*   Linear-scale transformation produces a set of scores whose shape distribution is identical with that of the raw scores, whereas normalised scale transformations force a non-normal distribution of raw scores into a normal distribution of scaled scores.

*   Sometimes when we suspect that a particular group is not really a representative group, we may prefer to have a population (not sample), e.g.

*   This is called equating because the final grades or marks are based on this year's statistics and on past statistics also.

(*a*) Statistical equating

(*b*) Equating grades

(*i*)   Some common tests like entrance examination, aggregate marks at previous level or it can be a special 'scaling' or a Link Test (e.g. Aptitude test) have to be used.

(*ii*)   For all entering examinees or students on the entrance test or scaling test, frequency distribution should be made.

*   There are other systems of marking having different purposes to serve, with their own advantages and limitations, as expressed below.

(*i*) Multiple marks on various aspects of achievement can improve marking but cost more in extra efforts and may not be worth the improvement expected.

(*ii*) The more the marks available in the system indicating different levels of achievement, the more reliable the marks will be, but the usability of this system is limited.

(*iii*) System of pass-fail marks does not serve the functions that marks are supposed to do, especially in finely graded marks.

(*iv*) Numerical marks lay emphasis as measurement and simplify calculations of grade-point averages.

- At present we assess students on their learning for three different purposes: (a) with reference to his own progress, i.e. to know his progress (self-referenced); (b) (With reference to his peer group (norm-referenced); and (c) with reference to criteria set up in terms of Intended Learning Outcomes (criterion-referenced).

- Moreover, many types of measurement errors creep in due to intra-examiner's and inter-examiner's subjectivity in marking.

18.10 Keywords

Marking : The activity of checking and correcting the written work.

Calibration : The units of measurement marked on athermometer or other instrument.

18.11 Review Questions

1. What is the Meaning of Marking System ?
2. Why is the need of Marking System Especially in Schools ?
3. What are Derived Scores ?
4. What are the Problems of Marking System ?

Answers : Self Assessment

1. (*i*)   Performance indicators   (*ii*)   Absolute marking
   (*iii*) Administrator   (*iv*)   Grade equivalent
   (*v*)   Numerical marks, grade point average

18.12 Further Readings

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*
2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*
3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*
4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 19 : Grading - Need, Problems, Components and Methods

CONTENTS

Objectives

Introduction

19.1 Concept of Grading

19.2 Need of Grading

19.3 Components of Grading

19.4 Methods of Assigning Grades

19.5 Implications for Teaching, Learning and Testing

19.6 Problems of Grading

19.7 Summary

19.8 Keywords

19.9 Review Questions

19.10 Further Readings

Objectives

The objectives of this unit can be summarised as below :

- To explain the concept of grading
- To discuss about the need of grading
- To describe about the components of grading
- To explain about the methods of assigning grades
- To describe about the implications for teaching, learning and testing
- To discuss about the problems of grading.

Introduction

The Indian education system has taken a step forward towards reviving the education system with the introduction of grading system in session 2009-10. It will help in reducing the pressure on students during exams. In the last five years the meaning of education has changed for students from imbibing knowledge to merely scoring marks, resulting in myriad forms of education policies.

As per reports, every day more than 17 students aged between 15-25 years commit suicide in India due to non-performance in the examination or an entrance test. Watching young children of the country succumbing to the undue pressure of scoring high marks is horrifying. One of the points to note here is the thinking of the society, which puts lot of pressure on students to 'to perform'.

This pressure from schools, parents, peer groups and society takes away the youthfulness of a child. Further, a health report also supports that this often causes health hazard such as fatigue, body aches, eye weakness, stress and in more severe cases, depression (neurotic/psychotic.)

Looking at today's education scenario, the Central Board of Senior Education has introduced educational counselors and child psychologists in schools to boost the confidence of young students and mentally prepare them for the board examination. This method has helped in reducing the stress and making them comfortable with the examination.

Adoption of grading system in India, the grading system was introduced in 2008-09 from class I-VIII, reducing the exam stress. Extending the concept to class IX and X has further reduced the pressure, giving students an opportunity to explore other avenues. Following the US model, the implementation of the grading system is to bring in more practical education than the current theoretical method. This model prescribes a varied range of opportunities, providing children of all levels a platform to showcase their talent and pursue their interests traditionally; class XI students were given subjects as per marks scored in Class X. This system often disappointed students if they scored low. Moreover, if a student didn't get the required percentage due to poor scoring in one subject, then the entire percentage gets affected. The grading system will give students relief. It will provide ample opportunities to students to excel in their choice filed.

19.1 Concept of Grading

Grading is a method of communicating students' achievement using a set of symbols supposed to be properly defined and understood clearly by students, teachers, parents and all other concerned people. Meaning of each symbol of grades must be clearly defined. The examiner has to adhere to the system of grading that is specified. A proper system of grading provides for better comparison of students' performance as well as quality of performance in terms of efforts and knowledge acquired at the end of the prescribed syllabus.

Basis of awarding grades differs from examiner to examiner and school to school. It is difficult to compare grades awarded in different institutions.

19.2 Need of Grading

A number of purposes are served by grading, as listed below :

(a) It provides an account of the level of instructional objectives achieved by students, which is useful to both teachers and parents.

(b) Grades provides a sort of permanent record of students growth, which is useful to institutes of higher learning for admissions, besides to the prospective employees.

(c) It serves the purpose of making decisions by school it self in terms of promotions and placement.

(d) Grades may be helpful in reviewing and adapting instructional strategies and appropriateness of the prescribed curriculum.

(e) Grades can also be used in determining Grade point average (G.P.A) for awarding scholarships in a number of institutions.

Notes Understanding the board exam system in India and its relation with students is of great importance in present times. While coping with the expectations of school, parents and society and keeping pace with their talent, students face a lot of hardships. Thus, the implementation of a   grading system and abolition of board exams is really a boon for students.

19.3 Components of Grading

The grade are to be based on homework, a case study, and fixed exams.

Homework

The only way to learn mathematical material is by working problems.

*   Homework is due at the beginning of class on the due date (see schedule). No late homework will be accepted.

*   You are encouraged to work exercises that are not assigned. Some can be found in the course notes. More can be found in the textbook by Winston, *Operations Research* : *Applications and Algorithms*, Duxbury Press (1994). Copies are on reserve in the Hunt library.

*   You should do the homework without consulting other students. If you need help, you may consult the instructor or a teaching assistant. For *nonassigned* exercises, it is a good idea to work them out with other students.

*   Every attempt will be made to return graded homeworks promptly. Solutions to the homework will be distributed in class after the due date.

*   The following rules apply to all homework :

    *   Use only 11 by 8.3 in. paper.

    *   Start each problem on a new page.

    *   Assemble the homework in proper order and staple it together,

    *   PRINT your name on the top right hand corner.

Case Study

The case study has two parts. The first part will require formulating a linear program. The second part requires solving the linear program and analyzing the results for managerial purposes.

Exams

*   There will be one midterm and one final exam. The precise dates of the exams will be announced by the GSIA administration.

*   The questions in the exams will be similar to the homework problems and the problems discussed in class both in form and content. The lectures may occasionally range over peripheral topics, but the content of the exercises is a reliable guide to which topics can appear on the exam.

*   All exams will be **open-notes.**

*   Except for unforeseen circumstances, makeup exams will not be given. A student is required to inform the instructor as soon as possible in such a case.

Self Assessment

1. *Fill in the blanks :*

(*i*) _____ is the method of communicating students achievement using a set of symbols, clearly understood by students, teachers, parents.

(*ii*) Grading provides a sort of permanent _____ of _____.

(*iii*) Grades can be used in determining _____ for awarding scholarship in a number of institutions.

19.4 Methods of Assigning Grades

There are different ways depending upon the reference point used. When the reference point is

the approach, grading may be Direct grading and Indirect grading. When the reference point is the standard of judgment, grading may be termed Absolute grading and Relative grading.

Direct Grading

When the performance of students is assessed in qualitative terms and the impression the examiner or the teacher makes is expressed directly in terms of letter grades, it is Direct grading. This method of direct grading can be used for assessment of both cognitive and non-cognitive learning outcomes. However, it is preferred for assessment of non-cognitive outcomes of learning. Non-cognitive outcomes can be listed stage-wise, assessed and reported separately in terms of letter grades. Depending on the nature and quality of the attribute assessed, 3-point or 5-point scale may be used for grading. This direct grading helps minimise inter-examiner variability besides its easiness to use compared with other methods. However, direct grading lacks diagnostic role, discourages competition and is devoid of transparency.

Indirect Grading

In this case students' performance is assessed in terms of marks, which are then transformed into letter grades, using different methods.

Absolute Grading

It is based on pre-determined standard, which becomes the reference point for assessing students' performance. This involve direct conversion of marks into grades, categorising students into different groups like first, second and third divisions as exemplified below.

(*i*)   Distinction : 75% and above

(*ii*)   First division : 60% to 74%

(*iii*)   Second division : 45% to 59%

(*iv*)   Third division : 33% to 44%

(*v*)   Unsatisfactory (fail) : Below 33%

In spite of same range of categories for different subjects, grades may not be comparable because marks are not comparable and distribution of marks varies from subject to subject and from year to year in a given subject. Thus a number of students placed in different categories would differ from subject to subject and year to year, making them uncomparable.

However, in criterion-referenced tests each student is assessed individually in terms of intended learning outcomes, usually stated in terms of competencies implied in instructional objectives.

An example of absolute grading is given in Table 1.

Table 1 : Absolute grading (9-point scale)

Q.No.	Letter grade	Marks Range	Description
1	A	90% and above	Outstanding
2	B	80% to 89	Excellent
3	C	70% to 79	Very good
4	D	60% to 69	Good
5	E	50% to 59	Above average
6	F	40% to 49	Average
7	G	30% to 39	Below average
8	H	20% to 29	Marginal
9	I	Below 20%	Unsatisfactory

This method involves simple procedure, and the meaning of each grade is understood. Criterion being the focal point; the student discover their strengths and weaknesses and each student can strive for attaining the highest grade. One limitation of this method is that scores are taken as such without considering the errors of measurement that creep in due to various subjective factors. Another limitation is that the cut-off scores for various categories are decided arbitrarily.

Relative Grading

This type of grading is called grading on the curve, which refers to normal distribution curve. In this case it is pre-determined what percentage of students are expected to receive the number of As, Bs, Cs, Ds and so on with reference to a particular group. That is why relative grading is also called 'norm-referenced grading'. Steps involved in grading are as follows.

(a) Pupils are expected to acquire new competencies, which become the basis for testing of sampled behaviour outcomes.

(b) Tests developed provide evidences on students' learning or achievement, which lead to award of marks.

(c) These marks when plotted on a graph leading to formation of distribution curve (obtained), which may be positively or negatively skewed. This obtained curve is converted into normal curve on the premise that there is a difference between the true level of achievement and the perceived level of achievement as yielded by tests. True level of achievement is expected to be normally distributed in a large population irrespective of curricular area.

(d) When perceived curve is transformed into a normal curve, the obtained scores can be classified into desired number of categories of grades. This is done by using mean of 5 and standard deviation of 2, and grade value will range from 1 to 9 categories. Two-tail categories at either end of distribution are combined to make 9 (Stanine scale) in place of 11 categories. In this way we will have grade-wise distribution of scores (Table 2).

Table 2 : Grade-wise distribution of cases (9-point scale)

Q.No.	Letter grade	Interval	No. of cases (%)	Grade value
1	A	1.75 σ to	4	9
2	B	1.25 to 1.75	7	8
3	C	0.75 to 1.25	12	7
4	D	0.25 to 0.75	17	6
5	E	– 0.25 to 0.25	20	5
6	F	– 0.75 to – 0.25	17	4
7	G	– 1.25 to – 0.75	12	3
8	H	– 1.75 to – 1.25	7	2
9	I	– to – 1.75	4	1

If 7-point grading is to be used, measurement scale may be divided into 7 equal parts. When 5-point grading is to be used, measurement scale may be divided into 5 equal parts and grade values will range from 1 to 5 with a mean of 3 and S.D. of approximately 1 unit. Grade-wise distribution of cases will be as shown in Table 3.

Table 3 : Grade-wise distribution of cases (5-point scale)

Q. No.	Letter grade	Interval	No. of cases (%)	Grade value
1	A	1.5 σ to	7	5
2	B	0.5 to 1.5	24	4
3	C	– 0.5 to 0.5	38	3
4	D	– 1.5 to – 0.5	24	2
5	E	– to – 1.5	7	1

Advantages

Grading on the curve has the following advantages :

(*i*)   Pass-fail concept is eliminated; there is no grade indicating failure.

(*ii*)   Grade of a student indicates his position relative to his group.

(*iii*)   Grades provide comparability across the curricular area, as normal distribution ensures uniformity in spread of scores.

(*iv*)   Grades are usable for recording growth of individual students.

(*v*)   Grades provide for meaningful additivity to arrive at G.P.A.

(*vi*)   Test difficulty will not affect distribution of grades.

Problems

(*i*)   One limitation of 'grading on the curve is that percentages of cases receiving different grades is pre-decided and the individual's grade is decided by the performance of the group, not by his performance alone.

(*ii*)   This system fails to reward improved teaching and learning. An extra ordinary group taught by good teacher leading to high achievement scores of all or majority of students will also get the same percentage of cases as A, B, C, D, E grades and not all getting As or As and Bs, because pre-conceived number of cases are to be placed in each group.

Comparison of Overall Performance of Students (G.P.A.)

The method of relative grading allows comparison of students' performance within as well as across the subjects. Difficulty arises when we have to compare overall performance (covering all the subjects) of students. For that we can compute the G.P.A. when the method of relative grading is used for assessment in various subjects (Table 4).

Table 4 : Grading On a 5-Point Scale (A-E) or (5-1)

Students	Grades awarded in different subjects					G.P.A.
	English	Biology	Chemistry	Physics	Maths	
P	A	B	A	B	A	4.6
R	B	B	C	A	A	4.2
S	C	B	A	B	B	4.0

(*i*)   For P the grade values are 5 + 4 + 5 + 4 + 5 = 23   5 = 4.6

(*ii*)   For R the grade values are 4 + 4 + 3 + 5 + 5 = 21   5 = 4.2

(*iii*)   For S the grade values are 3 + 4 + 5 + 4 + 4 = 20   5 = 4.0

Thus overall performance of student P is 4.6, that of R is 4.2 and of S is 4.0, showing that P performed better than R and S, and R performed better than S. Likewise the students' performance can be compared through G.P.A., whether it is on a grading of 5,7 or 9 points.7

Notes

Alternative Grading System

Here it is dichotomous two-point scale, which is with reference to development of Competency or non-acquisition of competency. Criteria for developing competency is pre-determined. If criteria are fulfilled, the student passes or else fails. Here the advantage is that student may work towards those criteria without pressure. Students and teachers can work together to develop the criteria for pass grade.

However, its limitation is that the information about competence of learner is not available to other teachers, because the learner might have scaled the criteria with the lowest acceptable level.

Moreover, this method does not provide information about strengths and weaknesses of students' learning.

Grading the Non-Cognitive Outcomes of Learning

The present system of evaluation emphasises mainly the scholastic aspects of pupils' growth, whereas the co-scholastic aspects are relegated to the background. This results in ignoring non-cognitive outcomes of learning, which are equally important for all-round development of pupils. These are neither tested regularly nor in a comprehensive manner. There is difficulty of assessing non-cognitive outcome because it is more time consuming and has less usable appraisal techniques and tools of assessing these learning outcomes besides availability of well-developed and validated criteria of assessment of non-cognitive outcomes of learning. This, however, does not deter us from improving the existing practices with more relevant criteria, tools and modes of assessment of those outcomes. What is needed to institutionalise assessment of these outcomes is to :

(a)   get clarity of the Intended Learning Outcomes;

(b)   develop simple usable tools of assessment;

(c)   decide about the mode of assessment to be used; and

(d)   develop profiles of students for recording results each term.

This would facilitate introduction and appraisal of non-cognitive outcomes of learning and help in identifying deficiencies that form the basis for remediation.

Steps Involved in Assessment

(*i*)   For proper assessment of these outcomes, first we have to list the non-cognitive learning outcomes stage-wise in terms of competencies or skills to be developed.

(*ii*)   Assess the various outcomes using direct grading.

(*iii*)   Report every term separately the results on intended outcomes listed.

(*iv*)   Responsibility of assessing non-cognitive outcomes rests with the individual institution even during the year of public examinations.

(*v*)   Results prepared may be conveyed to the boards for showing them in the certtificate to be issued along with the scholastic learning outcomes.

Pre-requisites for Proper Grading (Do's and Don'ts)

Before awarding grades it is expected that :

(a)   well-developed design and blueprints of balanced question papers are used ;

(b)   papers are set and moderated by experienced teachers ;

(c) detailed marking scheme is developed to minimise subjective marking.

(d) examination is conducted fairly without any malpractice;

(e) marking scheme is reviewed after marking a sample of answer scripts;

(f) Scripts are allotted to the examiner after randomisation of scripts to ensure parity of scripts for all examiners; and

(g) head examiner undertakes sample checking of scripts to ensure fair marking.

If all the above-mentioned measures are taken, it will lead to awarding of grades ensuring high reliability and validity.

19.5 Implications for Teaching , Learning and Testing

(a) In spite of various imperfections in numerical marking, the system goes on and everyone— the students, teachers and public at large—have developed a mind set that every one has accepted despite all types of shortcomings in the existing system. To come out of this mind set, all concerned schools, teachers, students, parents and the examining agencies have to work together to replace numerical scores by letter grading.

(b) When grading is introduced, misclassification of students based on unreliable marks will be minimised because similar students will get same grades and it would eliminate unhealthy competition among high achievers and the low achievers will be relieved of fear of failure in examination.

(c) Chance of improvement in grades will help reduce ill effects of poor results and avoid human-resource wastage. (50% stagnation)

(d) Teachers have to act as facilitator, and become conversant with the philosophy and mechanics of grading system. For that it is necessary to have pre-service and in-service training of teachers in grading procedures.

(e) Compatibility between grades of different schools in school-based assessment and external examinations both for curricular and non-cognitive areas will have to be ensured by developing common design, criteria and assessment procedures.

(f) Schools will have to evolve their own criteria on the basis of G.P.A. Some guidelines for maintaining uniformity in admissions by Boards of Secondary Education will have to be worked out.

(g) Boards will have to adopt a uniform grading system to help inter-board comparison of students' performance in terms of grades. This would also help students inter-board mobility to migrate from one board to the other.

(h) Results declared in terms of grades in place of marks will reduce pressure on students to compete, as parents would not pressurise their wards as in today's system. Situation will be further eased when flexibility in reviewing borderline cases is provided.

(i) For admission to institutes of higher learning either selection tests will have to be used as we have now in professional courses; or alternatively the selections can be made on the basis of relative grades in different subjects, for which G.P.A. can be usable indicator for selection of students for a specific course of studies.

If grading system is to be introduced in school system, its acceptability in institutes of higher learning has to be accepted. And for that the U.G.C. has to be persuaded to devise some workable strategy to remove bottlenecks for introducing grades in the university system.

19.6 Problems of Grading

The problem with traditional grading is that students have good reasons to worry about their grades, and yet being grade-oriented undermines the most important goals of liberal arts education. Students have good reasons to worry about their grades because of the powerful symbolic and social roles, that grades play in students' lives. Most undergraduate students are at an identity-forming stage of their lives, and so they are looking within and outside of themselves for clues about who they are and what they should do with their lives. It seems obvious to students to look to their grades in order to read what the world is telling them their strengths and weaknesses are. This way of thinking is often explicitly reinforced by parents, professors, and prospective employers.

Grades also have acquired increasingly powerful social force. Grades are not at all private communication between teachers and students but have a quasi-public role in students' lives. The

occasions in which students are asked to reveal their grades are frequently some of the most significant moments in students' lives and connect with some of students' most important relationships. There is so much that can rest on grades—parental approval; scholarships; students' being allowed to continue studying; their being allowed to participate in other meaningful, perhaps identity-forming activities such as athletic participation or study abroad; and their future opportunities such as eligibility for jobs or graduate school.

Grades are essentially numerical and thus can only be appropriately applied to what is measurable, but not everything that is measurable is always measured in a course of study. A Student too oriented towards getting good grades can miss or neglect those components of the course that are not graded. Furthermore, what is measurable is not always what is most important in liberal arts education.

There are many *qualitative* ideals underlying the purpose of liberal arts education that cannot be measured on a comparative scale of quantifiable achievement. In fact, some of these qualitative goals cannot be definitively judged by a teacher—students themselves are in a better position to evaluate these dimensions of their learning.

Research in cognitive science and developmental psychology reveals that human learning is extraordinarily complex. While professors can look in on some aspects of the learning process and judge whether students are putting certain words, numbers, or symbols together correctly— even reading past the words and numbers to more general conclusions about students' conceptual development—there is much about students' learning that remains invisible even to the most attentive professors. Students themselves are in a better position to judge many of the qualitative dimensions of their learning, as well as some quantitative dimensions, such as their sense of improvement, the intensity of their effort and engagement, whether they did all of the reading, how well they paid attention in class, and how significant their learning was for them. But traditional grading can *discourage* the development and refinement of students' abilities in these respects, because strong self-motivation and keen self-awareness of one's own learning can bring a student into conflict with professors' judgments. Ultimately, such conflicts are resolvable through thoughtful, mutually respectful dialogue, but our society and our educational system do not teach students how to work through such difficulties, and so the easiest psychological tactic is for students to suppress their self-motivation and subvert their intellectual self-awareness to the authority of their teachers. But developing self-awareness and developing self-motivation are exactly some of the qualitative ideals underlying the purpose of a liberal arts education.

When teachers strive explicitly to structure their classes in ways that foster the development of self-motivation and push students to engage authentically with their education, they can feel that their ideals are consistently undermined by their students' efforts to play it safe and try to please the teacher. Authentic engagement with the educational process is inherently frightening and difficult, exposing the student to a world larger and stranger than previously imagined, demanding

that the student reconstruct a sense of identity in order to find her or his place in this newly expanded world. In addition, authentic engagement with education demands that students push themselves to their very limits—a humbling enterprise requiring great personal strength. The only way that teachers can reasonably ask this of students is if the class can become a highly respectful and supportive environment: a context of mutual trust and commitment. Challenges and criticism offered in a context of trust can be perceived and accepted as exciting calls to growth. When students trust their teachers and their environment, they can open themselves to the (often difficult) personal transformation that authentic education inspires.

If, on the other hand, students feel that they cannot trust their teachers or their environment, they become guarded and try to play it safe. Their reaction to challenges or criticism becomes defensive. If they feel themselves to be in a hostile context, a context of distrust demanding that they prove their worthiness at every turn, then defensively protecting their ego becomes an important survival strategy. The fear-based learning that happens in a context of distrust may look effective at first, because fear evokes keen attentiveness, but such learning remains at the surface until the pressure is off, at which point much of what was learned promptly dissipates. The legacy of learning by fear is that many such students get better at manipulating the systems they become cynical about, and become better at appearing the ways others want them to appear. Such learning, however, does not help students to develop their subtle perceptual powers, or their depth of vision; it is not built upon respect, and it does not strengthen their character or expand their compassion.

Because so much outside of the classroom hinges on students' grades, traditional grading imposes an outermost context of distrust framing everything that then happens within the classroom. Even when individual teachers attempt to create a trusting and supportive environment for the students within the classroom, the knowledge that grades will be recorded on students' permanent, quasi-public transcripts at the end of the semester permeates students' consciousness at nearly every moment. Despite the teacher's good intentions, the grade in this class may have important ramifications in several different aspects of the student's life. Therefore, students feel embedded in a context of distrust, where their ability to maintain good relations with their parents, stay in school, keep their scholarships.

Task Task What is G.P.A ? Explain.

Self Assessment

2. Multiple Choice Questions :

Choose the correct option :

(*i*) _____ is based on predetermined standard.

    (a) Absolute grading   (b) Relative grading

    (c) Direct grading   (d) Alternative grading

(*ii*) _____ type of grading is called grading on the curve, which refers to normal distribution curve.

    (a) Direct grading   (b) Indirect grading

    (c) Relative grading   (d) Absolute grading

(*iii*) _____ is dichotomous two point scale, which is with reference to development of competency or non-acquisition of Competency.

(a) Direct grading   (b) Alternative grading

(c) Indirect grading   (d) Absolute grading

(*iv*) _____ is also called norm referenced grading.

(a) Direct grading   (b) Indirect grading

(c) Absolute grading   (d) Relative grading

19.7 Summary

- Adoption of grading system in India, the grading system was introduced in 2008-09 from class I-VIII, reducing the exam stress. Extending the concept to class IX and X has further reduced the pressure, giving students an opportunity to explore other avenues. Following the US model, the implementation of the grading system is to bring in more practical education than the current theoretical method. This model prescribes a varied range of opportunities, providing children of all levels a platform to showcase their talent and pursue their interests traditionally; class XI students were given subjects as per marks scored in Class X. This system often disappointed students if they scored low.

- Grading is a method of communicating students' achievement using a set of symbols supposed to be properly defined and understood clearly by students, teachers, parents and all other concerned people. Meaning of each symbol of grades must be clearly defined.

- Need of Grading

- A number of purposes are served by grading, as listed below :

  (a) It provides an account of the level of instructional objectives achieved by students, which is useful to both teachers and parents.

  (b) Grades provides a sort of permanent record of students growth, which is useful to institutes of higher learning for admissions, besides to the prospective employees.

  (c) It serves the purpose of making decisions by school it self in terms of promotions and placement.

- There are different ways depending upon the reference point used. When the reference point is the approach, grading may be Direct grading and Indirect grading. When the reference point is the standard of judgment, grading may be termed Absolute grading and Relative grading.

- Direct Grading

- When the performance of students is assessed in qualitative terms and the impression the examiner or the teacher makes is expressed directly in terms of letter grades, it is Direct grading. This method of direct grading can be used for assessment of both cognitive and non-cognitive learning outcomes. However, it is preferred for assessment of non-cognitive outcomes of learning.

- Indirect Grading

- In this case students' performance is assessed in terms of marks, which are then transformed into letter grades, using different methods.

- It is based on pre-determined standard, which becomes the reference point for assessing students' performance. This involve direct conversion of marks into grades, categorising students into different groups like first, second and third divisions as exemplified below.

- This method involves simple procedure, and the meaning of each grade is understood. Criterion being the focal point; the student discover their strengths and weaknesses and each student can strive for attaining the highest grade.

Notes
- Relative Grading
- This type of grading is called grading on the curve, which refers to normal distribution curve. In this case it is pre-determined what percentage of students are expected to receive the number of As, Bs, Cs, Ds and so on with reference to a particular group.
- Grading on the curve has the following advantages :
  (*i*) Pass-fail concept is eliminated; there is no grade indicating failure.
  (*ii*) Grade of a student indicates his position relative to his group.
  (*iii*) Grades provide comparability across the curricular area, as normal distribution ensures uniformity in spread of scores.
- The method of relative grading allows comparison of students' performance within as well as across the subjects. Difficulty arises when we have to compare overall performance (covering all the subjects) of students.
- Alternative Grading System
- Here it is dichotomous two-point scale, which is with reference to development of Competency or non-acquisition of competency. Criteria for developing competency is pre-determined. If criteria are fulfilled, the student passes or else fails. Here the advantage is that student may work towards those criteria without pressure. Students and teachers can work together to develop the criteria for pass grade.
- Grading the Non-Cognitive Outcomes of Learning
- The present system of evaluation emphasises mainly the scholastic aspects of pupils' growth, whereas the co-scholastic aspects are relegated to the background.
- Steps Involved in Assessment
  (*i*) For proper assessment of these outcomes, first we have to list the non-cognitive learning outcomes stage-wise in terms of competencies or skills to be developed.
  (*ii*) Assess the various outcomes using direct grading.
  (*iii*) Report every term separately the results on intended outcomes listed.
- Before awarding grades it is expected that :
  (a) well-developed design and blueprints of balanced question papers are used ;
  (b) papers are set and moderated by experienced teachers ;
- Implications for Teaching , Learning and Testing
  (a) In spite of various imperfections in numerical marking, the system goes on and everyone—the students, teachers and public at large—have developed a mind set that every one has accepted despite all types of shortcomings in the existing system.
  (b) When grading is introduced, misclassification of students based on unreliable marks will be minimised because similar students will get same grades and it would eliminate unhealthy competition among high achievers and the low achievers will be relieved of fear of failure in examination.
  (c) Chance of improvement in grades will help reduce ill effects of poor results and avoid human-resource wastage. (50% stagnation)
  (d) Teachers have to act as facilitator, and become conversant with the philosophy and mechanics of grading system. For that it is necessary to have pre-service and in-service training of teachers in grading procedures.
- The problem with traditional grading is that students have good reasons to worry about their grades, and yet being grade-oriented undermines the most important goals of liberal

arts education. Students have good reasons to worry about their grades because of the powerful symbolic and social roles, that grades play in students' lives.

- Grades also have acquired increasingly powerful social force. Grades are not at all private communication between teachers and students but have a quasi-public role in students' lives. The

- occasions in which students are asked to reveal their grades are frequently some of the most significant moments in students' lives and connect with some of students' most important relationships.

- Grades are essentially numerical and thus can only be appropriately applied to what is measurable, but not everything that is measurable is always measured in a course of study. A Student too oriented towards getting good grades can miss or neglect those components of the course that are not grade.

- Research in cognitive science and developmental psychology reveals that human learning is extraordinarily complex. While professors can look in on some aspects of the learning process and judge whether students are putting certain words, numbers, or symbols together correctly—even reading past the words and numbers to more general conclusions about students' conceptual development—there is much about students' learning that remains invisible even to the most attentive professors.

- When teachers strive explicitly to structure their classes in ways that foster the development of self-motivation and push students to engage authentically with their education, they can feel that their ideals are consistently undermined by their students' efforts to play it safe and try to please the teacher.

- Students feel that they cannot trust their teachers or their environment, they become guarded and try to play it safe. Their reaction to challenges or criticism becomes defensive. If they feel themselves to be in a hostile context, a context of distrust demanding that they prove their worthiness at every turn.

- Because so much outside of the classroom hinges on students' grades, traditional grading imposes an outermost context of distrust framing everything that then happens within the classroom.

19.8 Keywords

Grading   :   Marking with some particulr symbol or word

Components   :   One of several parts of which something is made

Absolute   :   Total and complete

19.9 Review Questions

1. Explain the concept of grading.

2. What is the need of grading ? Explain.

3. Describe different methods of grading.

4. Discuss various problems of grading

Answers : Self Assessment

1. (*i*) Grading   (*ii*)   record, student growth
    (*iii*) Grade point Average

2. (*i*) a   (*ii*) c   (*iii*) b   (*iv*) d

19.10 Further Readings

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 20 : Methods of Feedback for Students

CONTENTS

Objectives

Introduction

20.1 Meaning of Feedback

20.2 Objectives of Feedback for Students

20.3 Different Aspects of Feedback for Students

20.4 Types of Methods of Feedback for Students

20.5 Advantages and Disadvantages for Different Methods of Feed back

20.6 Summary

20.7 Keywords

20.8 Review Questions

20.9 Further Readings

Objectives

The objectives of this unit can be summarised as below :

- To explain about the meaning of feedback
- To discuss about the objectives of feedback for students
- To describe the levels of feedback
- To explain the types of methods of feedback for students
- To discuss about the advantages of feedback

Introduction

Feedback from students has always played an important role in the maintenance of quality and standards in higher education. As quality assurance arrangements have themselves become more formalised, so too have the arrangements for the collection, analysis and use of student feedback. However, there are a number of other contextual changes that have influenced these arrangements.

20.1 Meaning of Feedback

Feedback refers to the Information teachers give to learners about their performance on an assessment.

There are many different ways to do this. Most good teachers provide students with feedback. If we see a teacher's role as someone who is a facilitator of learning and someone who provides opportunities to learn, then we will make assessment feedback an opportunity for children to learn. This means that our feedback will provide students with constructive comments and suggestions where they may improve and acknowledgement of the learning and thinking that has been shown in the assessment exercise. Feedback can contain written or verbal communication to the student about their work. The feedback should contain suggestions on how to improve as well as feedback about their efforts.

Positive feedback suggests to learners what they might do to improve. Here are some examples of positive feedback a teacher could provide :

- "I can see that you have tried very hard on this task."

- "You have been seriously thinking about this topic."

- "Try writing this sentence so it is clear that you are disagreeing with the author."

- "You have not included any labels for the diagram, check your exercise book if you are having trouble."

Usually, negative feedback does not help learners and it should be avoided. Negative feedback makes learners feel unsuccessful, ashamed and unable to do the work. Instead of motivating learners, negative feedback tends to push learners away from accomplishing the desired tasks. Negative feedback may hurt some children deeply. Avoid negative feedback such as : "Pull up your socks!" "You can do better." "You have been lazy."

Feedback should be timely. That means that the learners should get back their assignments and projects with the feedback soon after they have handed them in to the teacher. If a teacher waits too long to mark and hand back the assignments that provide feedback, the learners may not show much interest in the feedback or they may forget some things about the work they did and not benefit as much from the teacher's feedback. Learners should receive feedback on the assessment activities while the memory of working on it is still "fresh" in their mind.

Here are some things to consider when providing learners with feedback.

- Feedback to learners from the teacher tells the student **how** to improve, Feedback is **instructive**. Learners can see their own **progress**. Learners start thinking about the quality of their work. All learners stay engaged in learning.

- Learners who are falling behind need extra support.

- Learners who are moving ahead of others need help to continue being challenged by school work.

- Learners come to know something about themselves, the way they learn best and how they can improve their own learning skills.

- Learners develop a positive feeling about themselves.

Notes Feedback data will be more useful if they contain or can be related to student profile data (for example, age, gender, mode of study, ethnic background). This is an important issue to consider at institution and programme levels and for modules with large and diverse student groups.

20.2 Objectives of Feedback for Students

The purposes of collecting and using student feedback are in the following matter:-

In responding to the consultation exercise, institutions referred to the main purposes, of student feedback as :

• enhancing the students' experience of learning and teaching

• contributing to monitoring and review of quality and standards.

Other purposes cited included :

- ensuring the effectiveness of course design and delivery

- enabling a dialogue with students

- helping students reflect upon their experiences

- as part of the teaching and learning process

- identifying good practice

- measuring student satisfaction
- contributing to staff development.

Although overwhelmingly concerned with a notion of quality enhancement, purposes nevertheless differed in emphasis : for example according to whether student feedback was considered to be a part of the student learning process or rather a commentary on that process; according to whether it was seen as being primarily about whether programme objectives were being achieved or providing an opportunity to critique those objectives.

Underlying these differences of emphasis could be different conceptions of student feedback. While some people equated student feedback with student satisfaction, this view was by no means universal. Student feedback could be accounts by students of their learning styles and study methods - clearly potentially valuable to their teachers but not calling for evaluation or opinion from the students. Student feedback could be student views about whether *their* objectives had been met. While such information is clearly related to 'satisfaction' it is not necessarily the same as feedback on satisfaction with the teaching and learning processes of specific programmes, modules or services.

20.3 Different Aspects of Feedback for Students

20.3.1 A range of uses and users

It is of course likely that different users will have different purposes for student feedback. The class teacher will look to feedback hopefully for endorsement of his or her teaching approach but also for guidance on what worked and what did not and way in which it might be improved next time. The class teacher might also be more interested to know how far the students understood what was being taught than whether they liked it or were satisfied. A course or programme team might be looking to student feedback for evidence that learning objectives had been met, that the programme as a whole cohered. Departmental or other committees might be looking at student feedback alongside a range of quality indicators as part of institutional quality assurance arrangements. They might be particularly interested in trend data or comparisons between programmes or in using feedback to understand better the implications of data from other sources, for example a higher than expected drop-out rate or some particular criticisms by an external examiner. For current students, the purpose might simply be to express a view, positive or negative. For students considering taking a programme or module, feedback from previous cohorts of students could aid module choice.

20.3.2 The need for discussion

To this range of perfectly legitimate purposes of different uses and users of feedback data is the possibility that the purposes of the exercise have been lost in time, that it has become just another of the institution's rituals, that engagement with the activity has become largely an act of compliance, that feedback is collected but little is done with it, and that those involved in the process see little point in it, have few expectations that change will occur as a result of it. Where such a situation occurs, there is of course little point in collecting feedback data at all.

Thus, we would emphasise the value of discussion about the purposes of student feedback - and discussion at suitable intervals to check that purposes have not changed. This should include all who are involved, whether providing data, analysing and interpreting data, or using data. The purposes of the users will have primacy and discussion will reveal whether different users have different purposes, whether they are compatible with each other, whether they can be achieved in relation to the time and resources available, including the time required of the data providers-the students themselves. Where institutions wish to have clarity and consistency about purposes across the whole institution, the need to obtain wide understanding and agreement will be essential. More devolved arrangements are probably easier to achieve and more likely to generate

local ownership and commitment but there may be costs in terms of efficiency and a loss of comparability and analytic capacity.

20.3.3 Level

In discussing the purposes of student feedback, a crucial issue is that of level. Levels include :

- An individual lecturer or class
- A module or unit
- A semester or year of study
- A programme of study
- A subject
- A department
- A faculty
- An institution.

As we have already noted, different users will require feedback at different levels and for different purposes. These may or may not be compatible. The purposes of users at the 'higher' levels of institutions are likely to require a degree of standardisation of data - both in regard to what is collected, how it is analysed and how it is presented (see below). Users closer to the 'chalk-face' have little need of standardisation and it may be in conflict with their needs. This might imply different data collections to meet the purposes of users at different levels, or perhaps reliance on some mechanisms rather than others to achieve certain purposes (for example, student representatives rather than feedback questionnaires).

20.3.4 Use at other levels

Recognition that purposes differ, especially at different levels within an institution, could of course result in a proliferation of feedback activity as every user emphasised the unique nature of his or her needs for feedback. In such situations, the question inevitably arises as to whether data collected at one level for one purpose can also be used at other levels for other purposes. Can data be aggregated or, for that matter disaggregated, to serve different purposes ? It is certainly the case that aggregating module feedback does not convert it into feedback on the student experience of the programme as a whole. It does, however, provide some information on the module 'set' that constitutes the programme and this is likely to be of interest to those with responsibilities at this level.

Nevertheless, if feedback is to be obtained about the student experience of a whole programme of study, this is likely to require feedback obtained at that level. This might be achievable through a one-off exercise during a 3 or 4 year programme and so does not represent a large additional load upon staff and students, who may also be obtaining feedback on each separate module. It is also likely that data obtained by the new national survey of graduates will provide institutions with some valuable information on the student experience at the broad subject level.

Did u know? Module level feedback can be of use at the programme level in conjunction with other relevant data, for example student profiles, progression data, external examiner reports.

20.3.5 Timing

Another discussion point should be that of timing. When should feedback data be collected ? When will it be needed ? When can it be used to best effect ? End of module questionnaires are common but they cannot provide information to improve the learning experiences of the particular

group of students taking the module. Maybe that is not important, as long as there are sufficient opportunities for informal exchange between students and teachers to render the use of formal mechanisms redundant. But this may not always be the case. A mid-point feedback questionnaire might be useful on large enrolment programmes but, if this is followed by an end of module questionnaire, it may be seen as overkill by students and staff alike, especially if the module is semester length.

A further point about timing of feedback concerns its links to action and change. At higher levels within institutions, ensuring that feedback is available to meet the timetable of committees may be essential if it is going to affect decision-making and action. Thus, it may be necessary to work back from the committee timetable to determine when feedback should be obtained.

20.3.6 Standardisation

A further discussion point is whether the purposes require feedback arrangements to be standardised. As indicated above, for some purposes this will undoubtedly be helpful. For other purposes, not only will standardised arrangements be unhelpful but a bespoke, one-off exercise might be called for. Thus, if the effects of a particular programme innovation are to be assessed, specific feedback on that innovation will need to be obtained. It is by no means clear that standardised arrangements would achieve the purpose. Or to take another example, the purposes of collecting feedback from students on an entirely new degree programme may differ from the purposes of collecting it on an established, long-running degree.

20.3.7 Clarity of purpose

The main point we would emphasise, however, is that the purposes of student feedback should be clear to all involved -especially students - if their commitment to the process is to be maximised.

We found a number of examples where purpose was stated explicitly - especially in the use of questionnaires. It is also important to state how feedback will be used and how results/actions will be disseminated to students. Such information is often stated in guidelines to staff and students. However, we believe it is especially important to state purpose and use, and how results and actions will be reported at the point when feedback is being requested.

Task   Why should standardisation is important in feadback mechanism.

Discussion about the purposes of student feedback inevitably leads on to discussion about the various mechanisms that can be used. For example, the existence of a well-publicised complaints procedure or a discussion during class may provide more effective ways of checking that a module or programme is going well than obtaining feedback through a questionnaire. The role of different kinds of feedback mechanisms is the subject of the next section.

This section considers the types of mechanisms that exist for collecting student feedback and those which are most commonly used by institutions. Each has implications for collecting students' views, analysing and interpreting the results, and feeding back the results and actions taken to students. Thus, we have also included at the end of this section a basic summary of the advantages and disadvantages of the different mechanisms.

20.4 Types of Methods of Feedback for Students

Institutions use many mechanisms to collect student feedback, both qualitative and quantitative. They include :

- Questionnaires
- Student representation on local and institutional committees
- Staff student liaison committees (or their equivalent)
- The lecture or seminar
- The tutorial
- Discussion groups (for example, focus groups, structured group discussions, nominal group technique and so on)
- Other informal mechanisms.

20.4.1 A combination of mechanisms

Any single mechanism has its drawbacks. Feedback through questionnaires is generally considered to be relative or indicative rather than absolute. It may for example be affected by factors such as the timing of the distribution of the questionnaire and whether it is distributed 'in-class' or by some other means. Student representation on the other hand cannot be relied upon to be truly representative of the student body. In recognition of the limitations of any single mechanism, all institutions use a range of mechanisms, generally ensuring a mix of qualitative and quantitative feedback. Quantitative feedback (for example, through questionnaires) can be used to provide 'evidence' that something is going well or not so well and such evidence will normally be required for quality assurance purposes. Qualitative information (for example, through open-ended response sections of questionnaires and from student representatives) can help explain *why* something is going well or not so well.

Thus, while questionnaires are by far the most commonly used mechanism, we found few situations where complete reliance was placed upon them, or indeed upon any one mechanism. Most institutions use a combination of formal mechanisms normally questionnaires, student representation and staff student liaison committees - although informal channels of feedback are also highly valued.

Indeed, a number of institutions require or encourage departments to use a combination of mechanisms.

Pack on student feedback provides details of a number of techniques, including the following :
- Structured feedback sessions
- Focus groups
- Self research
- Log books
- Tutorials
- Questionnaires and surveys
- Shadowing
- Course representatives
- Informal chats
- One minute methods
- Student diaries
- Suggestion boxes/books.

In addition to outlining the purpose, structure and outcomes of each technique, including the ideal group size, it describes the pros and cons of each mechanism.

20.4.2 Additional mechanisms

A combination of questionnaires, student representation and staff student liaison committees represents common practice in most institutions. Many individual teachers will employ additional mechanisms, for example by holding class discussions mid-way through a module.

Some institutions supplement the main mechanisms of gathering student feedback by using

discussion groups. For the purposes of this Guide, the term 'discussion group' is used generically; however, discussion groups can take various forms, as exemplified by the list outlined above in Nottingham Trent University's Staff Resource Pack.

Discussion groups are less commonly used in institutions because they can be resource intensive and require specialist training of staff to be effective. They provide an alternative to questionnaires

and student representation (although not necessarily a replacement) and can be a rich source of information. Indeed, some institutions use these mechanisms in order to find out what students think about the use of a particular mechanisms for gathering student feedback (for example, the London School of Economics has used focus groups to investigate students' views about the issues surrounding an online questionnaire) or what questions students believe should be part of a questionnaire survey (as used for the University of Central England's student satisfaction survey). We believe that the use of discussion groups in certain specific circumstances should be considered and Section four of this Guide includes a discussion of these techniques.

20.5 Advantages and Disadvantages for Different Methods of Feedback

Table 1 : A summary of some of the advantages and disadvantages of mechanisms used together student feedback		
Mechanism	**Advantages**	**Disadvantages**
Questionnaires	Useful for gathering responses on many issues at various levels	Need for some specialist skills (e.g. in questionnaire design and analysis)
	Inclusive (assuming a census and a good response rate)	Tend to be 'ex post'
	Can provide quantitative and qualitative information	Frequency and number may induce questionnaire fatigue
	Depending on the degree of standardisation, can provide comparisons and trends	Response rates may be low
	Does not require a great deal of time commitment from individual students	May be costly in terms of time spent on analysis
		Students may not get information on results and any actions taken
		Actions may not get taken Can become ritualistic
Student representatives/staff student liaison committees	Provides a means of on-going formal dialogue between staff and students	Possible lack of motivation and participation by students
	Can provide student input of a 'feed forward' nature Good opportunities for dialogue and testing out ideas	Representatives may be unrepresentative Representatives may not be visible to the student population they represent

	Actions can be taken immediately or relatively quickly	Lack of opportunity for representatives to communicate with the student population
	Provides personal development opportunities for students	Time taken to ensure that representatives are effectively trained in the necessary skills
Discussion groups	Useful for engaging students in dialogue and highlighting issues of concern	May be resource intensive (in terms of student /staff time, analysis and interpretation of results)
	Can provide a rich source of information	The student group may not be representative of the student population
	Good opportunities for dialogue and testing out ideas	May be intimidating for some students
	Enables exploration of an issue (s) in depth	Vocal students may dominate discussion
	Facilitator can be neutral	Compromises student anonymity and confidentiality
		Requires specialist input to secure effective facilitation and discussion
Lecture/seminar	Can find out immediately how things are going	May be intimidating for some students
	(Some) problems can be dealt with immediately	Vocal students may dominate discussion
		Compromises student anonymity and confidentiality
Tutorial	Useful for more personal exchanges of views on the student's experience	Anecdotal - does not provide hard data
	Can find out immediately how things are going	Resource intensive
	(Some) problems can be dealt with immediately	
	Sensitive issues can be discussed confidentially	
Informal	Useful for more personal exchanges of views on the student's experience	Anecdotal
	Can find out immediately how things are going	
	(Some) problems can be dealt with immediately	
	Sensitive issues can be discussed confidentially	
	Provides a means of on-going informal dialogue between staff and students	

20.6 Summary

- Use of a combination of mechanisms will be more effective than reliance on a single one.

- Choice of mechanisms should be determined by the purpose, level and context of gathering and using student feedback.

- Issues of timing and frequency should be taken into account, especially with regard to the intended uses of the data.

- All involved, including students, should understand the purposes and the potential benefits of taking part in the process.

- Staff may require support (including training) in order to make the most of some mechanisms.

Self Assessment

1. Fill in the blanks :

(*i*) _____ is the information which teacher gives to learness about their performance on an assessment.

(*ii*) Feedback can contain _____ or verbal Communication to the student about their work.

(*iii*) _____ are less Commonly used in institutions because they can be resource intensive and require specialist training of staff.

(*iv*) A _____ feedback questionnaire might be useful on large enrolment programme.

(*v*) Feedback through _____ is generally Considered to be relative or indicative rather then absolute.

20.7 Keywords

Feedback   :   Information about how good or useful something or somebody's work is

Instructive   :   Giving a lot of useful information

Discussion   :   A Conversation about somebody/something

20.8 Review Questions

1.   What do you understand by feedback for students

2.   What is the purpose of feedback ?

3.   What are the different methods of feedback ?

4.   Write short note on the following topics in reference of feedback

(*i*) Timing   (*ii*) Standardisation   (*iii*) Clarity

Answers : Self Assessment

(*i*) Feedback   (*ii*) written   (*iii*) Discussion groups

(*iv*) midpoint   (*v*) Questionnaire

20.9 Further Readings

Books

1.   Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2.   Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3.   Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4.   Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 21 : Semester System vs Annual System

Objectives

The objectives of this unit can be summarised as below :

- To discuss about the annual System of Examination

- To describe the merits and demerits of Annual Examination

- To explain about the improvement is Exam System

- To discuss about the Semester System

- To describe the merits and demerits of Semester System

- To explain about the Annual System VS semester System

Introduction

The purpose of examination is to judge and evaluate the intelligence and ability of a student. It may be oral or written, theoretical or practical. So it must be conducted in such a way that a student can easily express the originality of his thought. More emphasis should be the Study of Education the moral and cultural aspects of a student's character as without character-building education is of little or no use. There are two kinds of examination, internal or home Examination and external or boardor university examination. The internal examinations are conducted on monthly, quarterlyor annual basis and its aim is to evaluate the performance of students. Similarly external examinations are also conducted on semester or annual basis. But the aim of these examinations are to promote the students from lower classes to higher classes and award them certificates, diplomas or degrees. In our country different institutions follow different systems of examination and both the systems have their own merits and demerits.

21.1 Annual System of Examination

In annual system of examination the relevant, institution has to prepare questions papers, arrange for its marking and conducting examination only once a year. The students prepare themselves for the annual examination and their tension is relieved in a single chance but the problem is that as compared to semester system they have to make through study of the whole course included in their syllabus.

21.2 Merits of Annual Examination

The scholars who advocate internal and continuous evaluation, anticipate the following merits.

(1) The students shall remain regular in their studies in the internal and continuous evaluation, and they will emphasis on understanding in place of learning by rote.

(2) The teachers will correctly evaluate the student's achievements in internal and continuous evaluation, they will be able to know about their progress and on that basis of it they will be able to guide them properly.

(3) Chance factor will be eradicated in the internal and continuous evaluation, only those students will score better who will be eligible for it.

(4) The results of internal and continuous evaluation will be reliable.

21.3 Demerits of Annual Examination

Some universities in our country have adopted this system, but without any benefit, rather the following demerits took place. We can call them the demerits of the system.

(1) This system can be applied only where the ratio of teacher-student is less.

(2) In this system, the teachers have ample opportunity for favouritism.

(3) Generally, those students score better in internal and continuous evaluation who can arrange recommendations for them or may impress the teachers otherwise.

(4) In this system the teachers favour their students and award them more marks than they deserve. It causes downfall in the standard of education.

Suggestion About Annual Examination System

We have, in the meantime, done several experiments for the improvement in the examinations, including internal and continuous evaluation. There were no benefits from the application, of internal and continuous evaluations rather there could be seen misadministration in the field of education and the downfall in the standard of education. The truth is that there is no alternative to the public examinations, the only thing required is to improve upon them, to give weightage to the oral, practical and written examinations, and to give equal weightage to essay type, short answer type and objective type tests in the written examinations, to declare use of unfair means as legal offence and to exercise care in the evaluation of answer-books. In the public examinations, we are in favour of the administration of both types of examinations—oral or practical and written. We are also in favour of all three types of tests — essay type, short answer type and objective type. We are also in favour of giving 25% weightage to each of oral or practical, essay-type, short-answer-type and objective-type examinations. The oral or practical examinations at secondary level should be fully internal, and at the graduate and post graduate levels should be conducted with the help of internal and external examiners both. It would not be justified to give more than 25% weightage to internal evaluation.

21.4 Improvement in Examination System

Radhakrishnan Commission (1948-49) gave highest priority to the improvement of examinations in the field of higher education. It stated in clear words — If we have to give only one suggestion as regards improvement in higher education, it would be improvement of the examination. The commission was of the opinion that all other improvements would come by of their own if the examination system is improved. It gave several suggestions for the improvement in university examinations. After this, the Mudaliar Commission (1952-53) recommended for improvement of the public examinations being administered at the secondary level.

The Kothari Commission (1964-66) also considered this problem in depth. The National Educational Policy, 1968 and the National Education Policy, 1986 too have emphasised improvement in examinations. In the meantime, various Provincial Secondary Education Boards, National Council of Educational Research and Training (NCERT), University Grants Commission (UGC) and various committees under various university have given suggestion for the improvement of examinations. Some of the recommendations have also brought into force.

Notes   Some recommendations have proved to be very effective; such as administration of three types of examinations — oral, written and practical, use of all the three types of written examinations — essay type, short answer type and objective type, carefulness in the construction of examination papers, honesty in the administration of examinations, declaring the use of unfair means in examinations as legal offence, and centralised evaluation of answer-books.

On the other hand, some experiments have failed as well, such as, semester system, internal and continuous evaluation, question bank, book-examination system and grade system. However, some scholars still vote for them, so it is necessary to discuss them here.

Did u know?   In 1958, the Central Examination Unit of the Union Ministry of Education considered the merits and demerits of the public examinations being held at secondary level and gave several suggestions for their improvement.

21.5 Semester System

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

21.6 Merits of Semester System

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

(1)   The chief merit of this system is that no student (examinee) is declared to be failed in it. A whole year is not lost for the failure candidate to appear again at the examination, and there

comes a difference of only six months if such incident occurs in the final semester. It eradicates stagnation in the field of education.

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

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

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

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

21.7 Demerits of Annual System

The semester system has its own limitations and demerits :

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

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

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

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

21.8 Annual System VS Semester System

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

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

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

Task   What are the objectives of semester system ?

The semester system allows greater interaction with teachers and the students will be more focussed on preparing throughout the year.

Self Assessment

1. Fill in the blanks :

(*i*) The examinations are conducting only once a year called _____ examination.

(*ii*) Annual system cover more _____ and compels the student to remember all this till the end of the year.

(*iii*) The chief merit of _____ system is that no student is declosed to be failed in it.

(*iv*) The students attend the classes in the semester system involve them selves in the studies _____.

(*v*) Semester System allows greater _____ with teachers and students will be more focussed on preparing throughout the year.

21.9 Summary

- External examinations are also conducted on semester or annual basis. But the aim of these examinations are to promote the students from lower classes to higher classes and award them certificates, diplomas or degrees. In our country different institutions follow different systems of examination and both the systems have their own merits and demerits.

- The scholars who advocate internal and continuous evaluation, anticipate the following merits :

  (*i*) The students shall remain regular in their studies in the internal and continuous evaluation, and they will emphasis on understanding in place of learning by rote.

  (*ii*) The teachers will correctly evaluate the student's achievements in internal and continuous evaluation, they will be able to know about their progress and on that basis of it they will be able to guide them properly.

- Radhakrishnan Commission (1948-49) gave highest priority to the improvement of examinations in the field of higher education. It stated in clear words — If we have to give only one suggestion as regards improvement in higher education, it would be improvement of the examination. The commission was of the opinion that all other improvements.

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

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

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

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

- The semester system has its own limitations and demerits :

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

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

- Comparison between semester and annual systems is very often done. Both the systems have its merits and demerits. Annual system is the traditional system. Annual system covers more syllabuses at a stretch and compels the student to remember all this till the end of the year. Sometimes, two or more topics will be included in the same paper (very often, a paper will have to be set by two examiners under such situations), when specializations are there. Otherwise, certain topics will be omitted and the syllabus diluted. Since at the end of the year only the public examinations are conducted University gets enough time to prepare question papers and value answer papers. Number of examiners and examinations also can be reduced.

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

- Since examinations come within a few months student unrest also will be less in a semester system. There were many challenges to be faced by the under graduate colleges in preparing the students for the semester system. The semester system is a very proactive system as it engages both the faculty and the students throughout the year in academic activity.

21.10 Keywords

Annual   :   happening or done once every year.

Semester   :   One ofthe two periods that the school or college year is divided

Continuous :   happening or existing for a period of time

21.11 Review Questions

1.   Explain the term annual and semester system of examination.

2.   What is the difference between annual and semester system of examination.

3.   Give merits and demerits of semester system.

Answers : Self Assessment

1.   (*i*)   annual   (*ii*)   syllabus   (*iii*)   semester   (*iv*)   continuously

     (*v*)   interaction

21.12 Further Readings

Books

1.   Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2.   Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3.   Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4.   Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 22 : Continuous Assessment

CONTENTS

Objectives

Introduction

22.1 Meaning of Continuous Assessment

22.2 Need of Continuous Assessment

22.3 Purposes of Continuous Assessment

22.4 Types of Continuous Assessment

22.5 Principles of Continuous Assessment

22.6 Advantages of Continuous Assessment

22.7 Summary

22.8 Keywords

22.9 Review Questions

22.10 Further Readings

Objectives

The objectives of this unit can be summarised as below :

- To explain the meaning of Continuous assessment
- To discuss about the need of Continuous assessment
- To describe about the purposes of Continuous assessment
- To explain about the advantages and problems of Continuous assessment

Introduction

In today's policy environment, testing has become a critical component of education reform. Policy makers and education administrators often view test scores as a measure of educational quality and use test scores to hold schools accountable for teacher performance. Continuous assessment, an alternative or supplement to high stakes testing of pupil achievement, offers a methodology for measuring pupil performance and using those findings to improve the success of pupils.

22.1 Meaning of Continuous Assessment

Continuous assessment is a classroom strategy implemented by teachers to ascertain the knowledge, understanding, and skills attained by pupils. Teachers administer assessments in a variety of ways over time to allow them to observe multiple tasks and to collect information about what pupils know, understand, and can do. These assessments are curriculum-based tasks previously taught in class. Continuous assessment occurs frequently during the school year and is part of regular teacher-pupil interactions. Pupils receive feedback from teachers based on their performance that allows them to focus on topics they have not yet mastered. Teachers learn

which students need review and remediation and which pupils are ready to move on to more complex work. Thus, the results of the assessments help to ensure that all pupils make learning progress throughout the school cycle thereby increasing their academic achievement.

22.2 Need of Continuous Assessment

The continuous assessment process is much more than an examination of pupil achievement. Continuous assessment is also a powerful diagnostic tool that enables pupils to understand the areas in which they are having difficulty and to concentrate their efforts in those areas. Continuous assessment also allows teachers to monitor the impact of their lessons on pupil understanding. Teachers can modify their pedagogical strategies to include the construction of remediation activities for pupils who are not working at the expected grade level and the creation of enrichment activities for pupils who are working at or above the expected grade level. Hence, the continuous assessment process supports a cycle of self-evaluation and pupil-specific activities by both pupils and teachers.

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

In continuous assessment, teachers assess the curriculum as implemented in the classroom. It also allows teachers to evaluate the effectiveness of their teaching strategies relative to the curriculum, and to change those strategies as dictated by the needs of their pupils. In addition, continuous assessments provide information on achievement of particular levels of skills, understanding, and knowledge rather than achievement of certain marks or scores. Thus, continuous assessment enables pupils to monitor their achievement of grade level goals and to visualize their progress towards those goals before it is too late to achieve them.

Notes   One-to-one communication between the teacher and the pupil can motivate pupils to continue attending school and to work hard to achieve higher levels of mastery.

22.3 Purposes of Continuous Assessment

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

The purposes of continuous assessment are achieved in the following ways :

1.   **Longer time for collecting assessment information**

     To obtain accurate and reliable assessment data on a pupil, the assessments could be spread over a longer time, allowing the pupil to take tests and other assignments at different times

throughout the course. The average of the scores for the various assessments is a more reliable indicator of the pupil's performance in the subject than the score the pupil obtains in a one-shot examination.

2. **Use of different test forms and different test situations**

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

3. **Inclusion of more complex thinking skills in the testing programme**

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

4. **Teacher assistance and remediation**

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

22.4 Types of Continuous Assessment

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

Daily work

*   **Pros**

    *   This is likely to be the most sustained means of assessment, and in providing relatively prompt feedback, it serves to reinforce or correct learned responses.

    *   It assists in pacing learning. If a course is broken into units, each assessed on completion, students have regular feedback of what they have mastered (and what they need additional work on).

    *   As the most extensive means of assessment, it has a sustained impact on and improves the quality of student learning.

*   **Cons and possible solutions**

    *   As it is fairly labour intensive for both students and the teacher, decisions will have to be taken regarding an optimum level. Too many assignments may result in justice not being done to them all. It is probably better to have regular but smaller assignments than large-

scale but infrequent ones. Experience indicates that giving students more feedback results in faster learning and less confusion.

- Vigilance will be needed to ensure that plagiarism and other forms of cheating do not go undetected. Make clear to students that a very serious view will be taken of such offences. Point out that it is not only dishonest but short-sighted to abuse the important learning experience involved in preparing an assignment. As a countercheck, arrange for some of the assignments to be done and completed in class.

Projects/term papers

- **Pros**

  They provide a measure of the student's ability to :

  - collect, select and use data/information;

  - undertake independent study and enquiry;

  - plan and follow through a fairly large-scaled piece of work;

  - integrate theory and practice;

  - work with others, when the exercise involves teamwork.

- **Cons and possible solutions**

  - It is difficult to ensure parity in the projects. Careful coordination in the setting of topics is important.

  - Project work may lead to over-involvement at the expense of other aspects of the course. Tutors/supervisors should therefore guide students in defining tasks and in being selective with regards to content.

  - It may be difficult to maintain consistency and fairness in marking across a range of projects/papers. To mitigate this, project objectives should be clearly identified and some broad criteria for assessment set. Wherever possible, two examiners should be used.

  - Students may receive unequal supervision and this may affect the quality of the completed project. Agreement should be made among tutors/supervisors to standardise the degree of direction and guidance to be given to students. Criteria for evaluation should be available to students in advance as these provide direction.

  - It may be difficult to assign individual marks to students working in a group project, but this may be resolved by orally examining each student to establish their knowledge of the subject and the degree of his/her contribution. Alternatively, have students work on projects individually.

Practical work

- **Pros**

  It provides a measure of :

  - ability to relate theory to practice;

  - students' techniques, procedures and practical skills.

- **Cons and possible solutions**

  - The contribution to a student's final grade is low in relation to the amount of work involved. To give it due value, emphasise its importance as a learning process.

  - Too much emphasis may be given to the written report without sufficient consideration to the actual process and performance. To avoid this, instructors should scrutinise practical work closely and question students about their organisation of the work and their findings.

- The theoretical and practical aspects of the course may not be sufficiently linked. Obviously, it is important to communicate frequently with those involved in managing different parts of the course.

Self Assessment

1. *Fill in the blanks :*

   (*i*) _____ is a classroom strategy implemented by teachers to ascertain the knowledge, understanding and skills attained by pupils.

   (*ii*) Continuous assessment process supports a cycle of _____ and _____ by both pupils and teachers.

   (*iii*) _____ , _____ and _____ are the types of Continuous assessment.

Principles of Continuous Assessment

1. **Appreciative Inquiry : Continuous Assessment** uses the principles of appreciative inquiry as a basis for the development of the activities in each chapter. Appreciative inquiry is a method of figuring out how an organization or group can best undergo positive change. With appreciative inquiry, the starting point of change is acknowledging the strengths and skills you already have. The process of change begins by identifying assets (strengths) rather than problems. You will see that **Continuous Assessment** activities ask teachers and other educators to identify their strengths and what they are doing that works well.

   The next stage of appreciative inquiry which is also incorporated into many of the activities, is the dreaming stage. In this part of appreciative inquiry, those involved in the change are asked to think about how they might see their classroom, their teaching, or school in the future.

   The third stage of appreciative inquiry that you will see in this guide is designing the future. This is where teachers describe what they will do in order to make the changes they want happen. It is the road map or action plan for accomplishing the goals they have described.

2. **Adult Learning :** You will notice that each chapter focuses on activities. Many of the activities include working together with colleagues. We also include a series of discussion questions and actions that will help teachers to analyze their own teaching context in light of the information provided. These material embody the principles of Malcolm Knowles' theory of how adults learn (andragogy). The assumption has been made that every adult has had life and work experience and has gained some wisdom and insight from these experiences. Teachers, headteachers, district supervisors, and parents have important perspectives about their school's particular situation, These adults offer a rich source of ideas and talents that may be used to help their schools improve.

3. **Experiential Learning :** Adults learn through experience and reflection on that experience. **Continuous Assessment** asks teachers to try new techniques in their classrooms (experience) and then discuss what happened (reflection) with colleagues. The guide asks the reader or user to also spend time thinking individually about teaching experiences. The many activities in the guide are intended to provide teachers and others with experiences that will lead them to new ways of teaching and thinking about teaching. Teachers are encouraged to carry out the activities and answer the questions for each activity. In this way teachers will gain a deeper understanding of the teaching and learning strategies and ideas presented in this book.

4. **Understanding Reasons for Change :** It is important that adults have opportunities to analyze reasons for change. Most adults resist being told what to do. As adults we make choices about what we will do. For teachers, theory and reasons for implementing new strategies

and concepts are an important part of making a decision to try new techniques. Teachers need to understand why a particular innovation is beneficial to the learners.

5.   **Responsibilities for Change Rests with the Individual :** Ultimately, teacher themselves will decide whether or not to apply their new learning to the classroom. They may need to adapt their new skills and knowledge to fit their particular situation. Teachers are the agents of their own change.

Did u know? **Continuous Assessment** presents some background information that includes research findings from other contexts for many of the strategies, thus providing teachers with reasons for implementing a particular strategy. Teachers are also called upon to think of why a particular strategy might be beneficial in his or her context.

22.5 Principles of Continuous Assessment

Continuous Assessment Process

CAP is a system for supporting and evaluating all teaching and learning activities in the school.

CAP is geared towards :

*   ensuring that all children learn by utilizing a range of teaching and assessment

*   informing on students readiness for the next level of learning

*   identifying challenges to students' learning and providing support for appropriate intervention and remediation

Educators and Business agree that individuals need to be able to :

*   think critically

*   solve problems

*   work both in groups and independently

*   combine knowledge and skills from different subjects

*   communicate verbally and in writing

All children must therefore learn the basics and higher-level skills, since in the future, fewer opportunities for unskilled individuals would be available.

How Does CAP Relate to Students' Classroom Instruction ?

How will teaching, **learning** and **assessment** be carried out ?

*   A student-centred approach to instruction will be utilized with learning being activity based. Students are active participants in the teaching/learning process.

*   Students will be continuously assessed in all subject areas during the term through **Monthly Tests, Project and Term Test**.

*   **Student's Participation** is also assessed during this time.

*   Students in Standard 1 and III are also required to write the **National Test** in

    •   Language Arts and Mathematics

*   This test is constructed by the Ministry of Education and administered at the end of the school year.

Assessment is an integral part of the instructional process

- Teacher emphasize and assess skill in addition to knowledge

- Students learn to work both alone and in groups

- Students are actively involved in their own learning - they are taught to learn things on their own, as well as how to learn.

- Students learn to use many skills at the same time to solve complex tasks. Students are taught how to use skills they have learnt in many different subjects to complete an activity.

- Students learn how to solve problems, which require critical thinking and have more than one right answer.

Participation of Parents Guardians

(*i*) Be an active partner in the education of your child. Get involved in decision-making committees about school issues and problems, for example parent-teacher committees.

(*ii*) You can visit your child's teacher. A visit will give you an idea of what is expected of your child does at school and how he / she interacts with both the teacher and other children.

(*iii*) Support student events and performances by attending and assisting in whatever way you can.

Most teachers assign homework on a regular basis because practice is needed before children fully understand new skills and concepts. Homework also increases the amount of learning time available and allows students to do more in-depth learning.

- Help your child to develop a homework schedule that he / she can stick to.

  - Help your child to manage his/her work load.

  - Talk to your child each day about homework assignments; see if work is complete; ask questions about it. But do not do your child's homework.

  - Provide a suitable place for study. If possible, make it quiet and away from the distractions of TV, telephone or loud music).

- Avoid making homework a punishment

- Emphasize effort and achievement.

- Reward progress and display the child's good work.

- Have high expectations for your child's learning and behaviour, both at home and at school.

- Praise and encourage your child.

- Encourage your child to share information about school and respond with empathy.

- If you don't do anything else, read to your young child or have him/her read to you everyday.

- Encourage older children to read by having interesting and appropriate materials available.

22.6 Advantages of Continuous Assessment

One of the expected advantages of continuous assessment lies in its being guidance oriented. Since it will involve data gathering over a long period of time, it will yield more accurate data reaching the teachers early enough to modify instruction. This could play a vital role in diagnosing and remediating areas of learners' weaknesses if properly anchored in what occurs in classrooms. Continuous assessment is an approach that would capture the full range of learners' performance. Teachers and administrators would thus be able to assess learners' progress and would have time to correct the problems.

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

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

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

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

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

Task   What is Continuous assessment process ?

Self Assessment

2. State whether the following statements are 'true' or false'

(*i*) Continuous Assessment presents some background information that includes research findings from other contexts for many of the strategies.

(*ii*) Assessment is an integral part of the instructional process.

(*iii*) The Teacher involvement in continuous assessment is very less or not.

(*iv*) Record keeping is a problematic issue in continuous assessment for teachers.

(*v*) Continuous assessment does not help in capture the full range of learner's performance.

22.7 Summary

- Continuous assessment is a classroom strategy implemented by teachers to ascertain the knowledge, understanding, and skills attained by pupils.

- The continuous assessment process is much more than an examination of pupil achievement. Continuous assessment is also a powerful diagnostic tool that enables , pupils to understand the areas in which they are having difficulty and to concentrate their efforts in those areas. Continuous assessment also allows teachers to monitor the impact of their lessons on pupil understanding.

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

- The addition of continuous assessment in the instructional and testing process is intended to achieve two major purposes : to improve both the validity and reliability of the results of pupils' performance on tests and exercises, and secondly to help the pupil to develop effective learning and work habits. The present continuous assessment system is essentially based on frequent test taking and does not really serve the two critical purposes of continuous assessment.

- Classroom tests are based on assessment of lower level abilities and memorization. Where assessments are based on low level thinking skills i.e., "Knowledge" and "Comprehension", pupils complete their education still unable to analyze and apply their knowledge to solve problems.

- The central purpose of continuous assessment is to help the pupil to become a better learner and producer bv encouraging pupils to improve their knowledge and skills through learning, test taking and project undertaking in the critical and important objectives of the school curriculum.

- Types of Continuos Assessment

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

- Appreciative Inquiry : Continuous Assessment uses the principles of appreciative inquiry as a basis for the development of the activities in each chapter. Appreciative inquiry is a method of figuring out how an organization or group can best undergo positive change.

- Adult Learning : You will notice that each chapter focuses on activities. Many of the activities include working together with colleagues. We also include a series of discussion questions and actions that will help teachers to analyze their own teaching context in light of the information provided. These material embody the principles of Malcolm Knowles' theory of how adults learn (andragogy).

- Experiential Learning : Adults learn through experience and reflection on that experience. Continuous Assessment asks teachers to try new techniques in their classrooms (experience) and then discuss what happened (reflection) with colleagues.

- Understanding Reasons for Change : It is important that adults have opportunities to analyze reasons for change. Most adults resist being told what to do. As adults we make choices about what we will do. For teachers, theory and reasons for implementing new strategies and concepts are an important part of making a decision to try new techniques.

- Responsibilities for Change Rests with the Individual : Ultimately, teacher themselves will decide whether or not to apply their new learning to the classroom. They may need to adapt their new skills and knowledge to fit their particular situation.

- The Continuos Assessment Process

  - Continuous Assessment Process CAP is a system for supporting and evaluating all teaching and learning activities in the school.

- CAP is geared towards

  - ensuring that all children learn by utilizing a range of teaching and assessment

  - informing on students readiness for the next level of learning

- Educators and Business agree that individuals need to be able to

  - think critically

  - solve problems

  - work both in groups and independently

- How will teaching, learning and assessment be carried out ?

  - A student-centred approach to instruction will be utilized with learning being activity based. Students are active participants in the teaching/learning process.

  - Students will be continuously assessed in all subject areas during the term through Monthly Tests, Project and Term Test.

  - Student's Participation is also assessed during this time.

- Assessment is an integral part of the instructional process

  - Teacher emphasize and assess skill in addition to knowledge

  - Students learn to work both alone and in groups

  - Students are actively involved in their own learning - they are taught to learn things on their own, as well as how to learn.

- Be an active partner in the education of your child. Get involved in decision-making committees about school issues and problems, for example parent-teacher committees.

- You can visit your child's teacher. A visit will give you an idea of what is expected of your child does at school and how he / she interacts with both the teacher and other children.

  - Help your child to develop a homework schedule that he / she can stick to.

    - Help your child to manage his/her workload.

    - Talk to your child each day about homework assignments; see if work is complete; ask questions about it. But do not do your child's homework.

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

- Another advantage of continuous assessment is that it places teachers at the centre of all performance-assessment activities. It encourages more teacher participation in the overall

assessment or grading of his/her learners. Teachers must be given opportunities to select and review assessments so that they become involved and knowledgeable in the process.

- One of the important aspects of continuous assessment is the availability of valid and reliable tests which could be used in all schools.

- Apart from the skills of test construction measuring cognitive aspects of learning, teachers should also be able to measure the learners' affective attributes such as attitudes, motives, interests, values and other personality characteristics.

- For successful implementation of the continuous assessment approach, teachers need to give most tests, which means more marking. They need to observe the learners more keenly to assess their affective outcomes, and there will be more records to be kept on the learners.

- Another problem with continuous assessment is the issue of record keeping. Learners' records have to be adequately and meticulously kept over a long period of time.

22.8 Keywords

Continuous : Happening or existing for a period of time without interruption

Assessment : An opinion or a judgement about something that has been thought about very carefully

Inclusion : The fact of including something/Somebody

22.9 Review Questions

1. Explain the term "Continuous assessment".
2. Why is the need for Continuous assessment is evaluation ?
3. Give some ways by which purposes of Continuous assessment can be achieved.
4. What are the types of Continuous assessment. Give their advantages and disadvantages of Continuous assessment.
5. Give principles of Continuous assessment.
6. What are the advantages and disadvantages of Continuous assessment ?

Answers : Self Assessment

1. (*i*) Continuous assessment   (*ii*)   Self evaluation, pupil specific activities,
   (*iii*) Daily work, project papers, practical work
2. (*i*)   T   (*ii*) T   (*iii*) F   (*iv*) T   (*v*) F

22.10 Further Readings

Books

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*
2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*
3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*
4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 23 : Portfolio Assessment

CONTENTS

Objectives

Introduction

23.1 Concept of Portfolio Assessment

23.2 Need of Portfolio Assessment

23.3 Types of Portfolio Assessment

23.4 Steps in the Portfolio Assessment

23.5 Advantages of Portfolio Assessment

23.6 Disadvantages of Portfolio Assessment

23.7 Aims of Portfolio Assessment

23.8 Process of Portfolio Assessment

23.9 Summary

23.10 Keywords

23.11 Review Questions

23.12 Further Readings

Objectives

The objectives of this unit can be summarised as below

- To explain about the concept of portfolio assessment

- To discuss about the need of portfolio assessment

- To describe the types of portfolio assessment

- To discuss about the advantages and disadvantages of Portfolio assessment

Introduction

Student *portfolios* take many forms, as discussed below, so it is not easy to describe them. A portfolio is *not* the pile of student work that accumulates over a semester or year. Rather, a portfolio contains a purposefully selected subset of student work. "Purposefully" selecting student work means deciding what type of story you want the portfolio to tell. For example, do you want it to highlight or celebrate the progress a student has made ? Then, the portfolio might contain samples of earlier and later work, often with the student commenting upon or assessing the growth. The student and/or teacher might select items that illustrate the development of one or more skills with reflection upon the process that led to that development. Showcase the final products or best work of a student ? In that case, the portfolio would likely contain samples that best exemplify the student's current ability to apply relevant knowledge and skills. All decisions about a portfolio assignment begin with the type of story or purpose for the portfolio. The particular purpose(s) served, the number and type of items included, the process for selecting the items to be included, how and whether students respond to the items selected, and other decisions vary from portfolio to portfolio and serve to define what each portfolio looks like. I will describe many of the purposes and characteristics in the sections below.

23.1 Concept of Portfolio Assessment

A **student portfolio** is a systematic collection of student work and related material that depicts a student's activities, accomplishments, and achievements in one or more school subjects. The collection should include evidence of student reflection and self-evaluation, guidelines for selecting the portfolio contents, and criteria for judging the quality of the work. The goal is to help students assemble portfolios that illustrate their talents, represent their writing capabilities, and tell their stories of school achievement...

23.2 Need of Portfolio Assessment

In this new era of performance assessment related to the monitoring of students' mastery of a core curriculum, portfolios can enhance the assessment process by revealing a range of skills and understandings one students' parts; support instructional goals; reflect change and growth over a period of time; encourage student, teacher, and parent reflection; and provide for continuity in education from one year to the next. Instructors can use them for a variety of specific purposes, including :

- Encouraging self-directed learning.

- Enlarging the view of what is learned.

- Fostering learning about learning.

- Demonstrating progress toward identified outcomes.

- Creating an intersection for instruction and assessment.

- Providing a way for students to value themselves as learners.

- Offering opportunities for peer-supported growth.

Identifying specific goals or purposes for assigning a portfolio is the first and most critical step in creating such an assignment. Just as identifying a standard guides the rest of the steps of developing an authentic assessment, identifying the purpose(s) for a portfolio influences all the other decisions involved in producing a portfolio assgnment

23.3 Types of Portfolio Assessment

Process and product portfolios represent the two major types of portfolios. A **process portfolio** documents the stages of learning and provides a progressive record of student growth. A **product portfolio** demonstrates mastery of a learning task or a set of earning objectives and contains only the best work... Teachers use process portfolios to help students identify learning goals, document progress over time, and demonstrate learning mastery... In general, teachers prefer to use process portfolios because they are ideal for documenting the stages that students go through as they learn and progress

Self Assessment

1. *Fill in the blanks :*

   (*i*) A _____ is a systematic collection of student work and related material. Which shows the student's activities accomplishments and achievements in one or more school subjects

   (*ii*) A _____ demonstrate mastery of a learning task or a set of learning objective.

   (*iii*) A _____ documents the stages of learning and provider a progressive record of student growth.

   (*iv*) _____ is to show growth over time, to help develop process skills, to identify strengths and weakness, etc.

23.4 Steps in the Portfolio Assessment

First, the teacher and the student need to clearly identify the portfolio contents, which are samples of student work, reflections, teacher observations, and conference records. Second, the teacher should develop evaluation procedures for keeping track of the portfolio contents and for grading the portfolio... Third the teacher needs a plan for holding portfolio conferences, which are formal and informal meetings in which students review their work and discuss their progress. Because they encourage reflective teaching and learning, these conference are an essential part of the portfolio assessment process

23.5 Advantages of Portfolio Assessment

- Promoting student self-evaluation, reflection, and critical thinking.

- Measuring performance based on genuine samples of student work.

- Providing flexibility in measuring how students accomplish their learning goals.

- Enabling teachers and students to share the responsibility for setting learning goals and for evaluating progress toward meeting those goals.

- Giving students the opportunity to have extensive input into the learning process.

- Facilitating cooperative learning activities, including peer evaluation and tutoring, cooperative learning groups, and peer conferencing.

- Providing a process for structuring learning in stages.

- Providing opportunities for students and teachers to discuss learning goals and the progress toward those goals in structured and unstructured conferences.

- Enabling measurement of multiple dimensions of student progress by including different types of data and materials.

23.6 Disadvantages of Portfolio Assessment

- Requiring extra time to plan an assessment system and conduct the assessment.

- Gathering all of the necessary data and work samples can make portfolios bulky and difficult to manage.

- Developing a systematic and deliberate management system is difficult, but this step is necessary in order to make portfolios more than a random collection of student work.

- Scoring portfolios involves the extensive use of subjective evaluation procedures such as rating scales and professional judgment, and this limits reliability.

Scheduling individual portfolio conferences is difficulty and the length of each conference may interfere with other instructional activities.

Task   What is portfolio partner ?

23.7 Aims of Portfolio Assessment

Why might you use a portfolio assignment ? Portfolios typically are created for one of the following three purposes : to show growth, to showcase current abilities, and to evaluate cumulative achievement. Some examples of such purposes include

1. **Growth Portfolios**

   (a) to show growth or change over time

(b) to help develop process skills such as self-evaluation and goal-setting

(c) to identify strengths and weaknesses

(d) to track the development of one more products/performances

2.   Showcase Portfolios

(a) to showcase end-of-year/semester accomplishments

(b) to prepare a sample of best work for employment or college admission

(c) to showcase student perceptions of favorite, best or most important work

(d) to communicate a student's current aptitudes to future teachers

3.   Evaluation Portfolios

A few such possible samples of work in the following tables that could be included in each type of portfolio.

Growth Portfolios : What samples might be included ?

Purpose	Some possible inclusions
(a) to show growth or change over time	• early and later pieces of work • early and later tests/scores • rough drafts and final drafts • reflections on growth • goal-setting sheets • reflections on progress toward goal(s)
(b) to help develop process skills	• samples which reflect growth of process skills • self-reflection sheets accompanying samples of work • reflection sheets from teacher or peer • identification of strengths/weaknesses • goal-setting sheets • reflections on progress towards goal(s) • see more detail below under Process below
(c) to identify strengths/weaknesses	• samples of work reflecting specifically identified strengths and weaknesses • reflections on strengths and weaknesses of samples • goal-setting sheets • reflection on progress towards goal(s)
(d) to track development of one or more products or performances	• obviously, drafts of the specific product or performance to be tracked • self-reflections on drafts • reflection sheets from teacher or peer

Showcase Portfolios : What samples might be included ?

Purpose	Some possible inclusions
	• samples of best work • samples of earlier and later work to document progress

(a) to showcase end-of-year/ semester accomplishments	• final tests or scores • discussion of growth over semester/year • awards or other recognition • teacher or peer comments
(b) to prepare a sample of best work for employment or college admission	• cover letter • sample of work • reflection on process of creating sample of work • reflection on growth • teacher or peer comments • description of knowledge/skills work indicates
(c) to showcase student perceptions of favorite, best or most important	• samples of student's favorite, best or most important work • drafts of that work to illustrate path taken to its final form • commentary on strengths/weaknesses of work • reflection on why it is favorite, best or most important • reflection on what has been learned from work • teacher or peer comments
(d) to communicate a student's current aptitude	• representative sample of current work • match of work with standards accomplished • self-reflection on current aptitudes • teacher reflection on student's aptitudes • identification of future goals

Evaluation Portfolios : What samples might be included ?

Purpose	Some possible inclusions
(a) to document achievement for grading	• samples of representative work in each subject/unit/ topic to be graded • samples of work documenting level of achievement on course/grade-level goals/standards/objectives • tests/scores • rubrics/criteria used for evaluation of work (when applied) • self-reflection on how well samples indicate attainment of course/grade-level goals/standards/ objectives • teacher reflection of attainment of goals/standards • identification of strengths/weaknesses

(b) to document progress towards standards	• list of applicable goals and standards • representative samples of work aligned with respective goals/standards • rubrics/criteria used for evaluation of work • self-reflection on how well samples indicate attainment of course/grade-level goals/standards/objectives • teacher reflection of attainment of goals/standards • analysis or evidence of progress made toward standards over course of semester/year
(c) to place students appropriately	• representative samples of current work • representative samples of earlier work to indicate rate of progress • classroom tests/scores • external tests/evaluations • match of work with standards accomplished • self-reflection on current aptitudes • teacher reflection on student's aptitudes • parent reflection on student's aptitudes • other professionals' reflections on student's aptitudes

Notes   Portfolios provide an excellent vehicle for consideration of process and the development of related skills. So, portfolios are frequently included with other types of authentic assessments because they move away from telling a student's story though test scores and, instead, focus on a meaningful collection of student performance and meaningful reflection and evaluation of that work.

23.8 Process of Portfolio Assessment

Process : What processes will be engaged in during the development of the portfolio ?

One of the greatest attributes of the portfolio is its potential for focusing on the *processes* of learning. Too often in education we emphasize the products students create or the outcomes they achieve. But we do not give sufficient attention to the processes required to create those products or outcomes, the processes involved in self-diagnosis and self-improvement, or the meta cognitive processes of thinking. As a result, the products or outcomes are not as good as we or the students would like because they are often unsure how to get started, how to self-diagnose or self-correct or how to determine when a piece of work is "finished,"

Although a variety of processes can be developed or explored through portfolios.

Steps of Process

•   selection of contents of the portfolio;

- reflection on the samples of work and processes;

- conferencing about the contents and processes.

Selection of Contents

Once again, identifying the purpose(s) for the portfolio should drive the selection process. As listed in the tables above, different samples of student work will likely be selected for different purposes. Additionally, samples are selected might also differ depending on the purpose. For example, for an evaluation portfolio, the teacher might decide which samples need to be included to evaluate student progress. On the other hand, including the student in the decision making process of determining appropriate types of samples for inclusion might be more critical for a growth portfolio to promote meaningful reflection.

Furthermore, audiences beyond the teacher and student might have input into the portfolio, from team or department members, principals and district committees to external agencies to parents and community members. External audiences are most likely to play a role for evaluation portfolios. **However, it is important to remember there are no hard rules about portfolios. Anything can be included in a portfolio. Anyone can be involved in the processes of selection, reflection and evaluation of a portfolio.** Flexibility applies to portfolios as it does to any authentic assessment. That is, you should be true to your purpose(s), but you should feel no constraints on how you meet them with a portfolio assignment.

Notes   A showcase portfolio might be designed to include significant input from the student on which samples best highlight achievement and progress, or the teacher might primarily make those decisions.

How might the selection take place ?

What describe below are just a few of the many possible avenues for selecting which samples will be included in a portfolio. But these examples should give you a good sense of some of the choices and some of the decisions involved.

When

- **when a sample of work is completed :** at the point a piece of work is ready to be turned in (or once the work has been returned by the teacher) the student or teacher identifies that work for inclusion in the portfolio;

- **at periodic intervals :** instead of selecting samples when they are completed, the samples can be stored so that selection might occur every two (three, six or nine) weeks or once (twice or three times) every quarter (trimester or semester);

- **at the end of the :** unit, quarter, semester, year, etc.

By whom ?

- **by the student :** students are the most common selectors, particularly for portfolios that ask them to reflect on the work selected. Which work students select depends on the criteria used to choose each piece (see below).

- **by the teacher :** teachers may be the selector, particularly when identifying best pieces of work to showcase a student's strengths or accomplishments.

- **by the student and teacher :** sometimes portfolio selection is a joint process involving conversation and collaboration.

- **by peers :** a student might be assigned a "portfolio partner" or "portfolio buddy" who assists the student in selecting appropriate pieces of work often as part of a joint process involving conversation and collaboration. A peer might also provide some reflection on a piece of work to be included in the portfolio.

- **by parents :** parents might also be asked to select a piece or two for inclusion that they particularly found impressive, surprising, reflective of improvement, etc.

Based on what criteria ?

- **best work :** selection for showcase portfolios will typically focus on samples of work that illustrate students' best performance in designated areas or the culmination of progress made

- **evidence of growth :** selection for growth portfolios will focus on identifying samples of work and work processes (e.g., drafts, notes) that best capture progress shown on designated tasks, processes or acquisition of knowledge and skills. For example, students might be asked to choose

  – samples of earlier and later work highlighting some skill or content area

  – samples of rough drafts and final drafts

  – work that traces the development of a particular product or performance

  – samples of work reflecting specifically identified strengths and weaknesses

- **evidence of achievement :** particularly for showcase and evaluation portfolios, selection might focus on samples of work that illustrate current levels of competence in designated areas or particular exemplars of quality work

- **evidence of standards met :** similarly, selection could focus on samples of work that illustrate how successfully students have met certain standards

- **favorite/most important piece :** to help develop recognition of the value of the work completed and to foster pride in that work, selection might focus on samples to which students or parents or others find a connection or with which they are particularly enamored

- **one or more of the above :** a portfolio can include samples of work for multiple reasons and, thus, more than one of the above criteria (or others) could be used for selecting samples to be included

Reflection on Samples of Work

Many educators who work with portfolios consider the reflection component the most critical element of a good portfolio. Simply selecting samples of work as described above can produce meaningful stories about students, and others can benefit from "reading" these stories. But the students themselves are missing significant benefits of the portfolio process if they are not asked to reflect upon the quality and growth of their work. As Paulson, Paulson and Meyer (1991) stated, "The portfolio is something that is done *by* the student, not *to* the student." Most importantly, it is something done *for* the student. The student needs to be directly involved in each phase of the portfolio development to learn the most from it, and the reflection phase holds the most promise for promoting student growth.

In the reflection phase students are typically asked to

- comment on why specific samples were selected *or*

- comment on what they liked and did not like in the samples *or*

- comment on or identify the processes involved in developing specific products or performances *or*

- describe and point to examples of how specific skills or knowledge improved (or did not) *or*

- identify strengths and weaknesses in samples of work *or*

- set goals for themselves corresponding to the strengths and weaknesses *or*

- identify strategies for reaching those goals *or*

- assess their past and current self-efficacy for a task or skill *or*

- complete a checklist or survey about their work *or*

- some combination of the above

Reflection sheets

Probably the most common portfolio reflection task is the completion of a sheet to be attached to the sample (or samples) of work which the reflection is addressing. The possibilities for reflection questions or prompts are endless, but some examples I have seen include

Selection questions/prompts

- Why did you select this piece ?

- Why should this sample be included in your portfolio ?

- How does this sample meet the criteria for selection for your portfolio ?

- I chose this piece because ....

Growth questions/prompts

- What are the strengths of this work ? Weaknesses ?

- What would you work on more if you had additional time ?

- How has your ---------- (e.g., writing) changed since last year ?

- What do you know about ---------- (e.g., the scientific method) that you did not know at the beginning of the year (or semester, etc.) ?

- Looking at (or thinking about) an earlier piece of similar work, how does this new piece of work compare ? How is it better or worse ? Where can you see progress or improvement ?

- How did you get "stuck" working on this task ? How did you get "unstuck" ?

- One skill I could not perform very well but now I can is ....

- From reviewing this piece I learned ....

Goal-setting questions/prompts

- What is one thing you can improve upon in this piece ?

- What is a realistic goal for the end of the quarter (semester, year) ?

- What is one way you will try to improve your ---------- (e.g., writing) ?

- One thing I still need to work on is ....

- I will work toward my goal by ....

Evaluation questions/prompts

- If you were a teacher and grading your work, what grade would you give it and why ?

- Using the appropriate rubric, give yourself a score and justify it with specific traits from the rubric.

- What do you like or not like about this piece of work ?

• I like this piece of work because ....

Effort questions/prompts

• How much time did you spend on this product/performance ?

• The work would have been better if I had spent more time on ....

• I am pleased that I put significant effort into ....

Overall portfolio questions/prompts

• What would you like your (e.g., parents) to know about or see in your portfolio ?

• What does the portfolio as a whole reveal about you as a learner (writer, thinker, etc.) ?

• A feature of this portfolio I particularly like is ...

• In this portfolio I see evidence of....

As mentioned above, students (or others) can respond to such questions or prompts when a piece of work is completed, while a work is in progress or at periodic intervals after the work has been collected. Furthermore, these questions or prompts can be answered by the student, the teacher, parents, peers or anyone else in any combination that best serves the purposes of the portfolio.

Other reflection methods

In addition to reflection sheets, teachers have devised a myriad of means of inducing reflection from students and others about the collection of work included in the portfolio. For example, those engaging in reflection can

• write a letter to a specific audience about the story the portfolio communicates

• write a "biography" of a piece of work tracing its development and the learning that resulted

• write periodic journal entries about the progress of the portfolio

• compose an imaginary new "chapter" that picks up where the story of the portfolio leaves off

• orally share reflections on any of the above questions/prompts

Reflection as a process skill

Good skill development requires four steps :

• Instruction and modeling of the skill;

• Practice of the skill;

• Feedback on one's practice;

• Reflection on the practice and feedback.

Reflection itself is a skill that enhances the process of skill development and virtually all learning in innumerable settings. Those of us who are educators, for example, need to continually reflect upon what is working or not working in our teaching, how we can improve what we are doing, how we can help our students make connections to what they are learning, and much, much more. Thus it is critical for students to learn to effectively reflect upon their learning and growth.

As many of you have probably encountered, when students are first asked to respond to prompts such as "I selected this piece because..." they may respond with "I think it is nice." Okay, that's a start. But we would like them to elaborate on that response. The fact that they did not initially elaborate is probably not just a result of resistance or reluctance. Students need to learn *how* to respond to such prompts. They need to learn how to effectively identify strengths and weaknesses, to set realistic goals for themselves and their work, and to develop meaningful strategies to

address those goals. Students often have become dependent upon adults, particularly teachers, to evaluate their work. They need to learn self-assessment.

So, the reflection phase of the portfolio process should be ongoing throughout the portfolio development. Students need to engage in multiple reflective activities. Those instances of reflection become particularly focused if goal-setting is part of their reflection. Just as instruction and assessment are more appropriately targeted if they are tied to specific standards or goals, student identification of and reflection upon strengths and weaknesses, examples of progress, and strategies for improvement will be more meaningful and purposeful if they are directed toward specific goals, particularly self-chosen goals.

Did u know? As a skill, reflection is not something that can be mastered in one or two attempts. Developing good reflective skills requires instruction and modeling, lots of practice, feedback and reflection.

Self Assessment

2. *State whether the following statements are 'true' or 'false'*

   (*i*) Portfolio assessment is given at periodic level.

  (*ii*) The most Common portfolio reflection task is the completion of a sheet to be attached to the sample of work which the reflection is addressing.

 (*iii*) Teacher need not a plan for holding portfolio Conferences.

 (*iv*) Portfolio assessment does not help in student self evaluation, reflection, and critical thinking.

23.9 Summary

- A student portfolio is a systematic collection of student work and related material that depicts a student's activities, accomplishments, and achievements in one or more school subjects.

- Identifying specific goals or purposes for assigning a portfolio is the first and most critical step in creating such an assig. Just as identifying a standard guides the rest of the steps of developing an authentic assessment, identifying the purpose(s) for a portfolio influences all the other decisions involved in producing a portfolio assignment.

- Process and product portfolios represent the two major types of portfolios. A process portfolio documents the stages of learning and provides a progressive record of student growth. A product portfolio demonstrates mastery of a learning task or a set of earning objectives and contains only the best work

- First, the teacher and the student need to clearly identify the portfolio contents, which are samples of student work, reflections, teacher observations, and conference records. Second, the teacher should develop evaluation procedures for keeping track of the portfolio contents and grading the portfolio...

- Advantages of Portfolio Assessment

  • Promoting student self-evaluation, reflection, and critical thinking.

  • Measuring performance based on genuine samples of student work.

  • Providing flexibility in measuring how students accomplish their learning goals.

- Enabling teachers and students to share the responsibility for setting learning goals and for evaluating progress toward meeting those goals.

- Requiring extra time to plan an assessment system and conduct the assessment.

- Gathering all of the necessary data and work samples can make portfolios bulky and difficult to manage.

- Developing a systematic and deliberate management system is difficult, but this step is necessary in order to make portfolios more than a random collection of student work.

- Portfolios typically are created for one of the following three purposes : to show growth, to showcase current abilities, and to evaluate cumulative achievement. Some examples of such purposes include

- Growth Portfolios

  (a) to show growth or change over time

  (b) to help develop process skills such as self-evaluation and goal-setting

  (c) to identify strengths and weaknesses

  (d) to track the development of one more products/performances

- Showcase Portfolios

  (a) to showcase end-of-year/semester accomplishments

  (b) to prepare a sample of best work for employment or college admission

  (c) to showcase student perceptions of favorite, best or most important work

- Evaluation Portfolios

  - Process : One of the greatest attributes of the portfolio is its potential for focusing on the *processes* of learning.

- Steps of process

  - selection of contents of the portfolio;

  - reflection on the samples of work and processes;

  - conferencing about the contents and processes.

- Selection of Contents

  - Once again, identifying the purpose(s) for the portfolio should drive the selection process. As listed in the tables above, different samples of student work will likely be selected for different purposes. Additionally, samples are selected might also differ depending on the purpose.

  - External audiences are most likely to play a role for evaluation portfolios.

  - Reflectionons and samples of work

  - Many educators who work with portfolios consider the reflection component the most critical element of a good portfolio. Simply selecting samples of work as described above can produce meaningful stories about students, and others can benefit from "reading" these stories. But the students themselves are missing significant benefits of the portfolio process if they are not asked to reflect upon the quality and growth of their work.

  - Most common portfolio reflection task is the completion of a sheet to be attached to the sample (or samples) of work which the reflection is addressing.

- Reflection as a process skill

- Good skill development requires four steps :

- Instruction and modeling of the skill;

- Practice of the skill;

- Feedback on one's practice;

- Reflection on the practice and feedback.

23.10 Keywords

Growth : The process of growing physically, mentally or emotinally

Showcase : an event that presents somebody's abilities or good qualities

Evidence : The facts, signs or objects that make us believe that something is true.

23.11 Review Questions

1. What are the aims of portfolio assessment ?

2. Explain the process of portfolio assessment.

3. Give the types of portfolio.

4. Explain the advantages and limitations of portfolio assessment.

Answers : Self Assessment

1. (*i*) Portfolio   (*ii*) product portfolio

  (*iii*) process portfolio (*iv*) growth portfolio.

2. (*i*) T   (*ii*) T   (*iii*) F   (*iv*) F

23.12 Further Readings

Books

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 24 : Question Bank

CONTENTS

Objectives

The objectives of this unit can be summarised as below :

- To discuss the meaning of question bank
- To explain about the planning a question bank
- To describe about the item review
- To explain about the question bank in action
- To describe about the advantages and disadvantages of question bank.

Introduction

There is a persistent need to revise tests with some frequency to avoid giving the similar test year after year. Creating new set of question papers is a very tedious and time consuming job. This exercise when done manually requires the paper setter to compose questions for the test, determine their difficulty level, determine course coverage etc. so as to ensure that the test will neither be too hard nor too easy. By using computers for question banking the steps for the above processes are by passed. We will discuss about question bank in this unit.

24.1 Meaning of Question Bank

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

24.2 Planning a Question Bank

Planning for a question bank involves defining processes for preparation of individuals, preparatory work for the question bank and identifying what has to be established with the

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

24.3 Item Review

Review, editing and revalidation of items/questions submitted by the item-writers should be done in presence of item writers under the guidance of content and evaluation specialists. Generally the target of number of questions/items per course at the optimum level is taken as 10 times the total number of questions/items to be taken in the question paper. There is no upper limit to question bank size. Each question or item to be deposited in QB for a particular course must be well characterised in terms of the following aspects :

- The Block/Section/Chapter and unit number of the book

- Type of item and sub-type falls under it

- Estimated level of difficulty from the point of view of average learner

- Maximum marks it should carry

- Time (in Minutes) required for answers

- The synoptic points of answer and marking scheme for supply type questions and 'key' answers for selection type questions, and

- The level of educational objectives it is intended to possibly test (in the Taxonomy hierarchy)

Item-writer should indicate on an item-sheet, all the points which would be finalised in the editing process. Questions/items available for different courses should essentially pass through the revalidation process particularly content and evaluation-edition at the time of their being used for paper setting. Post-validation of questions/items used in evaluation involves determining the descriptive statistics such as mean, standard deviation, marks distribution, standard error of measurement of marks secured by the students (population/sample) in the examination of each course and item analysis (you will read it in detail in next section). Item analysis is the statistical study of the performance in each question by the student group as against the performance in question paper as a whole. This is a necessary feedback for the future improvement of the question bank. This can be achieved with the help of a computer and using standard method of analysis.

Descriptive statistics from the test analysis is extremely useful in making decisions such as pass/fail, grace marks for borderline cases, grading and so on. The results of item analysis are essentially important for improvement of question banks, for deciding on the reuse of 'good' questions for future examination and improvement or rejection of poorly functioning questions.

> *Notes* A question bank can store as many questions as possible so that generation of randomised tests is done without any difficulty. Question banking thus provides substantial savings of time and energy over conventional test development.

24.4 Question Bank in Action

A particular school wants to use a computerised question bank to generate its question paper for the term-end exam. It also wishes to have all types of questions in the paper. It wants to have a mix of essay type questions, short answers type questions and objective type questions in the question paper. Three sections are therefore incorporated in the paper each with two questions. Also, the Principal is very conscious of the educational objectives. How would they go about it ? A teacher who is a computer enthusiast has suggested the following method. The procedure for establishment of such a system requires a few steps such as preparing computerised question banks, doing test and item analysis while keeping taxonomy of educational objectives in mind and lastly, devising methods to construct various sets of question paper by simulation techniques using Random number generators. We hereby briefly discuss them.

> *Did u know?* Question banks also helps in providing a platform for discussing curriculum goals and objectives. The items put in the question banks can be made to inherit properties like common mistakes made by the students, their capabilities and incapabilities etc. This provides a way to discuss possible learning hierarchies and ways to better structure curriculum.

Taxonomy of educational objectives

A number of educators have divided the learning outcome into three non-overlapping domains – cognitive, affective and psychomotor. Theses efforts were mainly done because of finding some kind of relevant test items. We will concentrate on the first two domains.

DOMAINS			
Cognitive (Bloom, 1956)		**Affective (Krathwohle, 1964)**	
Knowledge	Remembering of Information	Interest	Level of involvement receiving
Comprehension	Relating of concepts	values	Responding valuing
Analysis	gathering and Arranging facts	attitudes	organisation
Evaluation	Judgement using Criteria remembered		Characterisation by value

Cognitive and affective domains

The cognitive taxonomy, has received the most attention from test constructors because it has been able to define the intellectual level at which instructional objectives and test items are written. So while making computerised question banks efforts would have to be done to translate

these educational objectives into test items. This is essential because the educational objectives may differ from program to program. We have intentionally left the discussion of the third domain i.e. psychomotor. Psychomotor domains deals with behaviours like gross bodily movements, finely co-ordinated bodily movements, non-verbal communications and speech behaviours which are more suited for the practical components of the exam.

Computerised question banks

One essential activity for the "On demand" examination system is the preparation of question banks. For this, let us try to relate the type of questions with the educational objectives. The table below shows the domains of learning outcomes and the question type, which will measure them effectively.

Table : Taxonomy of educational objectives vs. question types

Domains	Question Type
A : Cognitive	
1. Knowledge	True-False type, Multiple Choice,
2. Comprehension	Matching, Short Answer type
3. Application	
4. Analysis	Essay type
5. Evaluation	
B : Affective	
1. Interests	
2. Values	Essay type
3. Attitudes	

So, one can easily classify the question paper in three sections i.e. very short answer types, short answer types and long answer types. These sections will comprise of the following.

Very Short answer type $\longrightarrow$ True — False type Multiple choice, Matching

Short answer type $\longrightarrow$ Short questions

Long answer type $\longrightarrow$ Essay type questions

So, while preparing question banks one has to take care of these three sections. An ideal question bank will have all three of questions and when one attempts to draw a set of question paper from it, the only thing he has to decide is the number of questions to be taken from each section.

Using simulation and random number generators

Parallel sets can be made very easily with the help of random number generators. Arithmetic methods, particularly suited for computer operation are available for generating random numbers. The most common method is with the help of Linear Congruence Generator.

$$Z_1 = \left(aZ_{i-1} + c\right) \mod m, \ a,c,m,Z_0 \text{ are } + ve \text{ integers and } a,c,Z_0 \leq xm$$

The choice of numbers, depends upon the seed, which initialise the recursive computation. To make the calculations simple and to maintain a uniformity we have a set of 20 random numbers taken from Uniform (0,1) distribution on six columns.

Question Paper						
Sets	Section I		Section II		Section III	
	Q.1	Q.2	Q.3	Q.4	Q.5	Q.6
1.	0.270	0.280	0.276	0.015	0.938	0.091
2.	0.756	0.031	0.523	0.939	0.371	0.392
3.	0.589	0.923	0.483	0.426	0.441	0.107
4.	0.656	0.453	0.475	0.370	0.160	0.441
5.	0.061	0.629	0.386	0.290	0.091	0.169
6.	0.665	0.684	0.264	0.256	0.472	0.131
7.	0.213	0.502	0.002	0.021	0.096	0.472
8.	0.683	0.852	0.326	0.732	0.076	0.081
9.	0.443	0.608	0.797	0.584	0.569	0.165
10.	0.482	0.045	0.688	0.108	0.175	0.685
11.	0.297	0.674	0.122	0.519	0.958	0.649
12.	0.202	0.371	0.312	0.964	0.860	0.147
13.	0.251	0.444	0.102	0.587	0.577	0.470
14.	0.423	0.573	0.794	0.498	0.186	0.834
15.	0.899	0.708	0.311	0.347	0.279	0.258
16.	0.382	0.522	0.954	0.386	0.967	0.326
17.	0.755	0.295	0.343	0.308	0.916	0.847
18.	0.389	0.938	0.891	0.579	0.514	0.997
19.	0.695	0.338	0.444	0.421	0.277	0.719
20.	0.898	0.306	0.576	0.655	0.656	0.974

Each column represent a question and depending on the seed, one can pick up the digits say three digits after decimal or two digits after the decimal.

These sets when constructed could be checked by doing appropriate test and item analysis on pilot basis. The beauty of this kind of approach is that one can design as many parallel sets as one wish to. The ideal method would be that the question banks are available on CD-ROM.

Caution In order to avoid this, item review process must also include a careful evaluation of the skills assessed by each question and tests must be carefully formulated.

24.5 Advantages of Question Bank

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

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

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

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

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

24.6 Disadvantages of Question Bank

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

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

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

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

We are of the opinion that it is difficult to follow the method of construction of question banks; participation of students in the construction of question banks is at all impracticable. The setting of question papers with the help of question banks have both merits and demerits, however the demerits outnumber the merits. As far as the guidance of the teachers and students is concerned, the specialists have prepared question lists. Since we have experimented with the question banks without any benefit, the need of the hour is to train the question paper setters in the construction of good questions and to distribute them according to the aims and objectives of measurement and evaluation.

Task   What is Computerised question bank ?

Self Assessment

1. *Fill in the blanks :*

   (*i*) Question banks are large _____ of suitable questions that are coded by subject area, instructional objectives measured.

   (*ii*) _____ is an essential requirement for planning a question bank.

   (*iii*) Course wise blue print for the term and examination has defined on primarily three dimensions viz. _____ , _____ and _____ .

   (*iv*) _____ deals with behaviours like gross bodily movements, finely co-ordinated bodily movements non-verbal Communications and speech behaviours.

   (*v*) Parallel sets can be made very easily with the help of _____.

   (*vi*) _____ of the students is an impracticable Concept.

24.7 Summary

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

- Planning for a question bank involves defining processes for preparation of individuals, preparatory work for the question bank and identifying what has to be established with the question bank.

- Evaluation pattern of the program has to be specified e.g. details about courses and their contents, objectives of assessment, weightage of internal and external assessment, conflation and reporting procedure. Course-wise blue-print for the term-end examination has to be defined on primarily three dimensions viz. question types, content Blocks/Units and learner abilities.

- Review, editing and revalidation of items/questions submitted by the item-writers should be done in presence of item writers under the guidance of content and evaluation specialists. Generally the target of number of questions/items per course at the optimum level is taken as 10 times the total number of questions/items to be taken in the question paper. QB for a particular course must be well characterised in terms of the following aspects :

  - The Block/Section/Chapter and unit number of the book

  - Type of item and sub-type falls under it

  - Estimated level of difficulty from the point of view of average learner

  - Maximum marks it should carry

  - Time (in Minutes) required for answers

- A particular, school wants to use a computerised question bank to generate its question paper for the term-end exam. It also wishes to have all types of questions in the paper. It wants to have a mix of essay type questions, short answers type questions and objective type questions in the question paper. Three sections are therefore incorporated in the paper each with two questions. Also, the Principal is very conscious of the educational objectives.

- Taxonomy of educational objectives

- A number of educators have divided the learning outcome into three non-overlapping domains – cognitive, affective and psychomotor. Theses efforts were mainly done because of finding some kind of relevant test items. We will concentrate on the first two domains.

- Cognitive and affective domains

- The cognitive taxonomy, has received the most attention from test constructors because it has been able to define the intellectual level at which instructional objectives and test items are written.

- Computerised question banks

- One essential activity for the "On demand" examination system is the preparation of question banks. For this, let us try to relate the type of questions with the educational objectives.

- Using simulation and random number generators

- Parallel sets can be made very easily with the help of random number generators. Arithmetic methods, particularly suited for computer operation are available for generating random numbers. The most common method is with the help of Linear Congruence Generator.

  $$Z_1 = (aZ_{i-1} + c) \bmod m,\ a,c,m,Z_0 \text{ are } +ve \text{ integers and } a,c,Z_0 \leq xm$$

- The question banks have the following merits :

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

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

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

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

- Limitations or Demerits of Questions Banks

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

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

24.8 Keywords

Item   :   a single article or object

Cognitive   :   Connected with a mental process

24.9 Review Questions

1.   What is question bank ?

2.   Explain the planning of question bank.

3.   Give the advantages and disadvantages of question bank.

4.   Discuss the steps of preparing Computerised question bank.

Answers : Self Assessment

1.   (*i*)   database   (*ii*)   Computer expertise

     (*iii*)   question types, Content units, learner abilities

     (*iv*)   psychomotor domains   (*v*)   random number generators

     (*vi*)   Participation

24.10 Further Readings

Books

1.   Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2.   Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3.   Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4.   Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 25 : Use of Computer in Evaluation

CONTENTS

Objectives

Introduction

25.1 Use of Computer

25.2 Information Technology

25.3 Information and Communication Technology (ICT)

25.4 Use of ICT In Teaching

25.5 Use of ICT in Diagnostic Testing

25.6 Use of ICT in Evaluation

25.7 Use of ICT in Psychological Testing

25.8 Use of ICT in Remedial Teaching

25.9 Summary

25.10 Keywords

25.11 Review Questions

25.12 Further Readings

Objectives

The objectives of this unit can be summarised as below

- To discuss about the use of Computer
- To explain about the Information technology and information and Communication technology
- To describe the use of Computer in teaching
- To explain the use of ICT in diagnostic Testing
- To discuss about the use of ICT in Evaluation
- To describe the use of ICT in Psychological testing and remedial teaching
- To explain the use of ICT in Developing Reasoning & thinking.

Introduction

Gurukul System of Education was in vogue in India. The main characteristics of Gurukul System were dedicated and knowledgeable teachers, individualized and learner centre teaching, and self-motivated students eager to learn. This system changed due to increase in number of students. Consequently, the number of teachers increased. Some teachers are born but rests of them have to be given rigorous training so as to develop required competency to become a teacher. Teachers have been conscious about the quality of their teaching. To enhance the quality, some teachers use teaching aids, like, charts, models - static & working, specimen, slides, etc. because teachers are given training both in preparation and use of Audio-visual Aids. It is a known fact that majority of schools do not have appropriate teaching aids related to the school content. So teachers have no facility to use A - V Aids during teaching. The use of A - V Aids get further restricted due to unmotivated persons becoming teachers. Central Government realized the need of improving quality of education through the use of Television wherein most competent teacher

teaches the topic with the help of most appropriate teaching aids. This helped in improving the quality of teaching in schools having no teacher to teach the subject, less competent teacher, schools having poor or no facility of teaching aids, etc. Programmes offered through television were produced by different State Institute of Educational Technology (SIET) in different languages. Even the Video Instructional Materials were produced and made available to teachers; still majority of schools did not make use of them. Some of the reasons were no facility of TV and VCR, no electricity, TV and VCR not in working condition, not incorporated in the time table, lack of initiation on the part of teacher and Principal, etc. Along with A - V Aids, the print media has to go a long way in improving the quality of teaching and learning. Format in which the textbooks were written was not beneficial for teachers and students. Researchers started thinking and using different Theories of Learning for developing Instructional Material.

25.1 Use of Computer

The lust for quality is still on. This is the age of INFORMATION dominated by the Digital Technology. The Digital Technology has influenced all aspects of human life. Education is not an exception. At present majority of devices are based on Digital Technology. One such device is Computer. The Computer is an electronic device that has the capacity to store, retrieve & process both qualitative & quantitative information fast and accurately. The computers were never developed for improving quality of teaching - learning process. But researchers started using Computers for teaching purpose. It gave birth to Computer Assisted Instruction (CAI), Computer Managed Instruction (CMI), Computer Based Instruction (CBI), etc. People started developing CAI for teaching different subjects at School as well as Higher Education level. The developed CAIs were compared with the Lecture Method / Traditional Method and found that the developed CAIs were significantly superior to Lecture Method / Traditional Method in teaching different subjects Hart, 1992). In spite of benefit of CAI in different aspects of learning, CAI has not entered into the Classrooms as most of the developed CAIs were not based on sound Theories of Learning. People involved in developing CAI were not having the sound base of Instructional Design. Secondly, the courses are changing, the schools also do not have sufficient computer facility, teachers are not trained in the use of CAI, etc. The use of Computers was not only for teaching but also for Psychological Testing, Evaluation; database Management, Library Manaqement, etc.

Did u know? Now the technology is in the process of change from Digital to Photon. Shortly Photonic Technology will be available for the use of the society.

25.2 Information Technology

Networking of computers gave birth to Information Technology (IT). UNESCO considered Information Technology as "Scientific, technological and engineering disciplines and management techniques used in information handling and processing, their application, computers and their interaction with men and machines, and associated social, economical and cultural matters". According to Smith & Cambell (1982), a mosaic of technologies, products and techniques have combined to provide new electronic dimensions to information management. This mosaic is known by the name of Information Technology. OECD (1987) treated Information Technology as "a term - used to cover technologies used in the collection, processing and transmission of information. It includes micro-electronic and info-electronic based technologies incorporated in many products and production processes and increasingly affecting the service sector. It covers inter alias computers, electronic office equipment, telecommunication, industrial robot and computer controlled machine, electronic components and software products."

Darnton and Giacoletto (1992) defined IT as the systematic study of artifacts that can be used to give form or description to facts in order to provide meaning or support for decision making,

and artifacts that can be used for the organization, processing, communication and application of information. IT as use processing, communication, diffusion and sharing of information for social, economical and cultural upliftment.

The Information Technology leads to development of Websites. Government, Corporate sector, educational institutions, etc. started uploading the information on their websites. It provides facilities for Chat, e-mail, surfing, etc. It opens up a new source of Information which increased the limitation of access to information. Prior to IT, people were using only the print material for searching the information. It limited the search. This limitation has been overcome by the IT.

25.3 Information and Communication Technology (ICT)

IT was limited only to the textual mode of transmission of information with ease and fast. But the information not only in textual form but in audio, video or any other media is also to be transmitted to the users. Thus, the ICT = IT + Other media. It has opened new avenues, like, Online learning, e-learning, Virtual University, e-coaching, e-education, e-journal, etc. Third Generation Mobiles are also part of ICT. Mobile is being used in imparting information fast and cost effective. It provides e-mail facility also. One can access it anywhere. It will be cost effective. The ICT brings more rich material in the classrooms and libraries for the teachers and students. It has provided opportunity for the learner to use maximum senses to get the information. It has broken the monotony and provided variety in the teaching - learning situation. The ICT being latest, it can be used both at school and higher education levels in the following areas :

- Teaching
- Diagnostic Testing,
- Evaluation
- Psychological Testing
- Development of Virtual Laboratory
- Online Tutoring
- Development of Reasoning & Thinking
- Instructional Material Development

25.4 Use of ICT In Teaching

Teaching at School as well as Higher Education, mostly, concentrates on giving information which is not the sole objective of Teaching. Along with giving information, the other objectives are :

- developing understanding and application of the concepts
- developing expression power
- developing reasoning and thinking power
- development of judgment and decision making ability
- improving comprehension, speed and vocabulary
- developing self-concept and value clarification
- developing proper study habits
- developing tolerance and ambiguity, risk taking capacity, scientific temper, etc.

With the present infrastructure, class size, availability of teachers, quality of teachers, training of teachers etc., It is difficult to achieve all the objectives. Further, most of the teachers use Lecture Method which does not have potentiality of achieving majority of above mentioned objectives. The objectives are multi-dimensional in nature, so for their achievement multiple methods should be used in an integrated fashion. At present ICT may be of some use. It is a well known fact that

not a single teacher is capable of giving up to date and complete information in his own subject. The ICT can fill this gap because it can provide access to different sources of information. It will provide correct information as comprehensive as possible in different formats with different examples. ICT provides Online interaction facility. Students and teachers can exchange their ideas and views, and get clarification on any topic from different experts, practitioners, etc. It helps learners to broaden the information base. ICT provides variety in the presentation of content which helps learners in concentration, better understanding, and long retention of information which is not possible otherwise. The learners can get opportunity to work on any live project with learners and experts from other countries. The super highway and cyber space also help in qualitative improvement of Teaching -Learning Process. ICT provides flexibility to learners which is denied by the traditional process and method. Flexibility is a must for mastery learning and quality learning.

Notes   On INTERNET many websites are available freely which may be utilized by teachers and students for understanding different concepts, improving vocabulary, developing Reasoning & Thinking, etc. ICT can help in preparing students for SAT, GRE, TOEFL, etc.

25.5 Use of ICT in Diagnostic Testing

The common observation is that the quality of teaching in the classroom is on the decline. More and more students are depending on the private tutorial classes. The private tuition also has become a business. This phenomenon is not only in India but In other countries too. There are about 800 students from USA who have enrolled themselves for Private tuition in Mathematics. It means tuitions are also being outsourced. This is being done through the use of ICT. There are students who fail to understand certain concepts or retain certain information. This can be assessed by introducing the diagnosis in the process of teaching - learning. Today, this is not being done. The reasons might be large class size, non-availability of diagnostic tests in different subjects, lack of training, money and desire on the part of teacher, etc. This is the age of technology. These difficulties can be easily over come with the help of ICT. Sansanwal (2005) developed Computer Based Diagnostic Testing in Mathematics and Sansanwal and Lulla (2007) developed Computer Based Diagnostic Testing in Chemistry. Both these were tried out in CBSE affiliated schools situated in Indore. These developed Computer Based Diagnostic Tests work well and helped the teachers as well as students in identifying the gray area of each and every student. This can be put on the website of the school and the student can access it from home also. The student can prepare the topic / chapter and can take the test to find exactly what he has not understood ? The teacher cannot do this manually. The student progress can be monitored and his performance can be improved. This will develop confidence in students and may change their attitude towards the subject. It may also help in reducing the suicidal tendency among students. Students may start enjoying learning. Further, the following are the main advantages of Computer Based Diagnostic Test.

- They do not require any special setting or arrangement. The only requirement is computer systems and software.
- The student can use it even from home if made available on school website.
- They do not need any special assistance from teacher. Unlike the paper-pencil test, it does not require paper setting and paper correction on the part of the teacher.
- It saves time on the part of the teacher and students.

- The feedback is given immediately after the test is over, which gives an intrinsic reinforcement to the student.
- The student finds it more interesting and motivating as compared to the paper-pencil diagnostic test.
- It can be updated from time to time.
- It is economical in terms of money as it requires only one time investment.

There are some limitations of Computer Based Diagnostic Testing (CBDT). These are as follows :

- The learner might find it uninteresting or monotonous as compared to paper pencil test.
- The teacher might find CBDT difficult to administer if he / she is not a computer savvy.
- It faces certain constraints, like, power cut, when it is being administered.
- The learner might not take it seriously as he / she is used to the traditional paper and pencil tests.
- The development of CBDT is costly and tedious as compared to paper and pencil test.
- The use of CBDT requires many computers which may not be available in all the schools.
- The learners who are not computer friendly might not feel at ease while giving the test on Computer.
- Certain technical problems might crop up which can distract the learner while giving the test.
- All teachers may not be competent to develop diagnostic test and especially CBDT.
- Teacher may not know computer languages that may be used for developing CBDT.

25.6 Use of ICT in Evaluation

At present the paper pencil tests are conducted for evaluating the academic performance of students. These tests are conducted in the group setting. The content coverage is poor and students cannot use them at their own. These tests are evaluated by the teachers and they may not give feedback immediately to each and every student. It may be due to this that students are unable to know their weakness and do not make any attempt to improve upon them. The ICT can be made use in the evaluation. One such attempt has been made by Sansanwal and Dahiya (2006) who developed Computer Based Test in Research Methodology and Statistics. It has been titled as Test your Understanding : Research Methods and Statistics. This test can be used by individual student to evaluate his learning. The student can instantaneously get the feedback about the status of his understanding. If the answer is wrong, he even can get the correct answer. It goes a long way in improving the learning and teacher has no role to play in it. It is left up to students to use it. Such tests can be uploaded on the website for wider use. The students from other institutes can also make use of it. Not only the students even the teachers can also use it to assess their own understanding of the subject. If used by teachers before teaching the topic, they can prepare the topic properly. Such software can be used for internal assessment. Thus, ICT can be used to improve the quality of pre as well as in-service teacher's training.

Task   How does CBDT helps in education

25.7 Use of ICT in Psychological Testing

There are individual differences. Through research some correlates of academic achievement have been studied. Rarely this information is used by school / college teachers. Many of them even do not know about such researches. Even if they know, they do not make use of it at the time of forming the groups for different academic courses.

25.8 Use of ICT in Remedial Teaching

Once the ICT is used for diagnosis purpose, the next step is to organize Remedial Teaching Programme. The Remedial Teaching can be done by the teacher if some common mistakes are identified. It may not be feasible to organize Remedial programme for individual students. At this point, the ICT can be used for giving individual Remedial Programmer. It may be Online or off line. The instructional material if designed specifically for meeting the individual needs of students and uploaded on the School website and then the ICT can be used for providing Remedial teaching Programme.

Self Assessment

1. *Fill in the blanks :*

    (*i*) _____ provides online interaction facility students and teachers can exchange their ideas.

    (*ii*) The use of _____ is Costly and tedious, but it requires only one time investment.

    (*iii*) Computer based test can be used by individual student to _____ his _____.

    (*iv*) ICT can provide access to different sources of _____ , because it will provide correct information as comprehensive as possible in different formats.

25.9 Summary

- At present majority of devices are based on Digital Technology. One such device is Computer. The Computer is an electronic device that has the capacity to store, retrieve & process both qualitative & quantitative information fast and accurately. The computers were never developed for improving quality of teaching - learning process. But researchers started using Computers for teaching purpose. It gave birth to Computer Assisted Instruction (CAI), Computer Managed Instruction (CMI), Computer Based Instruction (CBI), etc. People started developing CAI for teaching different subjects at School as well as Higher Education level.

- Networking of computers gave birth to Information Technology (IT). UNESCO considered Information Technology as "Scientific, technological and engineering disciplines and management techniques used in information handling and processing, their application, computers and their interaction with men and machines, and associated social, economical and cultural matters".

- The Information Technology leads to development of Websites. Government, Corporate sector, educational institutions, etc. started uploading the information on their websites. It provides facilities for Chat, e-mail, surfing, etc. It opens up a new source of Information which increased the limitation of access to information.

- IT was limited only to the textual mode of transmission of information with ease and fast. But the information not only in textual form but in audio, video or any other media is also to be transmitted to the users. Thus, the ICT = IT + Other media. It has opened new avenues, like, Online learning, e-learning, Virtual University, e-coaching, e-education, e-journal, etc. Third Generation Mobiles are also part of ICT.

- The ICT being latest, it can be used both at school and higher education levels in the following areas :

    - Teaching

    - Diagnostic Testing,

    - Evaluation

- Psychological Testing

- Teaching at School as well as Higher Education, mostly, concentrates on giving information which is not the sole objective of Teaching. Along with giving information, the other objectives are :

  - developing understanding and application of the concepts

  - developing expression power

  - developing reasoning and thinking power

  - development of judgment and decision making ability

- With the present infrastructure, class size, availability of teachers, quality of teachers, training of teachers etc., It is difficult to achieve all the objectives. Further, most of the teachers use Lecture Method which does not have potentiality of achieving majority of above mentioned objectives.

- The ICT can fill this gap because it can provide access to different sources of information. It will provide correct information as comprehensive as possible in different formats with different examples. ICT provides Online interaction facility. Students and teachers can exchange their ideas and views, and get clarification on any topic from different experts, practitioners, etc.

- The common observation is that the quality of teaching in the classroom is on the decline. More and more students are depending on the private tutorial classes.

- Further, the following are the main advantages of Computer Based Diagnostic Test.

  - They do not require any special setting or arrangement. The only requirement is computer systems and software.

  - The student can use it even from home if made available on school website.

  - They do not need any special assistance from teacher. Unlike the paper-pencil test, it does not require paper setting and paper correction on the part of the teacher.

  - It saves time on the part of the teacher and students.

- There are some limitations of Computer Based Diagnostic Testing (CBDT). These are as follows :

  - The learner might find it uninteresting or monotonous as compared to paper pencil test.

  - The teacher might find CBDT difficult to administer if he / she is not a computer savvy.

  - It faces certain constraints, like, power cut, when it is being administered.

  - The learner might not take it seriously as he / she is used to the traditional paper and pencil tests.

- There are individual differences. Through research some correlates of academic achievement have been studied. Rarely this information is used by school / college teachers.

25.10 Keywords

Information	:	Facts or details about somebody/something.
Technology	:	Scientific knowledge used in practical ways in any field.
Communication	:	The activity or process of expressing ideas and feelings or of giving people information

25.11 Review Questions

1.   What is the use of Computer in teaching ?
2.   How does Information and Communication technology helps in evaluation ?
3.   What is the use of ICT in diagnostic testing ?
4.   Give advantages and limitations of Computer based diagnostic testing.

Answers : Self Assessment

1.   (*i*)   Information and Communication technology
      (*ii*)   Computer based diagnostic testing
      (*iii*) Evaluate, learning
      (*iv*)   Information

25.12 Further Readings

Books

1.   Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*
2.   Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*
3.   Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*
4.   Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 26: Achievement Test: Concept, Types and Construction

CONTENTS

Objectives

Introduction

26.1 Concept of Achievement Test

26.2 Aims of Achievement Test

26.3 Types of Achievement Test

26.4 Construction of Achievement Test

26.5 Significance of Achievement Test

26.6 Summary

26.7 Keywords

26.8 Review Questions

26.9 Further Readings

Objectives

The objectives of this unit can be summarised as below :

• To explain the Concept of Achievement test

• To discuss about the aims of Achievement test

• To describe about the types of Achievement test

• To explain about the construction of Achievement test

Introduction

As a teacher one is involved directly in the evaluation of the learner. The theory that you have learnt will have to be applied by you as a teacher in the classroom situation. It 'Learner's Evaluation in general and this unit in particular is concerned with this very important activity of teachers.

Teachers teach and help the learners to learn. The learning that takes place is assessed or evaluated not only for the learner's benefit but also for the teacher to evaluate his/her own work. At the end of a lesson or a group of lessons, the teacher needs to get feedback on what the learner has achieved, as a result of the teacher's efforts and also, indirectly to assess his/her own achievement as a teacher. This feedback comes with the help of a tool, generally an achievement test. An achievement test is designed to evaluate a unit during the teaching-learning process. The unit of teaching-learning may be, as has already been mentioned, one lesson or a group of lessons transacted in a particular time period. You have already read about achievement tests in the previous unit. In this unit we will discuss the same in detail.

26.1 Concept of Achievement Test

Teaching work in schools is carried out in a controlled manner. Under it, the teachers try to find out after a certain interval, how much the students have learnt. The tests which are prepared to find it out, are called achievement tests. In the words of **Ebel** :

An achievement test is one designed to measure a student's grasp or knowledge of his proficiency in certain skills.
— *Ebel*

At present, the teaching objectives are determined as changes in the cognitive, conative and affective domains of the students. So, the achievement tests should be defined in the following way :

The tests which are constructed to measure cognitive, conative and affective changes occurring as a result of teaching school subjects, are called achievement tests.

26.2 Aims of Achievement Test

The achievement tests are used in schools for the following functions :

(1)   To select the students for admission.

(2)   To classify the students.

(3)   To encourage the students.

(4)   To promote the students.

(5)   To issue certificates on the basis of public examinations and results.

(6)   To evaluate educational problems.

(7)   To evaluate teaching work of the teachers.

26.3 Types of Achievement Test

The common meaning of a test is to examine the traits of an object, person or activity. But in the field of education, tests are meant by those tools and methods by which the measurement of mental abilities and educational achievement of the students is carried out. These tests have their own definite form. In these tests questions, related to the mental ability or educational achievement to be measured, are asked which the students reply or related problems are presented to which the students respond. The mental abilities and educational achievement of the students are measured on the basis of these answers and responses. The main mental ability tests are—intelligence tests, aptitude tests and personality tests; and the main educational ability tests are — achievement tests and diagnostic tests. We can define educational tests briefly in the following words :

Educational tests are those tools and methods of measurement of various abilities of the students in which related questions about measurable ability are asked and of which the students have to answer and related problems are presented to which the students have to respond.

Classification of Educational Tests

The tests being used in the field of education are classified in different forms :

(I)   Classification on the Basis of the Field of the Tests

On the basis of measurement field, the tests can be divided into two classes as follows :

1.   Educational Tests : Educational tests are those tests by which educational achievements and educational problems of the students are measured; such as—achievement tests and diagnostic tests. They have been discussed in the related chapter in detail.

2.   Psychological Tests : Psychological tests are those tests by which mental abilities of the students are measured; for example—intelligence tests, aptitude tests and personality tests. They have been discussed in the related chapters in detail.

(II)   Classification on the Basic of Nature of the Tests

On the basis of the nature of the tests, they have been divided into three classes—oral, written and practical.

1. **Oral Tests :** Those tests in which the questions from the examinees are asked orally and the examinees have to answer orally are called oral tests. Such tests are generally comprised of short-answer and objective-type questions.

   The aim of teaching a subject is not limited to impart knowledge of the related facts, it is also taught to bring about a change in the thinking and working style of the students. And its measurement is done by oral tests. These tests are used for ascertaining their knowledge and practical use of linguistic abilities (pronunciation, lecture, expression style), their confidence and their knowledge of other subjects. The chief characteristic of these tests is that the examiner and the examinee are face-to-face with each other, so the examiner gets an opportunity to know the personal qualities of the examinees.

   These tests are subject to some shortcomings as well. First, the examiner can test only one student at one time which requires much energy and time. Second, these tests are subjective in nature and marking depends upon the examiner to a great extent. Third, change of behaviour as a whole, as a result of teaching of various subjects cannot be measured correctly. Fourth, the practical aspects of art and science subjects cannot be measured by these tests. Fifth and the last, these tests greatly depend on the ability and honesty of the examiners.

   It is evident that these tests should be made use of for measurement of students' pronunciation skill, linguistic ability, fearlessness, confidence and general knowledge of various subjects and their practice in life. But these tests can be beneficial only if they are executed honestly. Moreover, these tests should be used only in those subjects where it is necessary.

2. **Written Tests :** The tests in which the examinees are given questions in written form and which they respond in writing, are called written tests. In these tests any types of questions can be asked as essay-type, short-answer-type and objective-type. So it is necessary to discuss them independently which we have done in the succeeding paragraphs.

3. **Practical Tests :** The tests in which the measurement of skills and activities of the students is executed by asking them to display the skill or activity are called practical tests; such as asking a child to sing a song in order to measure his proficiency in singing. Practical tests are used for the measurement of practical skills and knowledge in art, music, science and other practical subjects.

   The chief characteristic of practical tests is that the practical aspect of the examinees is measured. These tests alone can test the practical knowledge of arts and science. If a doctor or engineer is not tested as such, a flaw in his theoretical knowledge may create havoc.

   But these tests have their own limitations. First, we cannot correctly measure the theoretical knowledge and emotional aspect of a person by these tests. Second, their execution needs comparatively more energy and time. Moreover, these too have been embraced in the ambit of immorality.

   However, the practical skills can be tested only by practical tests. The need is of course to improve these tests. There is only one advice worth giving—the examiners should work honestly.

(III) Classification on the Basis of Material Used

The tests are classified into two types on the basis of material used in them, which are :

1. **Verbal Tests :** This type of tests includes those tests in which words or language are used which may be written or oral. Such tests are mostly used in the field of education.

2. **Non-Verbal Tests :** This type of tests includes those tests in which words or language are not used, rather scenery, symbols, signs, pictures, etc are made use of these tests are used

to measure mental abilities of children and illiterate persons. These tests are also used in the measurement of intelligence and personality of all persons whether literate or illiterate.

(IV) *Classification on the Basis of the Construction of the Tests*

Tests are divided into two types on the basis of their construction process and characteristics:

1. **General Tests :** This type of test includes those tests which are constructed by the teachers, so these are also called Teacher-Made Tests.

   These tests are mostly used in the field of education, in all weekly, fortnightly, monthly, quarterly, half-yearly and annual examinations. Though attempts are being made to make these tests valid, reliable and objective, yet nothing much can be said with certainty about their validity, reliability and objectivity. Since fully valid, reliable and objective tests cannot be constructed for general tests, so much emphasis is laid on making these tests more and more valid, reliable and objective.

2. **Standardized Tests :** This type of tests includes those tests which are constructed by specialists in a specific way. They prepare a draft according to the measurable objectives, then they administer it on different groups of students, and then they delete unnecessary material from these tests and add necessary material to them. Thus, they make them valid, reliable and objective.

   Efforts are on to make the achievement tests more and more valid, reliable and objective, but success has not been tasted so far.

   But it should not be taken to mean that we should not standardize the achievement tests. We should keep endeavouring in this direction so that they are made more and more valid, reliable and objective. The standardisation of tests has been discussed in an independent chapter.

(V) *Classification on the Basis of Administration of the Tests*

The tests are also classified into two types on the basis of their administration :

1. **Individual Tests :** This type of tests includes those tests which are administered on one student at one time. Oral tests are generally administered individually. Some intelligence tests are also administered individually. The chief quality of these tests is that the attention of the measurer is wholly turned to the particular student, though it needs more time, energy and money. So these tests can be administered only in necessary situations.

2. **Group Tests :** This type of tests includes those tests which can be administered on a group of student at one time. Written tests are generally administered collectively. The chief characteristic of these tests is that a large number of students can be tested by these at one time, it saves time, energy and money. However, there is a shortcoming in these and, that is, the problem of a particular student cannot be paid attention to, for which individual tests have to be administered.

(VI) *Classification on the Basis of the Form of Measurement*

American psychologist, **Robert Glaser**, has classified measurement into two types—Norm referenced measurement and Criterion referenced measurement. He called the traditional measurement, by which the relative ability of a student in a subject is measured, as norm referenced measurement; and the measurement by which the real knowledge of a student in a subject is measured as criterion referenced measurement. On this basis, some scholars have divided the tests into two types :

1. **Norm Referenced Tests :** This type includes those tests which measure norm referenced measurement, that is they only measure the relative position of a student in a subject in group. The traditional essay-type tests, in which 5-6 questioned are to be answered out of 10-12 questions, fall in this type.

The chief characteristic of this type of tests is that by these tests the knowledge of the students pertaining to the subject as well as their insight can be measured, if the tests are valid, reliable and objective. Execution and evaluation of such tests is also easier.

However, since the questions are not asked on the whole subject matter, so it is not possible to measure the whole knowledge about the subject.

At present, efforts are being made to improve upon the norm referenced tests, and much success has been achieved in the direction of correct measurement and evaluation.

2.   **Criterion Referenced Tests :** Those tests are included in this type which commit criterion referenced measurement, that is, they measure the real condition of a student's ability in a particular subject in a group. However, Glaser has not clarified the meaning of criterion. As a result, different scholars use it in different meanings. Some scholars consider the meaning of criterion to be by contents. Their argument is that the ability of a student is not correctly measured by norm referenced tests, as such questions should be asked on the whole curriculum. Such scholars call criterion referenced tests as **Content Reference Tests** or **Ability Referenced Tests**. But most of the scholars attribute the meaning of criterion to teaching objectives and they take the criterion referenced tests as **Objective Referenced Tests**. Some educationists define teaching objectives in the form of behavioural change. They classify behavioural change into three domains : Cognitive domain, Affective domain and Psychomotor domain. These educationists take criterion referenced tests in the form of **Domain Referenced Tests**. We are of the view that the term—Objective Referenced Tests is the most suitable one. It includes all the three domains—cognitive, affective and psychomotor domains.

The chief characteristic of these tests is that they can measure the whole knowledge of a student in a subject if care is taken in their construction about their validity, reliability and objectivity. These tests, help to know how much the students have learnt and how much not, and accordingly planning can be made for future.

However, construction, administration and evaluation of these tests are difficult, execution needs special training. At present, much emphasis is given on Bloom's taxonomy in our country. We are, perhaps, not aware that it is now considered meaningless in America.

So it should suffice to say that we should divide the educational objectives into cognitive, affective and psychomotor domains, and , construct objective based tests to measure the achievement of the objectives for which teaching was conducted. Efforts are going on in this field at present; short-answer-type questions and objective-type questions are asked along with essay-type questions. Of course, we are not in favour of objective-type questions being asked along with essay-type and short-answer-type questions. Objective-type questions should be asked in independent tests.

Did u know? The psychologists have constructed some standardized tests for measurement of psychological abilities—intelligence, aptitude and personality, and they claim them to be fully valid, reliable and objective. But their claim does not seem to be fully correct.

(VII) *Classification on the Basis of Types of Questions Asked in the Tests*

Tests can be classified into three types on the basis of the construction of the questions and the form of their answers—essay type, short answer type and objective type tests. These tests are chiefly used in the field of education. So we shall discuss their forms and merits and demerits in the succeeding paragraphs.

Essay Type Tests

The test in which essay type questions are asked and their answers too have to be given in essay-type, they are called essay type tests. Such tests have been in vogue since ancient times, so some

scholars term them as **Old System of Tests**. As they have been prevalent continuously and continues even today, so some scholars call them as **Traditional Tests**. The chief shortcoming of these tests is that the measurement in them depends on the viewpoint of the measurers. From this viewpoint, some scholars call them as **Subjective Tests**.

Essay type questions are of several types—Descriptive, Explanatory, Discussion, Comparison, Illustrative and Criticism, etc. Here are given the examples of such essay-type questions :

1. **Short Answer Questions**

    Short answer questions generally require exact answers and, although taking many forms, they share the following distinctive features.

    (*i*)   They usually take less than five minutes to read and answer, many take lessc than a minute.

    (*ii*) They include some guidance on the extent of the answer, many take less than a minute. space or specific instruction such as "In not more than 20 words..."

    (*iii*) The answer is supplied by the pupil, not pre-selected as in objective questions.

    They can be grouped into two broad categories :

    (a) extended answer

    (b) insert and completion

    (a)   Extended Answer Type

    The extended answer version includes questions which require pupils to write a brief description, draw a map, make a list, perform a calculation, translate a sentence, write down a definition or formula and so on. They are probably the commonest form of questions used in schools and are frequently used by examining Boards. They are deceptively easy to set and usually difficult to mark with any degree of speed and consistency.

    Examples

    1. Give the titles of two novels of Prem Chand and for each the names of three of the principal characters.

        (a) Title : _____

             Characters : (*i*) _____   (*ii*) _____   (*iii*) _____

        (b) Title : _____

             Characters : (*i*) _____   (*ii*) _____   (*iii*) _____

    2. Describe briefly two factors which contributed to the outbreak of the First World War. Do not write more than thirty words about each factor.

    3. Name the two types of chemical compound which combine to form salts, and then name one inorganic and one organic salt.

    4. Make diagrams to show the structure of

        (i)   an atom

        (ii) a human cell

        (iii) Methyl alcohol

    5. Write two separate sentences, one containing the word 'eminent' and the other containing 'imminent'.

    6. Construct

        (a) An equilateral triangle with the sides measuring 4 *cm*.

        (b) A square with 5 *cm* side.

        (c) Find out the difference between the areas of the triangle and the square you have constructed.

(b) Completion Type

The commonest form of **completion** questions is one where the pupil is required to add one or two words to complete an incomplete statement correctly. Where the missing words are in the body of the statement to be completed it is usually called an insert type.

Examples

1. In the human eye light enters the (I)....., which is surrounded by the part called the (2)....., As the amount of light increases this part (3).....but (4).....again when the amount of light decreases. On reaching the (5).....at the back of the eye the light stimulates two types of nerve cell called (6) r......and (7) c......

    (1) ....   (2) ....   (3) ....   (4)................

    (5) ....   (6) ....   (7) ....

2. Complete the missing words in this paragraph.

    That night there was so little hotel a ...tion that they had to take an expensive......of rooms.

    After paying the bill they were almost p......less.

    (A useful technique for testing vocabulary and spelling).

3. Complete the following formulae :

    Ammonia : N ......

    Sulphuric acid : H ......

    Sodium carbonate : ...... CO

    (Incompleteness of formulae can be adjusted in accordance with what is to be tested.)

4. Complete the expansion by filling in the blanks :

    $(a + b)^2 = a^2 + b^2 + 2ab$

Notes   A **completion type** is where the words are required at the end of the statement. The use of insert or completion questions is not, however, limited to written statements and can be used to prepare extremely good questions based on incomplete maps, drawings, diagrams, formulae, calculations, and the like.

---

2. Objective Type Questions

What is an objective question ? Simply, an objective question is one which is free from any subjective bias–either from the tester or the marker. Confusingly, in educational jargon, the adjective 'objective' usually means 'not subjective' while the noun 'objective' usually means an aim, a goal, target or intention. This sub-section is not about course objectives-aims, intended learning outcomes, etc–but about testing which is free from subjective elements. There can only be one right or objective answer to an objective question. Objective questions can take various forms, but invariably they require brief answers with little or no writing. A simple tick or a quick oral answer may be enough.

(i) Simple Recall

The most common used objective type question by teacher as part of their day-to-day teaching is simple recall. The teacher asks a short question, expecting a quick one-word answer or a simple statement completed. Let us have a look at some examples.

(a) Direct question : After which battle was the Mughal empire established in India ?

Expected answer : First Battle of Panipat.

(b) Incomplete statement : A writer called Jane............wrote Pride and Prejudice.

Expected answer : Austen

This, of course, is the kind we have discussed under short answer questions also. There is a definite overlapping.

Self Assessment

1. *Fill in the blanks :*

    (*i*)   The tests which are Constructed to measure Cognitive Conative and affective changes occurring as a result of teaching school subjects, are called _____.

    (*ii*) There are three types of achievement tests _____, _____ and _____ on the basis of nature of the test.

    (*iii*) Tests can be classified into _____ , _____ and _____ on the basis of the Construction of the questions.

    (*iv*) The tests in which the measurement of skills and activities of the students is executed by asking then to display the skill or activity are called _____ .

    (*v*)   The most common used objective type question by teachers as part of their day to day teaching is _____ .

26.4 Construction of Achievement Test

Equal total from the two sides shows that the division of questions and of marks has been done according to the weightage given to the teaching objectives. If this total is not equal, it can be understood that division of marks is not done accordingly. It can be checked for correction.

1. **Constructing the Preliminary Draft of the Test :** At this step, the questions are constructed according to the blue print. The precautions which should be taken while constructing questions of various types have been discussed in Chapter 2 in detail. We are repeating them here in brief :

    (1)   The essay type questions, should be according to the level of the class (as descriptive, argumentative, analytical, etc.) They should be constructed in a simple and clear language. The questions should be pointed and clear. Separate marks should be allotted to different parts of a question.

    (2)   The language of the short answer type questions too should be easy and clear. These questions should be still more pointed and objective, that is, their answers should be almost definite. Model answers to these questions should be prepared so that they are evaluated objectively.

    (3)   There are many types of objective type questions, they should be selected according to measurable objectives. The questions should be such which may compel students to recall, reason and analyse.

    (4)   Instructions should be clearly given on the test (question paper). The classification of the questions should be such that the students can be classified as high, medium and low.

    (5)   The test should be checked several times to ascertain that questions have been constructed to meet the measurable objectives, the construction of the questions is good and they are objective to a great extent.

2. **Review of the Test :** At this step, the test so prepared is reviewed by specialists on the basis of which language errors and weightage given to different parts of the question are amended.

3. **Typing the Test :** At last, the test is typed or got printed and made available to administer on the students.

Standardisation of a test signifies the process by which a general or teacher made test is made valid, reliable, objective and discriminative and for which the norms are prepared. This is done in the following 6 steps : .

1. **Construction of the Test :** At first an achievement test is constructed following the definite steps. (Refer—Construction of Achievement Tests.)

The only extra care to be taken is that the number of questions of each type should be double that of the decided one. It should also be kept in mind that some essay-type questions are such which may only be answered by 20% students of high ability, and rest of the questions can be answered by all students. The short-answer-type questions are such which may only be answered by 20% students of high ability, and rest of the questions can be answered by all students. The short-answer-type questions too are constructed in the same way. The objective-type questions are constructed such that 27% questions can be answered by the students of higher level only and 27% cab be answered by all.

2. **Selection of Questions after Testing of Preliminary Draft :** In order to know of the errors of the test, the test is administered on some students of the required level. Then the questions are suitably amended keeping in view the difficulties felt by the students and the suggestions given by them, it is called **preliminary try-out stage**.

3. **Item Analysis :** It is necessary to understand the meaning of item analysis. The questions asked in a test are called item of the test. By item analysis is meant numerical analysis of the obtained scores of each item, the analysis of its validity and reliability, and deciding its suitability accordingly. Item analysis is such a process by which some items are selected for the test, some items are cancelled and some items are improved. This is done on two bases—difficulty level of the items, and discriminating power of the items.

   (a) **Difficulty Level of Items :** The difficulty level of a question or item is known from the percentage of students who succeed to respond it correctly. The difficulty level of an item indicates whether it is easy or difficult. The number of questions or items in a test indicates the number of difficulty levels in the test independently. For selection of an item in a test, generally, the percentage of students solving it correctly is taken into consideration. The questions being solved correctly by all students of a group are easy in nature, and on the opposite, the questions which are not solved by any student are extremely difficult. Both these type of questions are not accepted in a good test. Generally, if an item of a test is solved by 50% of the students, then it is taken to be of suitable difficulty level. It does never mean that only those items are selected for the test which have been solved by 50% students. Some such items too are selected which are solved by students of higher achievement only. Generally, the items falling in the range of 30% to 70% are selected.

   Finding Out Difficulty Level of the Items : All specialists are of the opinion that the level of difficulty of an item can be found out by comparing the number of students answering it correctly with those of answering it incorrectly. Different formulea are used in different circumstances to find out the difficulty level of an item.

   First-Method : The **ratio method** is the simplest method to find out the difficulty level of an item. In this method, the ratio of the number of students answering it correctly with the total number of the students is found out. The following formula is used to find it out:

   $$DL = \frac{R}{N}$$

   Here,   DL = Difficulty level

            R = Number of students answering correctly

            N = Total number of students.

   Example

   If 110 students out of a group of total 150 students are able to solve a test item correctly, what will be the difficulty level of that item ?

   Solution

   Here,   N = 150

            R = 110

So, $\qquad$ $DL = \dfrac{R}{N}$

$\qquad DL = \dfrac{110}{150}$

$\qquad\qquad = .73$

For the sake of convenience, it can be multiplied by 100 to present it in percentage.

$\qquad DL = .73 \times 100$

$\qquad\qquad = 73\%$

Notes The more is the percentage of difficulty level, the easier is the item.

Second Method : There is a probability in answering multiple choice questions that some students may give correct answers to some question by guess. In such a case, the above formula is amended to find out the difficulty level of an item. The amended formula is as under :

$$DL = \left(R - \dfrac{W}{K-1}\right)\dfrac{1}{N}$$

Here, the value of R and N is the same as given in the above formula.

$\qquad$ W = The number of students answering an item incorrectly

$\qquad$ K = The number of alternatives for solving an item.

Example

In the above example No. 1, if the number of alternatives in the multiple choice questions is four, the difficulty level of an item will be calculated as under :

Here, $\quad$ N = 150

$\qquad\quad$ R = 110

$\qquad\quad$ W = 150 – 110 = 40

$\qquad\quad$ K = 4

So, $\qquad\qquad DL = \left(110 - \dfrac{40}{4-1}\right) \times \dfrac{1}{150}$

$\qquad\qquad\qquad = \left(110 - \dfrac{40}{3}\right) \times \dfrac{1}{150}$

$\qquad\qquad\qquad = \dfrac{290}{450}$

$\qquad\qquad\qquad = .64$

$\quad$ Difficulty level (in %) $= .64 \times 100 = 64\%$

Third Method : In the above two methods, it has been taken for granted that all the students have solved all the items in the test. Sometimes, it may be possible that the students are not able to write answers to the last few questions due to lack of time or any other reason. In such a situation, the difficulty level calculated using the first two methods cannot be considered to be correct. Such questions or items are called unattempted items Unattempted items are those which have not been solved due to lack of time or any

other reason. The items in the test which have not been attempted because of inability are not called unattempted items. Suppose there are 50 items in a test. A student solved the items 1 to 20, did not solve 21 and 22, then solved the items from 23 to 45, and did not solve any item from 46 to 50 because of lack of time or any other reason. Under this situation, the items from 46 to 50 will only be called unattempted items, and not the items at 21 and 22. The difficulty level of such item is known by using the following formula :

$$DL = \left(R - \frac{W}{k-1}\right)\frac{1}{N - NR}$$

Where   NR = The number of students who have not reached to the item.

In the above example, if 90 students have solved the item correctly, 40 students have solved it incorrectly and 10 students have not reached to it, the difficulty level of the item will be found out as follows :

$$DL = \left(90 - \frac{40}{4-1}\right)\frac{1}{150-10}$$

$$= \left(90 - \frac{40}{3}\right)\frac{1}{140}$$

$$= \frac{230}{420}$$

$$= .54$$

   Difficulty level (in %) = 54%

Fourth Method : If the number of students is large, it needs more time and labour to find out the difficulty level of each item. In such a situation, **Kelley** has suggested to divide the total students, on the basis of their obtained marks, into three classes—high, medium and low.

High class : The top 27% of the total number of students

Medium class : The middle 46% of the total number of students

Low class : The lowest 27% of the total number of students

Only the data obtained from students of high and low classes is used to know the difficulty level of an item. Kelley has suggested that the test should be administered on at least 370 students to use the formula propounded by him so that about 100 students can be included in the 'high' and 'low' classes. The formula is as follows :

$$DL = \frac{1}{2}\left[\left(R_H - \frac{W_H}{k-1}\right)\frac{1}{N_H - NR_H} + \left(R_L - \frac{W_L}{k-1}\right)\frac{1}{N_L - NR_L}\right]$$

Where, R_H = The number of students in the high class answering correctly

              W_H = The number of students in the high class answering incorrectly

              N_H = Total number of students in the high class

             NR_H = The number of students in the high class not reached

              R_L = The number of students in the low class answering correctly

              W_L = The number of students in the low class answering incorrectly

              N_L = Total number of students in the low class

             NR_L = The number of students in the low class not reached

Example

Suppose, the following data are obtained in respect of question number 27 of a test :

$R_H = 60$, $W_H = 30$, $N_H = 100$, $NR_H = 10$,

$R_L = 20$, $W_L = 50$, $N_L = 100$, $NR_L = 30$, $K = 4$

So, the difficulty level of question number 27 will be :

$$DL = \frac{1}{2}\left[\left(60 - \frac{30}{4-1}\right)\frac{1}{100-10} + \left(20 - \frac{50}{4-1}\right)\frac{1}{100-30}\right]$$

$$= \frac{1}{2}\left[(60-10)\frac{1}{90} + \left(\frac{10}{3}\right)\frac{1}{70}\right]$$

$$= \frac{1}{2}\left[\frac{50}{90} + \frac{1}{21}\right] = \frac{1}{2}\left[\frac{38}{63}\right]$$

$$= \frac{19}{63} = .30$$

Difficulty level (in %) = 30%

Note :

Though no norm has been determined as regards difficulty level for the selection of items in the construction of a test, yet most specialists consider the selection of items of different level from 30% to 70% as suitable. Some of the specialists are of the opinion that it should be limited from 40% to 60%. **Katiz** has written in this context that if only proficiency of the students in a subject is to be found out, then the test should be constructed using less difficult items; but if the students have to be divided into different classes as high, medium and low then comparatively difficult items should be used to construct a test. He has demonstrated it by the following table :

Form of Item	Difficulty level (in %)
Essay type	50
Five choice type	60
Four choice type	62
Three choice type	66
Two choice type	75

(b)   Discriminating Power of Items : Discriminating power of a question or item intimates how far it can discriminate the students of higher and lower achievement. If a question has not this quality, it is not selected for the test. If a question or item in a test is solved by students of higher and lower levels in an equal or almost equal percentage, then its discriminating power is zero and it is not considered worth for inclusion in the test. However, if the percentage of the students of higher level is sufficiently more than the students of lower level, the positive discriminating power of the question is high and such a question is selected. The determination of high and low levels is done on the basis of obtained scores of the test.

Finding out Discriminating Power of the Items : The following methods and formulae are used to find out the discriminating power of the items :

First Method : When the group of students is of small size, the number of students solving an item is known in percentage and the discriminating power is found out using the following formula :

$$DP = PQ$$

Where,   DP = Discriminating Power

   P = The percentage of students solving an item correctly

   Q = The percentage of students solving an item incorrectly

From the form of the above formula, it becomes evident that if an item is solved correctly by all students or by none of them, the value of Q and P becomes zero respectively. In such a condition, the discriminating power will be zero. The specialists are of the view that the item with discriminating power 250 is considered suitable for the test.

Example

The item number 7 of an achievement test was solved correctly by 30 students and incorrectly by 20 students in a class of 50 students. Find out the discriminating power of this item.

Solution

The number of students solving the item correctly = 30

Percentage of students solving the item correctly

$$P = \frac{30}{50} \times 100$$

$$= 60\%$$

And percentage of students solving the item incorrectly

$$Q = 100 - P$$

$$= 100 - 60$$

$$= 40\%$$

So, discriminating power

$$DP = PQ$$

$$= 60 \times 40$$

$$= 2400$$

Second Method : When the number of students in a group is large, they are arranged in descending order according to the scores obtained in the test; and after this, two groups are formed as 'high' and 'low'. Different methods for forming these two groups are used which are shown in the following table :

Class	Number of Students			
	I	II	III	IV
High	25%	27%	33%	50%
Low	25%	27%	33%	50%

According to Geret and some other psychologists, the discriminating power can be better obtained by grouping 27% students in the high group and 27% in the low group. However, in research work where the number of students is 100 or less, the classes of 50-50% give better results.

In the method of 'high' and 'low' classes, the discriminating power is found out by using the following formula :

$$DP = \frac{R_H - R_L}{N}$$

Where, N = The number of students in the high or low class

R_H = The number of students in the high class solving an item correctly

R_L = The number of students in the low class solving an item correctly

Example

A test was administered on 200 students. (Using 27% method), the item number 23 was correctly solved by 45 and 31 students respectively out of a group of 54 students each. Find out the discriminating power of this item.

Solution

$$N = 54$$

$$R_H = 45$$

$$R_L = 31$$

So,

$$DP = \frac{R_H - R_L}{N}$$

$$= \frac{45 - 31}{54} = \frac{14}{54}$$

$$= .26$$

So, the discriminating power of the item number 23 is .26

Third Method : When some of the items of the test are not attempted, the following formula is used to find out the discriminating power.

$$DP = \frac{R_H - R_L - 1}{R_T \left(1 - \dfrac{R_T}{N_T - NR_T} \right)}$$

In which, $R_T = R_H + R_L$

$N_T = N_H + N_L$

$NR_T = NR_H + NR_L$

N_H = The number of students in the high class

N_L = The number of students in the low class

NR_H = The number of students in the high class, who did not attempt

NR_L = The number of students in the low class, who did not attempt

This formula is used only when the value of R_H is more than R_L. If the value of R_H is less than R_L, the following amended formula is used :

$$DP = \frac{R_H - R_L + 1}{R_T \left(1 - \dfrac{R_T}{N_T - NR_T} \right)}$$

When the value found out using these methods is 10 or more, the discriminating power of the item is considered satisfactory. The items with less than this discriminating power are not considered suitable for the test. If the value of DP comes in negative, the item is considered very poor.

Example

The analysis of obtained marks in an achievement test for item number 23 is as follows. Find out the discriminating power of the item.

Class	Number of Students	Correct Solution	Unattempted
High	54	40	10
Low	54	25	5

Solution

$$R_H = 40, \qquad R_L = 25, \qquad N_H = 54,$$

$$N_L = 54, \qquad NR_H = 10, \qquad NR_L = 5$$

Here the value of R_H is more than R_L, so the first formula will be used.

$$N_T = N_H + N = 54 + 54 = 108$$

$$R_T = R_H + R_L = 40 + 25 = 65$$

$$NR_T = NR_H + NR_L = 10 + 5 = 15$$

So,
$$DP = \frac{R_H - R_L - 1}{R_T\left(1 - \dfrac{R_T}{N_T - NR_T}\right)}$$

$$= \frac{40 - 25 - 1}{65\left(1 - \dfrac{65}{108 - 15}\right)}$$

$$= \frac{14}{65\left(\dfrac{28}{93}\right)}$$

$$= \frac{14}{\dfrac{1820}{93}}$$

$$= \frac{14}{19.5698} = \frac{14}{4.42} = 3.17$$

So, the discriminating power = 3.17

Note :

The following formulae are used to know the difficulty level and discriminating power of the teacher-made tests respectively :

$$DL = \frac{R_H + R_L}{2N}$$

$$DP = \frac{R_H - R_L}{N}$$

It is considered suitable to select the items from .3 to .7 difficulty level and about .5 discriminating power. The decision regarding selection of each item used in the test is displayed by writing yes or no in the last column.

Item number	Number of students correctly solving		Difficulty level (DL)	Discriminating Power (DP)	Suitability
	R_H	R_L	$\dfrac{R_H + R_L}{2N}$	$\dfrac{R_H - R_L}{N}$	Yes/No

Special Mention

At present, generally three types of questions are asked in the achievement tests—essay type, short answer-type and objective type questions. These questions are selected on different basis.

1. **Analysis of Essay Type Questions (Items) :** The marks should be adequately divided on the basis of recall, logic and analysis. It should be made clear in the instructions that the division of marks should be on the basis of division of marks given in the model answers. Only then some objectivity can be brought in the evaluation of essay type questions. The following procedure should be followed in selection of these questions (items) :

    (1) The test should at first be administered on a representative sample of students.

    (2) Then the calculation of the obtained scores of each student on the essay type questions be done.

    (3) Then the percentage of the obtained scores of the essay type questions be calculated on the basis of total obtained scores.

    (4) The questions on which the candidates secure less than 10% or more than 90% marks should be left out. Rest of the questions should be selected with due analysis (on the basis of the blue print). In our example, two essay type questions have to be selected. For it, 4 questions should be constructed. One question should be selected out of the questions 10% to 20% scores and the other, out of the questions between 50% and 80% scores.

2. **Analysis of Short Answer Type Questions (Items):** Construction and marking of short answer type questions should be done on the basis of suitable instructions. Then the similar procedure should be applied as for the selection of essay type questions. The questions in required number should be selected with scores between 10% and 80%.

3. **Analysis of Objective Type Questions (Items) :** The difficulty level arid discriminating power of these type of questions are known by statistical methods, and the questions are selected accordingly. For it, the following procedure is adopted :

    (1) The test is at first administered on a representative sample.

    (2) The obtained scores of each student is calculated.

    (3) The answer-books are put in descending order on the basis of obtained scores.

    (4) High group and low group are selected on the basis of 27% of the top and bottom respectively, and the answer books in the middle of these figures are left out. The number of answer-books of each level is displayed by *n*. If the test was administered on 100 students, then out of answer-books put in descending order on the basis of obtained scores, 27 from the top and 27 from the bottom are selected. So the value of *n* will be 27.

    (5) After this, RH (the number of right questions by the higher group) and RL (the number of right questions by the lower group) are found out.

(6) On the basis of values of RH and RL, the difficulty level and discriminating power of each question is found out with the use of the following formulae :

$$DL = \frac{R_H + R_L}{2n} \times 100$$

In which, DL = Difficulty level in percentage

R_H = The number of correctly answered questions by the high level group

R_L = The number of correctly answered questions by the low level group

And, $DP = \frac{R_H - R_L}{n} \times 100$

In which, DP = Discriminating power in percentage

R_H = The number of correctly answered questions by the high level group

R_L = The number of correctly answered questions by the low level group

(7) Definite number of questions is selected out of such questions whose DL and DP values are between 30% and 70%. The rest are left out.

4. **Evaluating the Test :** The questions selected on the basis of item analysis are arranged in the form of a test. After this, the following aspects of the test are evaluated :

(1) The difficulty level is of great importance for an achievement test. Generally, the items of 50% difficulty are considered adequate for a test. So, in the first step of evaluation, the difficulty level is studied. Such items should not be included in the test which both of the high and low groups are able to solve, or are not able to solve.

(2) At the second step of evaluation the validity of the test is examined. Generally, content validity is emphasised upon in an achievement test.

(3) At the third step, attention is paid to the criticism and analysis of the items by specialists and students to carry out suitable amendments in the test items.

(4) At the fourth step, norms are found out for analysis of the obtained scores of the test.

(5) Evaluation of the test is done on the basis of an external criterion, such as to find out the correlation with another standardized achievement test to find out the correlation with. The test with higher correlation is approved.

(6) The coefficient of reliability of the test is found out by using a suitable method. Generally, a test with .80 or more coefficient of reliability is considered a good test.

5. **Construction of the Final Draft of the Test :** After the preliminary and final evaluation of the test, the final draft is prepared. Only those questions are included in the test which are valid and are of required difficulty level. The related instructions of the test too are written clearly. The method of administration of the test, time limit, marking method, etc too are finalised at this stage. It is the stage when the test is examined from all aspects.

6. **Preparing Test Manual :** Alter the construction of the final draft of the test, the last task is to prepare the test manual. In this manual measurable objectives, blue print, item analysis, data, instruction pertaining to administration, methods, validity and reliability of the test, and method of marking are clearly described.

Special Mention

(1) Dividing the teaching objectives or the measurable objectives of the test into cognitive, conative and affective aspects does never mean that there is complete absence of conative and affective objectives in the cognitive aspect, or cognitive and affective in the conative aspect, or cognitive and conative in the affective aspect. This difference is done only on the basis of relative density.

(2)   No achievement test can be fully valid and fully reliable. Our aim should be to make it more and more valid and reliable.

(3)   The essay type questions cannot be made fully objective however. Our effort should be guided towards making them more and more objective. The same criterion applies to short answer type questions; of course, they can be made more objective than the essay type questions.

(4)   As far as the tests with only objective questions or items are, concerned, they can be made fully objective, but they cannot be made fully valid and fully reliable.

(5)   It is clear that no achievement test can be fully valid, reliable, objective and discriminating. We can standardise it by making it more and more valid, reliable, objective and discriminating.

Task   How is the final draft of the test Constructed ?

Writing of Questions

The next step after the finalization of the blueprint is writing appropriate questions in accordance with the broad parameters set out in the blueprint. One should take one small block of the blueprint at a time and write out the required questions. Thus, for each block of blueprint which is filled in, questions have got to be written one by one. Once it is done, we have all the questions meeting the necessary requirement laid down in the blueprint. While selecting each small block for writing a question, you can proceed in several ways.

(a)   either writing all questions (one by one) belonging to one objective at a time i.e. knowledge or understanding or application followed by other objectives, or

(b)   by taking up questions according to their form or type i.e. Essay Type followed by Short Answer and Very Short Answer Type or in any other order, or

(c)   by writing questions for one unit of the syllabus or portion to be covered by the test at a time.

Each approach has its advantages and disadvantages, too. Irrespective of the method followed, the questions then have to be arranged in a logical sequence.

Marking Scheme

The fifth step is to prepare the "Marking Scheme". The marking scheme helps prevent inconsistency in judgement. In the marking scheme, possible responses to items in the test are structured. The various value points for response are graded and the marks allowed to each value point indicated. The marking scheme ensures objectivity in judgement and eliminates differences in score which may be due to idiosyncrasies of the evaluator. The marking scheme, of course, includes the scoring key, which is prepared in respect of objective type questions. Let us discuss this in detail.

Apart from the quality of the question paper, reliability of assessment, to a great extent, depends on the degree of consistency of scores assigned to the students by different examiners or by the same examiner on two different occasions. Thus, variation can occur because of any one of two different reasons :

(i)   Due to inconsistency of the same examiner when he/she examines different answer scripts adopting different standards.

(ii)   Due to different examiners using different standards of judgement.

If an answer script is awarded the same grade or marks on repeated exposure to the same examiner, the examiner is said to be consistent in awarding the marks. As such, the assessment done by him/her could be said to be more reliable and consistent than the other examiner in whose case variation in award of marks is higher.

The factor contributing to variations in the standards of assessment, both at the intra-and the inter-examiner levels, can be controlled by supplying a detailed scheme of marking along with the expected answers so that every examiner may interpret the questions in the same way and attain the

same standard of marking without being too lenient or strict or varying in his/her assessment. Subjectivity, is thus minimised and it is believed to give a more reliable picture of the students' performance.

Highlights of a good marking scheme

(1) It is a three column statement showing serial number of the questions, their expected outline answer and the marks allotted to each value point under them.

(2) In respect of long answer or essay type questions, the expected outline answers should :

   (*i*)   be complete and cover all possible or major areas as demanded by the questions

   (*ii*) clearly indicate each expected point or the parts under the outlined major areas

   (*iii*) provide direction as to whether all points will count towards a complete or correct answer or a set of points will be adequate enough for full credit (All this should clearly reflected), and

   (*iv*) indicate marks for each expected point. Marks so distributed over expected points or their sets should be equal to the total marks assigned for a question.

(3) In respect of short answer questions a complete answer may be provided with its break-ups where ever necessary along with the break-up of marks.

(4) Out of the total marks assigned for a question, each point so enumerated/explained may be assigned marks according to their significance in the answer.

(5) In some situations, apart from the content, other qualities of answer may also matter significantly, particularly in long answer or essay type questions. These could be logical approach, coherence, lucidity of expression, the style of presentation etc. Some marks may also be set apart for such overall quality of answer which cannot be usually covered in enumeration of the content points.

(6) The scheme of marking needs to be comprehensive enough not to leave any point unexpected and thus should provide clear guidelines in respect of the break-up of marks over different points or parts of the answer.

(7) If a question entails some other points beyond one's expectation, a provision may also be made to take them into account and suitably reward them.

26.5 Significance of Achievement Test

1. Achievement tests are used in selection of candidates in different fields and for admission of students in schools.

2. These tests are important from the standpoint of determination of class and promotion.

3. In the field of medicine and counselling, achievement tests are used extensively. These tests are very important from the standpoint of identifying backward students diagnostic and remedial teaching.

4. Achievement tests are helpful in measurement of minimum abilities of an individual.

5. Achievement tests are used widely in different types of classifications and for appointment.

6. These tests are helpful in providing educational and vocational guidance.

7. These tests are used to evaluate a teacher's proficiencies and effectiveness.

8. The use of these tests is helpful in the amendment and revision of curriculum.

9. Construction of these tests and study of related literature helps teachers to develop their professional attitude.

10. These tests provide assistance in learning. A student knows well how much he has studied and how much he has yet to study. It gives him motivation for studying in future.

11. These tests help to give reports to guardians and issue certificates.

12. These tests evaluate the effectiveness of different methods of teaching and help to select the most suitable method of teaching.

13. The use of these tests is helpful in the amendment of content.

14. It is possible to undertake comparative study of educational levels of different schools on the basis of these tests.

15. These tests acquaint us with the all-round mental ability of a student.

Characteristics of Achievement Tests

1. The content of these tests is as per the students' level, abilities, interests and aptitudes.

2. The test items in these tests are objective, so there is no question of awarding partial marks.

3. These tests are discriminating besides being reliable and valid.

4. These tests are very economical from the standpoint of money, time and energy.

5. The purpose of these tests is predetermined.

6. These tests are useful from the practical viewpoint.

7. These tests have a wide content.

8. These tests possess all characteristics of standardized tests, such as scoring key, manual of instructions, norms etc. All these are prepared in advance and are got printed in the form of a booklet.

9. These tests are made separately for different classes.

10. The scoring, time limit etc of these tests are determined before their administration.

11. The number of questions in these tests is large, so there is no question of chance factor.

12. The results of these tests provide such material to a teacher by which he can construct his entire instructional plan.

Self Assessment

2.   State whether the statements are 'true' or 'false'.

   (*i*)   The ratio method is the simplest method to find out difficulty level of an item.

   (*ii*)   The difficulty level and discriminating power of objective type of questions are known by item analysis.

   (*iii*)   The questions asked in a test are called item of the test.

   (*iv*)   After the Construction of the final draft of the test, the last task is to prepare the test manual.

26.6 Summary

•   The tests which are constructed to measure cognitive, conative and affective changes occurring as a result of teaching school subjects, are called achievement tests.

•   The achievement tests are used in schools for the following functions :

   •   To select the students for admission.

   •   To classify the students.

   •   To encourage the students.

   •   To promote the students.

   •   To issue certificates on the basis of public examinations and results.

•   The tests being used in the field of education are classified in different forms

•   Classification on the Basis of the Field of the Tests

- On the basis of measurement field, the tests can be divided into two classes as follows :

  - Educational Tests : Educational tests are those tests by which educational achievements and educational problems of the students are measured; such as—achievement tests and diagnostic tests. They have been discussed in the related chapter in detail.

  - . Psychological Tests :

- Classification on the Basic of Nature of the Tests

  On the basis of the nature of the tests, they have been divided into three classes—oral, written and practical.
  - Oral Tests : Those tests in which the questions from the examinees are asked orally and the examinees have to answer orally are called oral tests.

- Written Tests : The tests in which the examinees are given questions in written form and which they respond in writing, are called written tests. In these tests any types of questions can be asked as essay-type, short-answer-type and objective-type.

- Practical Tests : The tests in which the measurement of skills and activities of the students is executed by asking them to display the skill or activity are called practical tests; such as asking a child to sing a song in order to measure his proficiency in singing.

- The tests are classified into two types on the basis of material used in them, which are :
  - Verbal Tests : This type of tests includes those tests in which words or language are used which may be written or oral. Such tests are mostly used in the field of education.

  - Non-Verbal Tests : This type of tests includes those tests in which words or language are not used, rather scenery, symbols, signs, pictures, etc

- The essay type questions, should be according to the level of the class (as descriptive, argumentative, analytical, etc.) They should be constructed in a simple and clear language.

- The language of the short answer type questions too should be easy and clear.

- Review of the Test : At this step, the test so prepared is reviewed by specialists on the basis of which language errors and weightage given to different parts of the question are amended.

- The short-answer-type questions too are constructed in the same way. The objective-type questions are constructed such that 27% questions can be answered by the students of higher level only and 27% cab be answered by all.

- Selection of Questions after Testing of Preliminary Draft : In order to know of the errors of the test, the test is administered on some students of the required level.

- It is necessary to understand the meaning of item analysis. The questions asked in a test are called item of the test. By item analysis is meant numerical analysis of the obtained scores of each item, the analysis of its validity and reliability, and deciding its suitability accordingly.

- The difficulty level of a question or item is known from the percentage of students who succeed to respond it correctly. The difficulty level of an item indicates whether it is easy or difficult.

- Finding Out Difficulty Level of the Items : All specialists are of the opinion that the level of difficulty of an item can be found out by comparing the number of students answering it correctly with those of answering it incorrectly.

- It is a three column statement showing serial number of the questions, their expected outline answer and the marks allotted to each value point under them.

- In respect of long answer or essay type questions, the expected outline answers should :
  - be complete and cover all possible or major areas as demanded by the questions

- clearly indicate each expected point or the parts under the outlined major areas

- provide direction as to whether all points will count towards a complete or correct answer or a set of points will be adequate enough for full credit (All this should clearly reflected), and

- indicate marks for each expected point. Marks so distributed over expected points or their sets should be equal to the total marks assigned for a question.

- Achievement tests are used in selection of candidates in different fields and for admission of students in schools.

- These tests are important from the standpoint of determination of class and promotion.

- In the field of medicine and counselling, achievement tests are used extensively. These tests are very important from the standpoint of identifying backward students diagnostic and remedial teaching.

- Achievement tests are helpful in measurement of minimum abilities of an individual.

26.7 Keywords

Achievement : The act or process of achieving something

Construction : The process or method of building or making something

Psychological : Connected with a person's mind

26.8 Review Questions

1. What do you understand by achievement tests ?
2. Explain the types of achievement tests on the basis of nature and types of questions asked in the tests ?
3. Write the process of Construction of achievement type test stop by step
4. What are the significance of achievement test ?

Answers : Self Assessment

1. (*i*)   Achievement test
   (*ii*)   oral, written, practical
   (*iii*)   essay type, short answer type, objective type
   (*iv*)   practical tests (v) simple recall
2. (*i*)   T   (*ii*)   F   (*iii*)   T   (*iv*)   T

26.9 Further Readings

Books

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*
2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*
3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*
4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 27 : Diagnostic Test : Concept and Construction, Remedial Teaching

Objectives

The objectives of this unit can be summarised as below :
- To explain the Concept of diagnosis test
- To discuss the objectives of diagnosis test
- To describe the Characteristics of diagnostic test
- To discuss about the Construction and standardization of Diagnostic test
- To explain the process of diagnosis test
- To describe about the Remedial teaching

Introduction

A diagnostic test is a form of an achievement test, under which specifics and shortcomings of knowledge acquired from learning experiences of a specific object or topic are evaluated. In these tests, the level of a student is not found out, rather the purpose of a diagnostic test for a teacher is to found out what difficulty is being faced by a student in understanding a topic, what can be its possible cause, how this difficulty can be eradicated etc. Instructional diagnosis is related not merely with students' individual abilities and strengths, but with the remedy of their weaknesses and difficulties too. A teacher has to bring about suitable changes in his instruction and method of teaching in order to remove any difficulty of a student, so that he can gain maximum learning experiences as per his ability. This type of teacher's process is called remedial teaching.

27.1 Concept of Diagnostic Test

The teaching in schools is carried out in a planned way. In it, the teachers try to find out at regular intervals how much the students have learnt out of what has been taught to them. The tests that are prepared to measure it are known as achievement tests. The results of these tests present a new problem before the teachers that which part of the subject matter taught, the

students could not learn properly and why. The tests which are prepared to find out the causes of it are known as diagnostic tests. In other words, diagnostic tests are those tests by which the hindrances and difficulties in the process of learning of a unit of subject matter of the students are found out. On the basis of analysis of the obtained data from diagnostic tests, the weaknesses of the students are identified and the remedial measures to be taken by the students, the teachers and the guardians are found out. The diagnostic tests are useful only when the causes found are removed. So long a diagnostic test is not concerned with remedial teaching or activity, it remains incomplete.

The diagnostic test can be defined as follows :

The diagnostic test is the test which is constructed in order to find out the causes of failure of a student or group of students in learning a particular part of subject matter taught to them and with the help of which findings remedial suggestions are given.

In order to understand the form of diagnostic tests, it is essential that the difference among diagnostic test, achievement test and action research is understood.

27.2 Objectives of Diagnostic Test

Following are the chief objectives of diagnostic tests :

1. To provide educational and vocational guidance to students and guardians on the basis of traits realized from specific subject-related learning unit.

2. To make teaching-learning circumstances effective.

3. To make evaluation process more meaningful and effective.

4. To arrange for remedial teaching.

5. To give advice to a teacher for proper improvement in his teacher process.

6. To know about the weaknesses, deficiencies and difficulties of a student.

7. To assist in the selection of different tests, techniques and tools for knowing the causes related to the problem.

8. To assist in the selection of different types of questions for the construction of different achievement tests.

9. To amend textbooks of different subjects on the basis of the specific points and shortcomings and to make them more useful for students.

27.3 Characteristics of Diagnostic Test

A diagnostic test should have the following characteristics in order to its beings useful for teachers and students :

1. These tests are standardized, but some specialists opine that diagnostic tests should not be standardized.

2. Time limit for these tests is not specified.

3. These tests form an essential part of the curriculum.

4. These tests reveal the form of mental process of a learner.

5. These tests are according to the specific objectives.

6. The basis of these tests is formed by such facts or norms which are established on the basis of experiments.

7. These tests do not measure a child's ability, but diagnose his weakness in order to provide for remedy.

8. These tests are analytical and analyze all parts of a process fully.

9. Scores obtained by students are not attached any importance in these tests. They only see the level of difficulty of questions which a student can solve.

10. These tests test the progress of a student objectively.

If diagnostic tests are related with the remedy, then they become very beneficial for students. It has been proved on the basis of experiments that they have proved effective in instructional diagnosis and remedial teaching for all subjects. On the basis of his experiments in mathematics, Stone has said that diagnostic tests are very useful for small classes. Diagnostic tests and remedial teaching are closely related. Bukner and Melbi have written that the chief purpose of diagnostsic tests is to find out some specific weakness of a student so that the causes of weaknesses can be found out in order to take steps for remedial teaching.

Notes The word 'diagnosis' is extensively used in medical science. The patient is prescribed medicine on the basis of the diagnosis. If the diagnosis is proper, the medicine prescribed will cure the patient soon. In the same way, in the field of education, if the proper causes of not learning a subject matter correctly by the students are found out, remedy too can be done properly.

27.4 Construction and Standardization of Diagnostic Test

The subject for which a diagnostic test has to be constructed, is at first analyzed minutely and it is determined what concepts, principles, theories, rules, facts, formulae or processes have to be attached importance. Besides, those mental faculties of students have to be analyzed which are directly related to such knowledge, such as reasoning, logic, thinking power, observation power etc. For example, ability in mathematics requires to link to many fundamental theories, rules and formulae, such as multiplication, division, addition, subtraction etc. Besides, many mental abilities are also very important in it such as logic, thinking power and observation power etc. After this, different mental processes at application in that subject are analyzed and noted down, and then on the related concept, theory, rule, formula or process or mental processes, several questions are formulated. These questions are very simple and are written in the ascending order of difficulty. In order to find out the difficulty spots of a student, the question is noted down after which the student is not able to solve any question, by which it is predicted that the student is unable to solve the questions of the given difficulty. For example, questions on simple equation in algebra diagnostic test can be somewhat as follows

1. $6x + 7 \quad = 35$
2. $27x - .19x \quad = 12$
3. $x + 5 \quad = 20$
4. $4x + 2 \quad = .9x + 5$
5. $5x + .6 \quad = 7.8$

Now supposing a student is able to solve the first two equations correctly, but is not able to solve the questions that follow, then we can conclude that he can solve the simple questions of equations, but is unable to solve fraction or point related questions. Therefore, now the teacher focuses his attention on this aspect and diagnoses the causes due to which he is weak in this type of questions. Though the construction of diagnostic tests and their standardization process is similar to the achievement tests, yet these tests are somewhat different. A diagnostic test is an original skill test, while an achievement test is constructed not with a view to its relation in the realization of a skill, rather it only tells about the general achievement in a subject. A diagnostic test can be individual-oriented or group-oriented. An individual-oriented test comprises of test items related to the weaknesses and abilities of a particular student, while in a group-oriented test, test items are constructed keeping in view the specific group errors or weaknesses. For construction of test items, common and specific errors of the students are collected. Stems and distractors of test items are specified keeping in view the type of errors and their direction. The minuteness with which the distractors are determined will bring diversity to the test. Therefore, individual and group errors of

students should be minutely studied in order to convert observations into test items. Interpretation of scores is done under individual and group errors respectively, and different remedial teaching methods are adopted for these two types of errors.

Notes

27.5 Process of Diagnosis Test

Following are the five important steps of the process of instructional diagnosis :

1. **Selection of Students for Diagnosis :** At first those students are located who find it difficult to adjust in the school. These are those students who are weak in one or more subjects, and are not able to adjust well with some other activities of the school. The following methods can be used to locate such students :

    (a) Each student has peculiar causes for his weakness. Therefore, the students who need diagnosis should come to the teacher themselves.

    (b) Such students can be selected on the basis of test results held in school from time to time.

    (c) Teachers can select such students on the basis of their experience.

    (d) The students giving unsatisfactory performance in achievement should be given achievement tests and intelligence tests in order to locate them.

    (e) Some students have very high IQ but their performance is very low, and there can be some students who stand first in the class but are not able to adjust with the social environment. Such low-achievers and misfits should be located very carefully.

    (f) The interview method can prove to be very effective for locating such students.

2. **Identify Difficulty Points :** To know what student is facing what type of difficulty, diagnostic and performance or achievement tests can be administered which can be teacher-made or standardized. However, the most suitable tool for this are the diagnostic tests, which present the correct picture of a student's weaknesses and abilities. Though the scope of each diagnostic test is limited and narrow, yet the manner in which these test the narrow and limited field, is unique. The diagnostic importance of a test depends more on the teacher, and less on the form of the test. Sometimes, informal tests too prove helpful in this field.

    "Such informal and observation charts usually indicate the correct level on which to start remedial instruction."   — *Durrell*

    In brief, by diagnostic tests is seen the difficulty level of which a student can solve questions based on concepts, laws, processes, formulae and principles etc, and at what places he finds it difficult.

3. **Analysis of Difficulty Points :** It is very difficult to ascertain why a student is committing a particular type of error again and again, because the mind of each student functions in a peculiar manner. Having ascertained the nature of subject-related weakness and level of difficulty of a student, a teacher sets out on the task of locating its causes. A teacher tries to locate its cause amongst one or more of the following causes : physical defects, emotional instability, disinterest, indifference, bad habits, lack of general or specific intelligence, school-related causes (lack of good teaching, bad behaviour by teachers or classmates, failure in tests, disinterest of teacher in the student, prejudiced behaviour, harsh discipline etc.), domestic causes (bad behaviour or ill-treatment by parents and other family members or relatives, lack of proper resources for study, lack of suitable environment, poverty, business in domestic tasks etc.). Sometimes, he can estimate the causes and tries to ascertain them on the basis of his experience or interview.

4. **Remedial Procedures :** Having diagnosed a student's weakness, now is the time to undertake procedures for its remedy. A suitable plan for the removal of weakness and error is constructed which comprises of the details of the weakness and its causes, and measures to be undertaken for its remedy. If a number of students have committed the same type of error, then they may be treated collectively, and if the error is individual, then it should be remedied individually. A teacher should keep the following aspects in mind while using remedial measures :

Notes

(a) These methods are not foolproof against idiots.

(b) If remedial work does not succeed, then the form of the programme should be changed.

(c) The students who are making progress should be admired for motivation.

(d) These tests are not so successful with brilliant students as they are effective with dull students.

(e) It may take time to eradicate bad habits of brilliant students, therefore, a teacher should work with patience.

(f) There should be a graph and chart of each student's progress.

(g) The subject in which a student is weak, he should be given related literature, magazines, periodicals etc. to read.

(h) In order to make work more lively, attractive aids and effective method of teaching should be used.

5. **Preventive Measures :** If we desire that a student may not commit any error in the subject in future, then we should effect such changes in his school and domestic environment so that his problem of maladjustment can be permanently eradicated. For this, we have to construct a multidimensional project, so that the student can be given a suitable environment; such as, improvement in school environment, amendment to curriculum, improvement in examination system, construction of suitable ability tests, better environment etc. Prognostic tests can be administered to see whether a student is ready to learn abstract topics or not. When a student is weak in learning arithmetic as well as lacks intelligence, it would be proper to teach him generalized arithmetic. A prognostic test is constructed with this in view. These tests predict a student's learning in future.

Example

Construction of Diagnostic Tests in Algebra :

Name of student : XYZ

Class and section : 7A

School's name : Government Inter College XYZ city

Date : October, 2010

Instructions

1. Some questions based on simple equations are given. Find out the value of 'x'.

2. Write the answers on the specified places.

3. Pay specific attention to neatness and cleanliness. If the space is inadequate for calculation, you may take extra sheet for rough work.

4. There is no time limit for the test, still solve all questions as soon as possible.

Questions	Answers	Place for Calculation
1. $x + 15 = 35$	(1)	
$x =$		
2. $7x + 14 = 35$	(2)	
$x =$		
3. $2x + 23/5 = 74/9$	(3)	
$x =$		
4. $3x + .6 = 4.5$	(4)	
$x =$		
5. $.05x + 7 = 14$	(5)	
$x =$		

6. $5x - 12 = 23$ (6)

      $x =$

7. $7x - 13x = 49$ (7)

      $x =$

8. $.023x - 17 = 19x - 30$ (8)

      $x =$

9. $14x + 6 = 4 - 5x$ (9)

      $x =$

10. $4x + 2 = .8x + 1$ (10)

      $x =$

11. $x/4 - 3x/5 = 6$ (11)

      $x =$

12. $2x + 3/9x = 8$ (12)

      $x =$

13. $x + 2x/3 = 18$ (13)

      $x =$

14. $x + 17 = 32$ (14)

      $x =$

15. $5x = 25$ (15)

      $x =$

16. $4x = 2/25$ (16)

      $x =$

17. $3x - 9 = x = 15$ (17)

      $x =$

18. $5x + 7x = 36$ (18)

      $x =$

19. $x/4 + 1/4 = 5/4$ (19)

      $x =$

20. $5(x + 3) = 20$ (20)

      $x =$

The above questions comprise of those based on simple equation of different forms. All questions have been constructed on the basis of maxim 'from simple to difficult'. The sole purpose of the test is to find out the difficulty level up to which the students of class VII student can solve the questions. If they are not able to solve them all, then what and where their difficulties are. The above format of test can be helpful to locate the students' difficulties in simple equations. The following diagnostic chart contains the errors committed by seven students, on the basis of which we can locate their difficulties and weaknesses.

On the left side are written the students' names who are administered the diagnostic test, and on the right, different columns are noted with the different types of questions. Now, the difficulties faced by each student are marked in the related column, with details of number of times a student commits that error. The total number of questions is 20, and the number of different types of questions included in the test have been indicated in the brackets.

S. No.	Students' Names	Subtraction in equation	Multiple	Equation and simple fraction	Equation and decimal fraction	Equation and simple addition	Side difference	Scores	Absence in class	Remarks
1.	Rajendra	–	–	–	–	–	–	20	1	I
2.	Nitin	–	–	–	2	1	–	17	5	III
3.	Usha Pande	–	–	–	–	–	1	19	3	II
4.	Karuna Singh	1	–	4	2	–	1	12	7	V
5.	Imran	1	–	–	1	2	6	10	12	VI
6.	Jitendra	–	2	–	–	1	–	17	4	III
7.	Nishi	3	–	–	–	–	1	16	6	IV
8.	Total errors in the entire class	9	–	–	12	14	17			
9.	Number of right responses in the entire class	46	32	58	43	86	74			
10.	Total number of questions	60	35	60	70	100	90			
11.	Percentage of right responses	76	91	96	62	86	82			

Did u know? **Prognostic Test of Algebraic Ability** made by **Lee** and **Orleans** is available.

Diagnostic Chart

In the above chart, the weaknesses of seven students have been analyzed. It is clear from the chart that Imran is the weakest student of the class. He has obtained only 10 marks out of 20. Secondly, his absence in the class has been the highest. This can be the cause of his weakness. He has incorrectly done all six of the questions on side difference. It appears that his disinterest in the subject, carelessness, misunderstanding or absence in the class has caused him to remain weak in the subject. Rajendra is the most brilliant student of the class, who has scored 20 out of 20, which indicates that he has understood the subject well. Usha Pande too is a talented student of the class, who has mistaken in only one question. Maybe she has committed the mistake by chance. In the same way, the chart gives information about other students in the group. From the standpoint of difficulty level of questions, the percentage of students solving decimal fraction questions is very low, which is 62%. From this viewpoint, the teacher should give more exercise pertaining to these questions, and should teach them more carefuly. From the standpoint of remedial teaching, Imran requires remedial teaching, else he will lag behind in the class. Thus, the above diagnostic chart is a blueprint of the weakness and difficulties of students. In indicates about a particular student what error he has committed and where, and what remedy can be arranged for him.

27.6 Remedial Teaching

Remedial Measures

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

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

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

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

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

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

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

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

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

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

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

Task   What is prognostic test ?

Educational Diagnosis and Remedial Teaching

Diagnostic test is constructed for educational diagnosis, and remedial teaching is conducted on the basis of diagnosis. We shall clarify it by an example.

For example, take teaching of a sub-topic of mathematics : square root. Suppose, after the teacher has taught the method of finding out square root, some of the students still commit error in it. The teacher will conduct diagnostic and remedial work in this order:

1. First, the teacher will construct different questions of decimal numbers to find out square root; such as :

    (a) Find out square root of the following numbers :

        (*i*) 1355.23   (*ii*) 25.2143   (*iii*) 44.135678

    (b) Find out square root of the following numbers up to two decimal points :

        (*i*) 210.7   (*ii*) 12345.071   (*iii*) 7.0357

2. He will give this question paper to the group of students which commits errors.

3. He will analyse the students' answers and then arranges their causes.

    Suppose, the students commit the following errors in solving the above questions :

    (a) Some students do not take down the pairs of numbers while calculating square root.

    (b) Some students do not pay attention to decimal point while making pairs.

    (c) Some students do not add zero after decimal point to make pairs.

4. On the basis of the above errors, the teacher will classify the students and prepare their lists; such as — the list of students who do not take down the pairs correctly, the list of students who overlook the decimal point, and the list of students who do not add zero after decimal point to make pairs.

5. After this, the teacher will formulate different types of hypotheses. These hypotheses will be formulated on the basis of student-related hindering factors, teacher-related hindering factors, home-related hindering factors and culture related hindering factors. Some of the student-related and teacher-related hypotheses are given here, for illustration :

    (*i*)   The student is not attentive in the class.

    (*ii*)   The student thinks that mathematics is a difficult subject and overlooks to learn it.

    (*iii*) There is no coordination between the rate of students' learning and teacher's teaching.

    (*iv*)   The student has not done sufficient exercise.

    (*v*)   The teaching method is not suitable.

    (*vi*)   The teacher has not made the students practice enough.

    (*vii*) The teacher has not improved the errors in checking work.

    In the same way, if the teacher considers it necessary, he can also formulate hypotheses about curriculum-related, home-related, guardians-related and culture-related hindering factors.

6. The teacher will construct questions to test various types of hypotheses.

7. The teacher will select correct hypotheses by analysis of the answers to these questions.

8. At last, the teacher will conduct remedial teaching on the basis of correct hypotheses.

Self Assessment

1. *State whether the following statements are 'true or 'false'.*

    (*i*)   The diagnostic test is the test which is constructed in order to find out the causes of failure of a student in learning a particular part of subject matter.

    (*ii*)   There is no difference between diagnostic test and achievement test.

    (*iii*) A diagnostic test has no meaning until it suggests remedial teaching.

    (*iv*)   The time limit for diagnostic test is specified.

    (*v*)   There are many factors which hinder in learning.

27.7 Summary

*   The teaching in schools is carried out in a planned way. In it, the teachers try to find out at regular intervals how much the students have learnt out of what has been taught to them.

*   The diagnostic test is the test which is constructed in order to find out the causes of failure of a student or group of students in learning a particular part of subject matter taught to them and with the help of which findings remedial suggestions are given.

*   Following are the chief objectives of diagnostic tests :

    *   To provide educational and vocational guidance to students and guardians on the basis of traits realized from specific subject-related learning unit.

- To make teaching-learning circumstances effective.

- To make evaluation process more meaningful and effective.

- To arrange for remedial teaching.

- To give advice to a teacher for proper improvement in his teacher process.

- The subject for which a diagnostic test has to be constructed, is at first analyzed minutely and it is determined what concepts, principles, theories, rules, facts, formulae or processes have to be attached importance. Besides, those mental faculties of students have to be analyzed which are directly related to such knowledge, such as reasoning, logic, thinking power, observation power etc. For example, ability in mathematics requires to link to many fundamental theories, rules and formulae, such as multiplication, division, addition, subtraction etc.

- Selection of Students for Diagnosis : At first those students are located who find it difficult to adjust in the school. These are those students who are weak in one or more subjects, and are not able to adjust well with some other activities of the school.

    - Each student has peculiar causes for his weakness. Therefore, the students who need diagnosis should come to the teacher themselves.

    - Such students can be selected on the basis of test results held in school from time to time.

    - Teachers can select such students on the basis of their experience.

- Identify Difficulty Points : To know what student is facing what type of difficulty, diagnostic and performance or achievement tests can be administered which can be teacher-made or standardized.

- Analysis of Difficulty Points : It is very difficult to ascertain why a student is committing a particular type of error again and again, because the mind of each student functions in a peculiar manner. Having ascertained the nature of subject-related weakness and level of difficulty of a student, a teacher sets out on the task of locating its causes.

- Remedial Procedures : Having diagnosed a student's weakness, now is the time to undertake procedures for its remedy. A suitable plan for the removal of weakness and error is constructed which comprises of the details of the weakness and its causes, and measures to be undertaken for its remedy. If a number of students have committed the same type of error, then they may be treated collectively, and if the error is individual, then it should be remedied individually.

- A teacher should keep the following aspects in mind while using remedial measures :
    - These methods are not foolproof against idiots.
    - If remedial work does not succeed, then the form of the programme should be changed.
    - The students who are making progress should be admired for motivation.
    - These tests are not so successful with brilliant students as they are effective with dull students.

- The following factors should be kept in view while conducting remedial measures :

- The weak students should be asked to sit on the front seats in the class.

- The development of the subject matter should be done with the help of examples and illustrations.

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

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

27.8 Keywords

Diagnostic : Connected with identifying something especially an illness

Remedial : aimed at solving a problem

Procedure : The way of doing something

Task   Prepare a diagnostic test on the topic 'percent' in mathematics for class VII.

27.9 Review Questions

1.   What is a diagnostic test ? What is its utility ?
2.   How are diagnostic tests constructed ? Illustrate.
3.   Write the objectives of diagnostic tests.
4.   How is a diagnostic test different from achievement test ?
5.   What is the difference between diagnostic test and action research ?
6.   Describe the process of remedial teaching.
7.   What factors should be kept in mind while conducting remedial teaching ?

Answers : Self Assessment

(*i*)   T   (*ii*)   F   (*iii*)   T   (*iv*)   F   (*v*)   T

27.10 Further Readings

Books

1.   Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*
2.   Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*
3.   Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*
4.   Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 28 : Objective Type Test : Advantages and Limitations

CONTENTS

Objectives

Introduction

28.1 Meaning of Objective Test

28.2 Objective Test Formats

28.3 Construction of Objective Test Items

28.4 Advantages of Objective Test

28.5 Limitations of Objective Test

28.6 Summary

28.7 Keywords

28.8 Review Questions

28.9 Further Readings

Objectives

The objectives of this unit can be summarised as below

- To explain the meaning the objective test
- To discuss about the objective test formats
- To explain about the Construction of objective test items
- To discuss about the advantages and limitations of objective test
- To describe the effect of objective test items of teaching and learning

Introduction

Objective tests began to be used early in the twentieth century as a means of evaluating learning outcomes and predicting future achievement, and their high reliability and predictive validity led to the gradual replacement of the essay test.

One common misconception about the objective test item is that it is limited to testing specific, often trivial, factual details, which would sometimes lead to the use of an essay or performance test to assess students' comprehension of broader principles or their ability to apply them. However, as Robert Ebel pointed out, well written objective tests (especially multiple choice tests) can actually assess such higher-order abilities to some extent. While it is true that some types of knowledge or abilities cannot be assessed by objective tests, educators also should keep in mind that what test items can assess depends largely on the skills and effort of the test constructor, rather the test format per se.

28.1 Meaning of Objective Test

An objective test item is defined as one for which the scoring rules are so exhaustive and specific that they do not allow scorers to make subjective inferences or judgements; thereby, any scorer that marks an item following the rules will assign the same test score.

Several types of questions are asked in them. These questions are generally classified into two types—Recall type questions and Recognition type questions. Recall type questions have two types—Simple recall type and Sentence completion questions. Recognition type questions have serveral forms in which are included Classified questions, True/False questions, Multiple choice type questions, Matching type questions, Multiple choice and completion questions, and Analogy questions. We shall clarify their form and difference here by examples :

28.2 Objective Test Formats

A variety of different types of objective test formats can be classified into two categories : a selected response format, in which examinees select the response from a given number of alternatives, including true/false, multiple choice, matching test items; and a constructed response format, in which examinees are required to produce an entire response, including short answer test items. This distinction is sometimes captured in terms of recognition and recall. These two general categories are further divided into basic types of objective tests, illustrated in the following examples.

The true/false test is the simplest form of selected response formats. True/false tests are those that ask examinee to select one of the two choices given as possible responses to a test question. The choice is between true and false, yes and no, right and wrong, and so on. A major advantage of the true/false test is its efficiency as it yields many independent responses per unit of testing time. Therefore, teachers car cover course material comprehensively in a single test. However, one apparent limitation of the true/false test is its susceptibility to guessing. It should be noted, however, that test givers can attenuate the effects of guessing by increasing the number of items in a test. In addition, guessing might reflect partial knowledge, which would provide a valid indication of achievement.

28.3 Construction of Objective Test Items

Basically, scoring objective test items is easy : It only requires one to follow the scoring rules. However, constructing good objective test items requires much more skill and effort. The first step is to develop a set of test specifications that can serve to guide the selection of test items. A table of specifications (or test blueprint) is a useful tool for this purpose. This tool is usually a two-way grid that describes content areas to be covered by the test as the row headings and skills and abilities to be developed (i.e., instructional objectives) as the column headings.

After specifying the content and ability covered by the test using the table of specifications, the appropriate test item format is selected for each item. At this point, not only objective test items but also other types of test items—essay test or performance assessment—should be considered, depending on the learning outcomes to be measured.

The next step is to create specific test items. Typically, it is particularly important for objective test items to be written in clear and unambiguous language to allow examinees to demonstrate their attainment of the learning objectives. If complex wording is used, the item simply reflects reading comprehension ability. It is also important for each objective test item to focus on an important aspect of the content area rather than trivial details. Asking trivial details not only makes the test items unnecessarily difficult, it also obscures what the test constructor really wants to measure. Similarly, relatively novel material should be used when creating items that measure understanding or the ability to apply principles. Items created by copying sentences verbatim from a textbook only reflect rote memory, rather than higher order cognitive skills.

Caution Test constructors must be very careful that examinees with little or no content knowledge cannot arrive at the correct answer by utilizing the characteristics of the test format that are independent of specific content knowledge.

28.3.1 True/False Items

The format of items that represent true-false and allied varieties requires examinees to select the answer from two alternatives, which remain the same for the whole series of items. These items are usually in the form of statements (sometimes in question form) and examinees are to make judgment about their truth or falsity. These items may also take the form of Yes-No, agree-disagree, synonyms-antonyms, correct-wrong etc. Statement is preceded by a clear direction regarding stipulation of responses and mode of responding.

Forms of Constant Alternative Items

(a) *True-False variety* (T/F)

(*i*) Temperature remaining the same, pressure in a gas varies directly proportional to its volume.

(*ii*) Star fish belongs to the class Pisces. (T/F)

(b) *Right-Wrong variety* (R/W)

(*i*) Chandigarh is the capital of both Haryana and Punjab. (R/W)

(*ii*) Isosceles triangle has all the three sides equal. (R/W)

(c) *Agree-Disagree variety* (A/D)

(*i*) Every religion teaches hatred for other religions. (A/D)

(*ii*) All religions emphasise the same tenets of good living. (A/D)

(d) *Yes-No variety* (Yes/No)

(*i*) Is potato an underground stem ?

(*ii*) Does a water pump work on the principle of air pressure ? (Yes/No)

Construction of Objective-based True-False Items

It looks easy to frame T/F items, but construction of good T/F items requires insight into their structure. It involves : (a) selection of good concept. principle proposition; (b) restating the essence of the idea; (c) finding its implications; (d) couching its anthesis; and (e) writing the item using the same essential point for one true and one false version.

Observe the following 5 versions of the same idea, principle or concept, reflecting the five steps.

(*i*) Plants floating in water have abundant stomata in their leaves.

(*ii*) Water plants with floating leaves have more stomata on the upper surface than on the lower surface.

(*iii*) Abundant stomata on the upper surface of the floating leaves are meant to increase transpiration.

(*iv*) Rate of transpiration in floating leaves is directly proportional to the number of stomata on the upper side.

(*v*) If all the stomata on the upper surface of floating leaves of a water plant are smeared with vaseline, the leaves would rot.

(*vi*) If all the stomata on the lower surface of floating leaves of a water plant are smeared with vaseline, the leaves would rot.

If observed carefully the competency tested in these 5 items would vary from ability to recall, compare, interpret and establish relationship to that of ability to predict. Thus a T-F item can be constructed to test higher-order objective provided one develops the technique reflected in a-e steps mentioned above.

Class-room tests show some advantages that they offer simple, direct tests of students' knowledge, their efficiency is better in terms of scorable response per hour of testing time, writing of items is

simpler and less time consuming, and are good for testing cause- and-effect relationship as well as for discrimination between facts and opinion.

Notes The commonly visualised shortcomings of T-F items are encouraging blind guessing, testing of trivials, verbal memory, lack of qualifiers, and negative suggestion effect; and favouring willing in aggressive students to take chance.

28.3.2 Matching-Type Items

Design

Matching-type items are prescribed as set of terms, events, phrases, definitions etc., called the premises, which are written on the left-hand side, say column I. Another set of name pictures, statements etc., called the responses, are placed on the right-hand side under column II. Students are asked to match each item with the corresponding response, which is considered as one test item. Relationship may be between a term and definition, object and its functions, inventor and inventions, author and work, dates and events, problems and solutions etc. If the number of responses are equal to the number of premises, it is termed perfect matching. If the number of responses are more than the premises. it is called imperfect matching. If single criterion for matching the items is used, it is called simple matching; but if more than one criterion is used as the basis of matching, i.e. more than two lists are given, it is termed compound matching. While attempting the matching-type items, the student has more choices (2, 3,4, 5, 6 or even more) to choose from for each item (premises). The number of choices (responses) goes on decreasing as one goes on attempting one item after another. Thus every matching-type item in a set is indeed a multiple choice-type item. The following examples would clarify the concept of different varieties of matching-type items.

28.3.3 Rearrangement-Type Items

Design

Examinees are required to re-arrange the randomly presented material into some specified order. Material may be presented in the form of a series of statements one after the other or responses may be given of the multiple choice type. Direction is to be provided whether the responses are to be rearranged by writing them in specified order; to serial them into particular order; or indicate the serial number of each response etc.

Types of Re-arrangement Items

Chronological Order

Rewrite the serial number of the Indian Presidents listed below from past to present according to chronological order in the space provided.

| 1. | A.P.J. Abdul Kalam | 3. | S.D. Sharma | (4-2-3-1) |
| 2. | Giani Zail Singh | 4. | Rajendra Prasad | |

Functional Order

Rearrange the following steps involved in the manufacture of food by plants in order of their occurrence, giving the serial number in the bracket provided against each.

1.	Splitting of water	(3)
2.	Formation of starch	(5)
3.	Excitation of chlorophyll	(1)
4.	Evolution of oxygen	(4)
5.	Formation of A.T.P.	(2)

> *Task* What is meaning of the term response ?

28.3.4 Structured Questions

Concept

Such questions are rooted in a given stipulated situation, providing the needed introductory statement in the form of a passage, experimental data, table, or diagram, followed by a number of subquestions based on the subject matter. Initial question followed by subquestions are framed to test different abilities, as depicted in Chart 1.

Basic introductory material based on selected theme

Subquestion → Initial subquestion

Subquestion

Subquestion } Successive subquestions
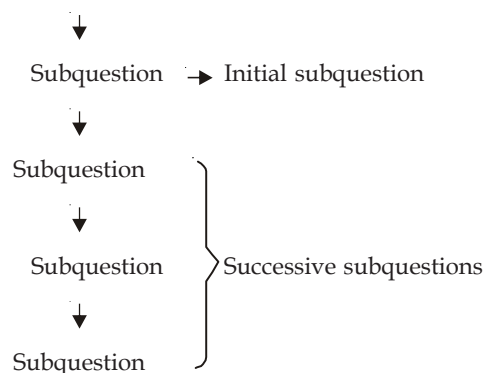
Subquestion

Chart 1 : Format of structured questions

Principles of Construction

(a)   Same introductory paragraph or material is to be used for each question.

(b)   Introductory material should neither be inadequate nor have redundant matter.

(c)   Subquestions may be of short-answer or objective type. Usually the same form of question is used in one set.

(d)   Each subquestion must demand reading of introductory paragraph for answering.

(e)   Each subquestion must be independent and should not depend on knowledge of previous questions.

(f)   Each question should test different abilities as far as possible.

(g)   All questions must test high-order abilities – not simply recall.

(h)   Subquestions may be arranged according to abilities tested, difficulty level, sequence of events or concepts involved.

(i)   Such questions should be time effective by keeping limited number of questions.

28.3.5 Multiple-Choice Questions (MCQ)

The most potential and usable form of objective tests is the MCQ. These questions are either used exclusively as in some selection tests, or in combination with other forms of questions as in most of the examinations conducted by the examination boards or other agencies.

If the number of candidates is very large and results are to be declared within a few days, the use of MCQ cannot be overemphasised. Exclusive use of MCQ is warranted in all such situations where

100% scoring objectivity, time constraint, computerisation of results, machine scoring, ranking, post-examination statistical moderation and record of psychometric properties of items are considered the necessary conditions or requirements.

The multiple-choice items are based on response-directed stimulus, in which responses or options may be arranged in different ways. The choice of the correct answer among the given options or responses can be made from the independent set of responses given for each item or from the same set of responses given for different items. Accordingly, we have different varieties of multiple choice items in terms of design or format of the items as given in the classification flow-chart.

Did u know? When the purpose of examination is to rank the candidates for selection or awarding scholarships as in National Talent Search Examination conducted by NCERT in India, MCQ may be used exclusively to ensure complete objectivity in scoring.

Self Assessment

1. *Fill in the blanks :*

   (*i*) The _____ is the simplest form of selected response formats, in which one of the two choices give as possible response to a test question.

  (*ii*) In _____ items require the examinee to select one or more responses from a set of options.

 (*iii*) The correct alternative in each item is called the _____, and the remaining alternatives are called _____.

 (*iv*) The multiple choice items are based on _____, in which responses or options may be arranged in different ways.

  (*v*) In matching type items the terms which are written on the left hand side are called _____ and on the other side are called _____.

28.4 Advantages of Objective Test

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

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

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

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

4. **Comprehensiveness :** As has been stated in the beginning, small questions are asked in these tests and the answers to these questions are given in one mark, number or word. From the viewpoint of availability of time, the time to be taken on to write one essay type answer or 4-6 short-answer-type answers can suffice to think and write the answers of 40-50 objective questions. Thus, these tests are comprehensive in nature.

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

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

28.5 Limitations of Objective Test

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

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

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

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

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

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

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

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

Self Assessment

2. State whether the following statements are 'true' or 'false'

(*i*) Objective tests are generally made to measure the knowledge of the students.

(*ii*) The questions in objective type tests do not cover the whole course.

(*iii*) Objective type questions do not help in the measurement of language skill.

(*iv*) These tests can be easily copied in exams.

28.6 Summary

• An objective test item is defined as one for which the scoring rules are so exhaustive and specific that they do not allow scorers to make subjective inferences or judgements; thereby, any scorer that marks an item following the rules will assign the same test score.

- A variety of different types of objective test formats can be classified into two categories : a selected response format, in which examinees select the response from a given number of alternatives, including true/false, multiple choice.

- The true/false test is the simplest form of selected response formats. True/false tests are those that ask examinee to select one of the two choices given as possible responses to a test question. The choice is between true and false, yes and no, right and wrong, and so on. A major advantage of the true/false test is its efficiency as it yields many independent responses per unit of testing time.

- Another selected response format type is the multiple-choice test, which has long been the most widely used among the objective test formats. Multiple-choice test items require the examinee to select one or more responses from a set of options (in most cases, 3–7). The correct alternative in each item is called the answer (or the key), and the remaining alternatives are called distracters.

- Scoring objective test items is easy : It only requires one to follow the scoring rules. However, constructing good objective test items requires much more skill and effort. The first step is to develop a set of test specifications that can serve to guide the selection of test items. A table of specifications (or test blueprint) is a useful tool for this purpose.

- The next step is to create specific test items. Typically, it is particularly important for objective test items to be written in clear and unambiguous language to allow examinees to demonstrate their attainment of the learning objectives.

- Many other specific rules exist for constructing objective test items. Jason Millman and his colleagues called this skill of the examinees "test-wiseness." For example, in multiple-choice test items, all options should be grammatically correct with respect to the stem (questions or incomplete statements preceding options), and key words from a stem, or their synonyms, should not be repeated in the correct option.

- Matching-type items are prescribed as set of terms, events, phrases, definitions etc., called the premises, which are written on the left-hand side, say column I. Another set of name pictures, statements etc., called the responses, are placed on the right-hand side under column II. Students are asked to match each item with the corresponding response, which is considered as one test item.

- If the number of responses are more than the premises. it is called imperfect matching. If single criterion for matching the items is used, it is called simple matching; but if more than one criterion is used as the basis of matching, i.e. more than two lists are given, it is termed compound matching.

- Steps in Construction
  - Decide about the intended aspect of measurement.
  - List the premises to be used.
  - Identify responses including the distracters.
  - Arrange items in some order.
  - Write clear direction for students.
  - Indicate space and mode of responding.

- Examinees are required to re-arrange the randomly presented material into some specified order. Material may be presented in the form of a series of statements one after the other or responses may be given of the multiple choice type.

- Principles of Construction
  - Same introductory paragraph or material is to be used for each question.

- Introductory material should neither be inadequate nor have redundant matter.

- Subquestions may be of short-answer or objective type. Usually the same form of question is used in one set.

- Each subquestion must demand reading of introductory paragraph for answering.

- STQ can be used in selection tests like those of N.T.S. examinations and Jawahar Navodaya Vidyalaya admission tests, especially in languages for testing reading comprehension.

- In larger heterogeneous groups of students, STQ can be used with advantage for catering to all ability ranges of students by setting questions on wider range of abilities to be tested.

- The most potential and usable form of objective tests is the MCQ. These questions are either used exclusively as in some selection tests, or in combination with other forms of questions as in most of the examinations conducted by the examination boards or other agencies.

- The multiple-choice items are based on response-directed stimulus, in which responses or options may be arranged in different ways. The choice of the correct answer among the given options or responses can be made from the independent set of responses given for each item or from the same set of responses given for different items.

- An ordinary multiple-choice item consists of an incomplete statement or a question, called the *stem* of the item, followed usually by four or five suggested answers (choices, called the responses or alternatives) out of which usually one is correct. The correct answer is called the *key* and the incorrect answers are called the *distractors*.

- Single-response variety

- Multiple-responses variety

- Best, least or worst variety

- All-or-none variety

- Multiple completion or multiple selection

- Combined-response variety

- Classification variety

- Negative-response variety

- Sequenced-response variety

- Cause-and-effect variety

- Analogy-type items

- Assertion-reasoning variety

28.7 Keywords

Objective : Something that you are trying to achieve

Blueprint : A plan which shows what can be achieved and how it canbe achieved

Discrimination : The ability to judge what is good, true etc.

28.8 Review Questions

1. Give the Construction of different types of objective type test.

2. Explain the format of Multiple Choice Questions.

3. What is the usablility of matching items of objective type test ?

4. What are the advantages and limitations of objective type tests ?

Answers : Self Assessment

1. (*i*)   true/false test   (*ii*)   Multiple choice test

    (*iii*) answer, distractors   (*iv*)   response directed stimulus

    (*v*)   premises, response

2. (*i*)   T   (*ii*)   F   (*iii*)   T   (*iv*)   T

28.9 Further Readings

Books

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 29 : Short Answer Type Test : Advantages and Limitations

---

CONTENTS

Objectives

Introduction

29.1 Meaning of Short Answer Type Test

29.2 Format of Short Answer Type Test

29.3 Principles of Construction Short Answer Type Test

29.4 Editing and Reviewing of Short Answer Type Test

29.5 Advantages of Short Answer Type Test

29.6 Disadvantages of short Answer Type Test

29.7 Summary

29.8 Keywords

29.9 Review Questions

29.10 Further Readings

Objectives

The objectives of this unit can be summarised as below

- To explain the meaning of short answer types test

- To discuss the types of short answer type test

- To describe the format of short answer type test

- To discuss the principles of Construction short answer type test

- To describe about the Editing and reviewing of short answer type test

- Advantages of Short Answer Type Tests

- Disadvantages of short Answer Type Tests

Introduction

Short answer questions are typically composed of a brief prompt that demands a written answer that varies in length from one or two words to a few sentences. They are most often used to test basic knowledge of key facts and terms. An example this kind of short answer question follows :

Short answer questions have many advantages. Many instructors report that they are relatively easy to construct and can be constructed faster than multiple choice questions. Unlike matching, true/false, and multiple choice questions, short answer questions make it difficult for students to guess the answer. Short answer questions provide students with more flexibility to explain their understanding and demonstrate creativity than they would have with multiple choice questions; this also means that scoring is relatively laborious and can be quite subjective. Short answer questions provide more structure than essay questions and thus are often easy and faster to mark and often test a broader range of the course content than full essay questions. For Example :

"Will you include short answer questions on your next exam ? Please justify your decision with two to three sentences explaining the factors that have influenced your decision."

29.1 Meaning of Short Answer Type Test

Form of Short Answer Type Tests : The tests in which such questions are asked, which can be answered briefly are called short answer type tests. Short answer type questions are generally of two types–short answer type and very short answer type. We are illustrating the both types here.

Short Answer Type Questions : The questions which can be answered in comparatively less time and in less words than the essay type questions are called short answer type questions. The time taken to answer one essay type question can suffice to answer 4-6 short-answer type questions. Following are the examples :

1. What is the difference between measurement and evaluation ?
2. What are the steps of measurement ?
3. What do you understand by essay-type questions ?
4. Enlighten any two characteristics of transcendental poetry.
5. What are the figures of speech ?

Very Short Answer Type Questions : The questions which can be answered even more briefly and in shorter time than the short answer type questions are called very short answer type questions. The time taken to answer one essay type question or 4-6 short answer type questions can suffice to answer 8-10 very short answer type questions. Here are given the examples of very short answer type questions :

1. What do you mean by measurement ?
2. What do you mean by evaluation ?
3. Enumerate the steps of measurement.
4. Enumerate the steps of evaluation.

Notes   Short answer questions can also be used to test higher thinking skills, including analysis or evaluation.

29.2 Format of Short Answer Type Test

Basically we have VSQ and SAQ. Questions of both types can be set using different formats, as illustrated below.

(a) VSQ

   (*i*)   *Direct question*

            What is the chemical formula of common salt ?

   (*ii*)   *Selection form*

            Which of the two 'come' or 'coming' is the present participle ?

   (*iii*) *Completion form*

            $(a + b)^2$ = _____ .

   (*iv*)   *Identification type*

            In the sentence given below underline the word which is an adjective :

            I like black pen.

   (*v*)   *Analogy type*

            Stamen : Carpel : : Pollen : _____ .

   (*vi*)   *Classification variety* (odd man out).

Which of the following animals can be classified as different from others ?   **Notes**

(a)   Pigeon   (b)   Bat   (c)   Crow   (d)   Sparrow

(b)   SAQ

The following types of SAQ can be developed for use in a question paper. Most of the questions are verbal in nature and are developed using appropriate directional words and different language patterns like the following :

(*i*)   *Interrogative format*

SAQ can be developed in verbal form ending with a question mark, using such words as what, how, why, where, when, which etc.

Question : Why does a toy ship float in water but not an iron needle ?

Question : What is the difference between a partial and total parasite ?

(*ii*)   *Statement format*

Question is in the form of a statement, followed by a directional word indicating nature of response expected.

Question : Religious policy of Aurangzeb was the major cause of downfall of Mughal empire. Give two arguments to support this statement.

(*iii*)   *Numerical format*

Most of the mathematical problems and some in Physics, Chemistry and other subjects involve the use of number or symbols, as in the following question.

Question : Calculate or find out the density of a substance whose mass is 40 and volume is 10 *cc*.

(*iv*)   *Tabular format*

Some content areas are better tested economically for testing time by requiring examiners to first fill in the table, as shown below.

	Enzyme	*Acts on*	*to produce*
1.	_____	starch	_____
2.	_____	fat	_____

(*v*)   *Graphic format*

We may use different forms of graphs like line graph, pie graph, bar graph etc.

Question : What does this graph indicate about the population trend ? (Graph is given)

(*vi*)   *Figural or diagrammatic format*

Here either the question demands the production of a diagram or a diagram is used as a means to frame a question like the following.

Question : Draw a labelled diagram showing 3 stages of working of a force pump.

Question : Observe the given diagram of 3-bean experiment and explain why do we use seed A and seed C ?

In accordance with the blueprint, stipulated number and types of SAQ are to be prepared. It is assumed that while developing the blueprint care has been taken to insert fixed-response (VSQ) and free-response (SAQ) questions in appropriate columns and rows that suit best the content of the unit and the assessment objectives. VSQ are more difficult to frame than SAQ, because they demand specificity of the response that must be fixed and objectively scorable.

(a)   *VSQ*

Such questions can be used when

(*i*)   testing of specific technical terms, concepts, events, principles, formulae, laws etc. is required;

(*ii*) complete scoring objectivity is to be ensured;

(*iii*) more questions are to be set to encompass more topics or content elements;

(*iv*) more valid responses through supply-type responses are preferred;

(*v*) multiple-choice questions are to be combined with VSQ in one section of the question paper, ensuring maximum objectivity in scoring;

(*vi*) examinees are not well conversant with the multiple-choice questions; and

(*vii*) assessment objectives are not of higher order like application or analysis objective.

(b) *SAQ*

SAQ of free-response type can better be used when :

(*i*) freedom of response is to be provided instead of a fixed response;

(*ii*) more than one credit point are to be tested, say 2 to 4 (not unlimited);

(*iii*) brief, precise and to-the-point answers are expected;

(*iv*) acceptable answers can be scored quite objectively;

(*v*) a functional and accurate marking scheme is desired;

(*vi*) more number of questions are to be included in the test for better sampling of content to increase reliability;

(*vii*) each and every competency must be tested in a limited time;

(*viii*) bluffing, redundant and superfluous responses are to be discouraged; and

(*ix*) a particular content area or element is neither amenable to essay-type nor multiple-choice items.

Thus we find that SAQ have indeed a great potentiality for use in a question paper. They can ensure more validity and reliability of a test by adequate sampling of content and competencies, ease of construction, better scoring objectivity and inclusion of more number of questions in the test.

29.3 Principles of Construction Short Answer Type Test

Very Short Answer Questions

Most of the paper setters have wrong notion that the easiest questions to be constructed are the very short-answer questions (VSQ). Make any statement or an interrogative sentence asking for a word or phrase that is quite specific, factual of trivial and you get a question. Such questions are certainly easy to construct when you are interested in testing rote memory, factual information and textual statements based on knowledge objective. If VSQ are to be, developed to test comprehension or understanding of concepts, it requires a great skill. Still more difficult and sometimes impossible to write a VSQ testing higher objective like Application. It is indeed a tough, difficult and time-consuming exercise to develop such questions of VSQ type. That is why one would seldom find entry of VSQ under application objective in a blueprint prepared even by experienced subject specialists trained in item writing. In fact in actual practice such items have usually low difficulty level. Nevertheless, if principles listed below are followed, the quality of VSQ can be improved to a large extent.

Specific hints are given below.

(a) Direct question forms evoke more favourable response from students compared with statement or incomplete stem with fill-in-the-blank type.

(*i*) What is the chemical formula of hydrogen peroxide ?   (Better)

(*ii*) Chemical formula of hydrogen peroxide is ____ .   (Poor)

(b) Only key and significant words should be omitted in blanks to pinpoint the answer.

(*i*) Adjective is a ____ word that qualifies the noun.   (Poor)

(*ii*) Adjective is a describing word that qualifies a ____ .   (Better)

(c)   It should admit only one correct fixed response.

    (*i*)   Name one mammal. (Poor) _____ Multiple response

    (*ii*)   Name one mammal that flies. (Better) _____ Fixed response

(d)   Specify the intended response through proper wording.

    (*i*)   Who put forward the theory of Organic evolution ?   (Poor)

    (*ii*)   Write the name of the biologist who put forward the theory of Inheritance of acquired characteristics?   (Better)

(e)   Use unambiguous language in wording of questions.

    (*i*)   An animal that has hair on its body and suckles its young is a _____ .   (Poor)

    (*ii*)   An animal that has hair on its body and suckles its young can be classified as a _____.
                                       (Better)

    (*iii*)   In the first question answer could be a man, tiger, dog, cow etc.

(f)   Use concise language in wording the question.

    (*i*)   When a plant cell is put in salt water its cell contents are plasmolysed due to the process of _____.   (less concise)

    (*ii*)   Cell contents of a plant are plasmolysed due to the process of _____ .

                                               (more concise language)

(g)   Avoid using textbook phrases and stereotypes.

    (*i*)   The higher we go the _____ it is.   (Poor)

    (*ii*)   Shimla is cooler than Chandigarh because its altitude is _____ than of Chandigarh. (Better)

(h)   Avoid indefinite statements

    (*i*)   When was Mahatma Gandhi born ?   (Poor)

    (*ii*)   In which year Mahatma Gandhi was born ?   (Better)

(i)   Avoid grammatical clues (and other clues).

    (*i*)   Group of tissues form an _____ (organ). It gives article clue 'an'

    (*ii*)   An organ is formed by a group of _____ (tissues).   (Better)

(j)   Develop key that contains all acceptable answers.

    (*i*)   Which method is used for carrying ripe fruits to long distances?

Here acceptable responses could be refrigeration, preservation, cooling by ice etc. Unless VSQ are used for testing the understanding of concepts, major ideas or a theme, there is no use in including such questions in a question paper just to test merely recall of specifics, trivial details, factual information, name, dates, events etc. because such questions would never evoke the exercise of intended mental processes implied in assessment objective.

Short Answer Questions

Specific guidelines for construction of SAQ (free response) include all those enunciated for VSQ. Besides, the following principles if observed by the framer of SAQ would lead to construction of questions more relevant and of better quality. List is quite long, but here only 8 major principles are listed.

1.   **Use SAQ whenever you fail to test a particular competency or specific objective with any other type of questions such as MCQ or essay type.**

    Each and every assessment objective and competency is amenable to testing through SAQ. Let us take one concept, say 'osmosis'.

    Ability to :

    (*i*)   recall   :   What is osmosis ?

    (*ii*)   translate   :   Define osmosis in your own words.

Notes	(*iii*)	illustrate	:	Give one example of osmosis from your daily life.
	(*iv*)	relate	:	Which organelle in Amoeba is concerned with osmo-regulation ?
	(*v*)	compare	:	What is the difference between endosmosis and exosmosis ?
	(*vi*)	classify	:	List two biological phenomena of endosmosis and two of exosmosis from your daily life.
	(*vii*)	detect	:	Diagram of a plasmolysed cell of Spirogyra is given in the opposite side. Detect the mistake if any and rectify. (diagram to be provided)
	(*viii*)	interpret	:	A fresh-water species of Amoeba has a contractile vacuole, but it is absent in a marine species. What does it indicate ?
	(*ix*)	explain	:	A potato chip when put in water becomes turgid. Why ?
	(*x*)	analyse	:	Species X of Amoeba when put in sea water lost its contratile vacuole (CV). On what basis can you say that it is a marine or fresh-water species ?
	(*xi*)	hypothesise	:	When marine Amoeba adapts to fresh water, it develops CV but when fresh-water Amoeba adapts to sea water it loses its CV. Propose a hypothesis to explain these observations.
	(*xii*)	suggest	:	Suggest an experimental procedure to test the above-mentioned hypothesis, (xi above)
	(*xiii*)	reason	:	When after heavy chemical manuring a grassy lawn is watered, the grass dries up after some days. Why ?
	(*xiv*)	infer	:	In which case the pickle would last longer: the one with the salt or the one without salt ? Why ?
	(*xv*)	predict	:	What will happen to a freshwater species of Amoeba when it is transferred to sea water and gets adapted ?
	(*xvi*)	judge	:	When a freshly cut potato chip is kept in salt water, it limps after a few hours because of the phenomenon of endosmosis.
				Is the statement (a) valid and (b) the reasoning given consistent ?

2. **Ensure maximum possible coverage of content elements**

Since expected response in SAQ is short, care may be taken not to test bits of information but a good chunk of content.

(*i*)   Give one characteristic of plants.   (least coverage)

(*ii*)   Give one characteristic of water plants.   (more coverage)

(*iii*) Give one characteristic of water plants different from that of desert plants.

(still more coverage)

Mark (1) remaining the same, the last SAQ is the best for syllabus coverage.

3. **Never make a SAQ completely free-response type.**

SAQ should remain a restricted variety of open-ended questions,

(*i*)   Describe uses of medicinal plants, (not SAQ)   (Poor)

(*ii*)   Write short note on allotropy.   (not SAQ)   (Poor)

(*iii*) Name the scientist who invented telephone,   (not SAQ)   (Poor)

(*iv*)   List three chemical properties of ammonia gas.   (SAQ) (Good)

4. **Indicate clearly the scope of intended answer.**

(*i*)   Give two examples of modified stems.   (More options for answering)

(*ii*)   Give two examples of underground stems.   (Clear intended scope)

5. **Use appropriate wording.**

    (*i*)   Describe three properties of carbon dioxide.   (Poor direction)

    (*ii*) List three chemical properties of carbon dioxide.   (Better direction)

6. **Allot marks in terms of credit points.**

    (*i*)   List 3 factors affecting the climate of a place. - 2 marks   (defective)

    (*ii*) List 2 factors affecting the climate of a place. - 2 marks   (functional)

7. **Use precise language and concise wording.**

    (*i*)   Write a brief description of your pet dog.   (Ambiguous)

    (*ii*) Write four sentences on your pet dog.   (Better)

8. **Write model answer.**

    This helps you to

    (*i*)   pinpoint the expected value points;

    (*ii*)   allocate proportional weightage in term of marks;

    (*iii*) take care of all acceptable answers; and

    (*iv*)   improve wording of question in conformity with the intended response.

Sometimes it become very difficult to classify a question into VSQ or SAQ, because of the flexibility in the range of responses that vary from a single specific fixed response with no flexibility to that of freedom of response and flexibility within some arbitrarily stipulated restrictions ranging from 1 to 4 or 5 value points. Being on a continuum, both varieties can be termed short-answer type.

Task   When and where to use SAQ

29.4 Editing and Reviewing of Short Answer Type Test

If we are aware of the various principles of construction of SAQ (as discussed earlier), it will not be difficult to apply those principles to check and verify whether still there is any defect that needs improvement.

The following Table 1 indicates some of the common omissions and commissions usually observed during editing of questions and their moderation.

Table 1 : Common mistakes during editing of questions

Original question	Improved version	Nature of improvement made
(a) Name any two dicot seeds.	(a) Name any two dicot seeds, one endospermic and one non-endospermic.	(a) Two content elements are tested ensuring better content coverage.
(b) List one chemical property of oxygen.	(b) List one chemical property of oxygen that differs from that of carbon dioxide.	(b) Content coverage is doubled and it tests understanding in place of factual information.
(c) Draw a right-angled triangle with 10 *cm* base.	(c) Draw a right-angled triangle with 10 *cm* base without using a compass.	(c) It becomes a thinking question, testing higher level of assessment objective.

(d) Write short note on the mechanism of photosynthesis. (SAQ)	(d) List major chemical reactions that take place during dark phase of photosynthesis.	(d) Original is not a SAQ but modified version is.
(e) Explain why does an iron needle sink in water.	(e) Why does an iron needle sink in water.	(e) Wording is made more precise by removing superfluous directional word (Explain).
(f) Besides air, list two other conditions necessary for germination of a seed.	(f) List two conditions necessary for germination of a seed.	(f) Instructional aside, the lead (air) is removed.
(g) Describe briefly 3 theories of organic evolution. (SAQ)	(g) Name three theories of organic evolution.	(g) More appropriate directional word is used to make the question SAQ; the original one was ETQ.
(h) Describe main features of Theory of Natural Selection.	(h) List three characteristic features of Theory of Natural Selection.	(h) Scope of expected response is made more clear by appropriate wording.
(i) Explain various modifications of stem with the help of examples.-4 marks	(i) Describe any two stem modifications giving one example of each.-4 marks	(i) More relevant directional word. (ii) Scope of response is delimited. (iii) More precise language used.
(j) Draw 3 lines of symmetry in the given figure.-2 marks	(j) Draw 2 lines of symmetry in the given figure.-2 marks	(j) Credit points reduced to 2 to conform to marks that are not divisible by value points. (now 2)

Did u know? It is only for convenience and practical reasons that this dichotomy of two types of questions, VSQ and SAQ, is introduced. This would further ensure quality of each type of questions, and its advantage can be taken of in consolidation of questions into a particular format leading to more than one section for separate administration.

29.5 Advantages of Short Answer Type Test

The short-answer-type tests are considered better than essay type tests ways. These have the following merits :

1. **Possible to Set Questions on the Whole Curriculum :** Short answer type questions are very short in comparison to essay type questions. The time taken to answer one essay type questions can suffice to answer 4-6 short answer type and 8-10 very short answer type questions. Thus, questions can be asked on the whole course.

2. **More Valid than Essay Type Tests :** Tests are generally used to measure change in the cognitive, affective and conative aspects. Though short answer type tests can measure only cognitive aspects, yet they so extensive that these are more valid than essay type tests.

3. **More Reliable than Essay Type Tests :** The answers to short answer type questions are generally definite, as such their measurement is objective and it is generally the case. It is evident that these tests are more reliable in comparison to essay type tests.

4. **More Objective than Essay Type Tests :** Although caution may be exercised in the construction of essay type questions and evaluation of answers, but they remain subjective to a great extent. However, the answers to the short answer type questions are generally definite, so their evaluation too is often objective; at least much more as compared with essay type tests.

5. **More Comprehensive than Essay Type Tests :** In short answer type tests questions can be framed on almost the whole curriculum, so these tests are more comprehensive than essay-type tests, and accordingly more valid and reliable.

6. **More Discriminatory than Essay type Tests :** More number of questions can be asked in these tests, almost on the whole curriculum, so the students have to memories and understand the whole curriculum. The more a student is able to understand, the more is he able, to score and higher is his rank. Thus, the examinees can be classified on the basis of these tests.

7. **Simple Administration :** These tests are administered just like essay type tests. The personnel who administer the essay type tests can also administer them.

29.6 Disadvantages of Short Answer Type Test

These tests are constructed in order to do away the demerits of essay type tests, and they have succeeded in some measure, but they do not-possess the qualities of the essay type tests. The non-possession of these qualities is the shortcoming of these tests. These tests have the following demerits or shortcomings :

1. **Inability to Measure Skill and Attitude :** The short answer type tests often provide complete facts but not skill and attitude. Moreover, the students are not fully free to answer the questions, so their attitudes cannot be measured.

2. **Inability to Measure Linguistic Skill and Power of Expression :** The answers to the short answer type questions have to be given briefly, so the students' linguistic style and power of expression cannot be measured by these tests as proficiently as can be done by essay type tests.

3. **Inability to Measure Higher Mental Faculties such as Logic :** Short answer type questions are often factual in nature. In answering them, the students have to present neither reasoning nor examples in favour of their opinion, so the faculties of logic and decision etc. cannot be measured by them.

4. **Inability to Measure Insight and Foresight :** How far the insight and foresight of the students have been developed in a particular field can be measured by critical and problematic questions, and it is not possible to ask such questions in short answer type tests, in which the students have to comprehensively think about a problem and suggest of their own. Then how can the insight and foresight of the students be measured by them?

5. **Less Helpful to Measure Personality :** First, hardly any questions related to students' interest and attitude are asked in such tests; secondly, the answers to these questions are generally definite and the students have no freedom to answer them in their own. So, the personality of the students cannot be measured by them.

6. **Emphasis on Rote Learning :** In short answer type tests, emphasis is often laid on facts by which knowledge is chiefly tested. So, the students are inclined to learn by rote.

7. **Neither fully Valid, nor Reliable, nor Objective :** It is true that the questions on almost the whole course may be asked in short answer type tests and it is also true that their evaluation is more objective than the essay type questions, but not fully objective. Evidently, these tests are neither fully valid nor reliable nor objective.

Self Assessment

1. *Fill in the blanks :*

(*i*)   The questions which can be answered even more briefly and in shorter time than the short answer type questions are called _____ .

(*ii*)   The answers to short answer type questions are generally _____ , as such their measurement is _____ .

(*iii*)   Short answer type test can measure only _____.

(*iv*)   _____ and _____ can not be measured by short answer type test.

(*v*)   In short answer type test, emphasis is often laid on facts by which _____ is chiefly tested.

29.7 Summary

- The tests in which such questions are asked, which can be answered briefly are called short answer type tests. Short answer type questions are generally of two types–short answer type and very short answer type. We are illustrating the both types here.

- Short Answer Type Questions : The questions which can be answered in comparatively less time and in less words than the essay type questions are called short answer type questions.

- Very Short Answer Type Questions : The questions which can be answered even more briefly and in shorter time than the short answer type questions are called very short answer type questions.

- VSQ
  - Direct question
  - Selection form
  - Completion form
  - Identification type
  - Analogy type
  - Classification variety (odd man out)

- SAQ
  - Interrogative format
  - Statement format
  - Numerical format
  - Tabular format
  - Graphic format
  - Figural or diagrammatic format

- It is assumed that while developing the blueprint care has been taken to insert fixed-response (VSQ) and free-response (SAQ) questions in appropriate columns and rows that suit best the content of the unit and the assessment objectives. VSQ are more difficult to frame than SAQ, because they demand specificity of the response that must be fixed and objectively scorable.

- Most of the paper setters have wrong notion that the easiest questions to be constructed are the very short-answer questions (VSQ). Make any statement or an interrogative sentence asking for a word or phrase that is quite specific, factual of trivial and you get a question.

- Specific guidelines for construction of SAQ (free response) include all those enunciated for VSQ. Besides, the following principles if observed by the framer of SAQ would lead to construction of questions more relevant and of better quality. List is quite long, but here only 8 major principles are listed.

- Use SAQ whenever you fail to test a particular competency or specific objective with any other type of questions such as MCQ or essay type.

  - Ensure maximum possible coverage of content elements
  - Never make a SAQ completely free-response type.
  - Indicate clearly the scope of intended answer.
  - Use appropriate wording.
  - Allot marks in terms of credit points.
  - Use precise language and concise wording.
  - Write model answer.
- The short-answer-type tests are considered better than essay type tests ways. These have the following merits :
  - Possible to Set Questions on the Whole Curriculum
  - More Valid than Essay Type Tests
  - More Reliable than Essay Type Tests
  - More Objective than Essay Type Tests
  - More Comprehensive than Essay Type Tests
  - More Discriminatory than Essay type Tests
  - Simple Administration

29.8 Keywords

Numerical   :   relating to numbers

Tabular   :   Presented or arranged in a table

Analogy   :   The process of comparing one thing with another thing that has similar features in order to explain it.

29.9 Review Questions

1. What is the meaning and type of short answer type test ?
2. When and where to use Short answer Question ?
3. Give the format of short answer type test ?
4. What are the principles of Construction of short and very short answer type question
5. Explain the merits and demerits of short answer type test.

Answers : Self Assessment

1. (*i*)   very short answer type test   (*ii*)   definite, objective,
    (*iii*)   Cognitive aspect   (*iv*)   linguistic style, logic faculties
    (*v*)   knowledge

29.10 Further Readings

Books

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*
2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*
3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*
4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 30 : Essay Type Test: Advantages and Limitations

CONTENTS

Objectives

Introduction

30.1 Meaning : Essay Type Test

30.2 Extended and Restricted Response Essay Type Test

30.3 Principles of Constructing Essay Type Test

30.4 Grading of Essay Type Test

30.5 Improving in Reliability of Scoring Essay Type Test

30.6 Advantages of Essay Type Test

30.7 Limitations of Essay Type Test

30.8 Summary

30.9 Keywords

30.10 Review Questions

30.11 Further Readings

Objectives

The objectives of this unit can be summarised as below :

- To know about the meaning of essay type test

- To discuss about the extended and Restricted Response of Essay type test.

- To explain the principles of Constructing Essay type test

- To describe the grading of essay type test

- To discuss about the improving in Reliability of scoring Essay type test

- To explain the advantages and limitations of essay type test

Introduction

For efficient-testing, use of ETQ at secondary and university stages is not only necessary but also the most difficult exercise for paper setters. In spite of their frequent use in testing, they are the ones that are criticised the most for lack of their validity and reliability. Taylor of Guwahati University and Edwin Harper of Ewing Christian College, Allahabad have done tremendous research work on the validity and reliability of grading and other aspects of ETQ, highlighting the inadequacies of these questions. But this does not mean that these questions have little potential of measuring the educational outcomes. Regarding their relevance in daily life, adults and students do encounter situations where they have to respond in a manner one likes to express, explain or react, like listening and reciting stories, communicating and receiving messages from others, presenting new points in a summarised way etc. Therefore no system of examinations can ignore this reality of assessing students on skills and abilities that are necessary for disciplining the life of every one of us.

30.1 Meaning : Essay Type Test

An essay test is an assessment technique that requires students to thoroughly respond to a question or prompt by developing, organizing, and writing an original composition. The purpose of an essay test is to assess students abilities to construct a logical, cohesive, and persuasive writing piece.

30.2 Extended and Restricted Response Essay Type Test

Most of the higher-order abilities demand integration of intra-disciplinary and inter-disciplinary subject-mater. Critical analysis of content, text and reports, creative synthesis for developing plans, hypotheses and set of principles or abstractions, making critical appraisals, and production of evaluation reports can better be tested more validly by using ETQ alone. Therefore it must be admitted that in spite of their inherent weaknesses in grading and objective scoring, their role in testing higher-order competencies cannot be over-emphasised.

The nomenclature of long-answer questions suits better for content subjects like history, geography, biology etc. and of essay-type questions for language subjects. However, both are used here interchangeably. Since ETQ is more commonly accepted nomenclature, this term is used for further discussion. To appreciate the difference between the two types of ETQ, let us examine the following questions demanding responses ranging from unlimited to restricted responses, depending upon the nature and scope of the expected response.

(*i*)   Write an essay on India's struggle for independence.

(*ii*)   Discuss in details the various theories of organic evolution.

(*iii*) Describe various allotropic forms of carbon.

(*iv*)   In what way Kalinga war influenced King Asoka ?

(*v*)   Describe the laboratory method of preparing oxygen gas.

(*vi*)   Write in about 150 words an essay on 'A Village Fair'.

In questions (i) to (iii) the freedom of response is almost unlimited and examinees can write-as long response as they like. Such questions are classified as extended-response questions and ETQ. But in questions (iv) to (vi) there is restriction by limiting the scope of content tested or the length of the expected answer. Freedom of writing, organising and presenting the answers still remains with the examinees.

Since restriction is imposed and examinees cannot be expected to write extended responses, these types of questions are termed restricted-response, essay-type questions. The paper-setter has to set or can set both types of questions depending upon maturity level of the students. Extended-response variety suits better at higher levels like the colleges or university, whereas the restricted-response variety of ETQ is favoured more for secondary and senior secondary stages. Moreover, it lends to less subjectivity in marking compared with extended-response questions. In the former an examinee is free to select, organise, integrate, evaluate and express in any way he deems appropriate.

When thorough understanding of a set of topics or units is the objective of testing and best presentation of the subject-matter is to be appreciated, this variety of extended response of ETQ is the best to take advantage of. But if the focus of measurement of learning outcomes is on interpretation, application of data or outcomes that are more specific and that clearly define the nature of the intended response and attempt to reduce subjectivity in marking, the restricted response variety of ETQ can be preferred. ETQ can also be classified on the basis of directional words used in such questions. Weidman (1941) categorised such tests into 11 types of questions ranging from the simple to the most complex using directional words like (i) what, when, who,

which, where; (ii) list; (iii) outline; (iv) describe; (v) contrast; (vi) compare; (vii) explain; (viii) discuss; (ix) develop; (x) summarise and (xi) evaluate. Likewise, Monroe and Carter divided ETQ into 21 categories, ranging from selective recall to ...cause and effect...application of principles...criticism... new procedure to inferential thinking.

30.3 Principles of Constructing Essay Type Test

General Hints

(*i*)   Use ETQ where you must and avoid where you can.

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

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

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

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

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

Specific Hints

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

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

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

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

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

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

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

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

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

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

30.4 Grading of Essay Type Test

Effectiveness of ETQ depends largely on how well they can be graded. A poorly developed ETQ

cannot be salvaged by the most carefully devised method of grading. Improper grading, however can spoil the well-constructed essay question. Unreliable grading is one of the most valid criticisms against ETQ. We should aim at optimum reliability only and try to reduce intra-examiner and inter-examiner variances. Scoring uniformity and scoring objectivity should be ensured while setting the question and developing the marking scheme. The following two methods can be used in grading.

30.4.1 Analytical Method

It is also called the 'Point score method', in which a model answer is broken into specific, value points each of which is given proportional weightage in the allotment of marks.

30.4.2 Global Scoring Method

It is also called 'Holistic or rating method' Ideal answer is split into component parts. Students' responses are graded on the quality continuum, which has a number of standards or anchor points. Rater reads the response, forms a general impression and assigns a rating, using some standard on quality continuum. Response can be graded on a 3- or 5-point scale like excellent, good, average, satisfactory and poor. Such scale values can be established by preparing a variety of answers corresponding to the various scale points. The reader after rapid reading can assign scripts into five piles. For more objectivity two independent raters can assign independently. In homogeneous group, assign the scripts into one of the piles. Each question can then be re-read to re-classify the scripts that are classified. Outline answers and detailed marking scheme; when prepared along with framing of questions it helps improve wording, unrealistic time, difficulty level etc. This method is more suitable to apply in restricted-response variety of ETQ. The extended-response variety of ETQ can be scored by this method more reliably for the same reason.

Self Assessment

1. *Multiple choice Questions :*

   (*i*) The nomenclature of essay type question test is best for _____.

      (a) language subjects   (b) Practical subjects

      (c) Mathematical subjects   (d) Science Subjects

   (*ii*) Weidman (1941) Categorised essay type test into _____ types of questions ranging from simple to the most complex using directional words.

      (a) 9   (b) 10

      (c) 11   (d) 12

   (*iii*) Monroe and Caster divided Essay type test into _____ Categories ranging including selective recall, new procedure to inferential thinking, cause and effect.

      (a) 11   (b) 21

      (c) 31   (d) 41

   (*iv*) Analytical method is called _____.

      (a) Point score method   (b) Holistic method

      (c) Rating method   (d) Scientific method

30.5 Improving in Reliability of Scoring Essay Type Test

The following measures may be adopted to control the reliability of ETQ :

(a) Use precise, unambiguous and understandable language to enable the examinees to think about the expected scope of answer in the same way as intended by the framer of the question.

(b)   Reshuffle answer scripts after marking each question to reduce the carryover effect.

(c)   Use double grading when the test is being used for selection or awarding scholarship.

(d)   Marks of the scripts may be totalled at the end after marking all the scripts to avoid carryover of any impression formed about the quality of the script, when the marks are totalled.

(e)   Avoid contamination of scores by avoiding extraneous factors like spelling, handwriting, punctuation, neatness etc.

(f)   To reduce inter-examiner reliability, examiners may meet together after marking some scripts to discuss the marking scheme and compare marking.

(g)   Allocation of scripts question-wise to each examiner improves scoring objectivity, because the examiner scores the same question in all the scripts. Of course, provision of separate sheets for each question is costly.

Despite various limitations, ETQ have come to stay in spite of their low validity and reliability. It is neither possible nor desirable to dispense with essay-type questions. In spite of significant role they play in educational testing, steps taken to improve ETQ have never been commensurate with their importance and use at the secondary and university stages of education. Even in developed countries the money spent on improving the quality of ETQ is discouraging when compared with objective tests. The reason is the facility of computerisation of results and ease of application of psychometric devices for interpretation and improvement. If similar efforts are made to develop question-banks of ETQ and training courses are organised to imitate teachers, paper setters and markers, the stigma of subjective marking, low scorer reliability and grading difficulties can be washed away to a large extent.

Notes   Well-designed ETQ with a well-structured scoring procedure could enable the teachers to grade students' responses more objectively, besides improving their validity and reliability.

30.6 Advantages of Essay Type Test

There are following advantages of Essay type test

(*i*)   Such tests induce good study habits in the pupils.

(*ii*)   The guess work can be eliminated to large extent.

(*iii*)   Such tests are easier to prepare and administer.

(*iv*)   By this type of tests it is possible to measure all degrees of comprehensiveness and accuracy.

(*v*)   Such tests can be used by all type of schools.

(*vi*)   These help in developing the po0wer of logical thinking, critical reasoning, systematic presentation etc. in the students.

(*vii*)   Such tests provide an opportunity to the child to show his initiative, originality of thought, fertility of their imagination, etc.

(*viii*)   This type of tests are considered to be best for measuring ability to organise idea effectively, ability to criticise or justify a statement, ability to interpret etc.

30.7 Limitations of Essay Type Test

At present new type of tests are taking the place of traditional or Essay type of tests. This trend is gaining importance because traditional type of examinations suffer from following defects.

(a)   Defects from the Point of View of Students

(*i*)   The essay type tests are less objective and so they lack of validity. This type of test can reveal child's cramming capacity only.

(*ii*) These tests lack reliability. A student is compelled to have a selective reading. He depends more on guess papers and so there is an element of chance.

(*iii*) It keeps the students busy and fall of nervous tension. The study does not spread over the whole year and is limited to a short period just before the examinations. Thus a habit of irregular study is developed in the student.

(b)   Defects of Essay-type Tests from the Point of View of the Teacher

(*i*)   The teacher covers only a limited and important portion of course because his aim is to see that maximum number of his students pass the examination.

(*ii*) The teaching programme of the teacher is wholly examination oriented and the basic principle of teaching his students are given least consideration.

(*iii*) The teacher is compelled to encourage his students to cramming which is not a psychological method of teaching.

(*iv*) Since a teacher is judged by the results of his students so everything becomes subservient to the examinations.

(*v*)   To show good results sometimes the teacher devotes a good deal of his time to indulge in guess work which affects his teaching

(c)   Defects from the Point of View of Achievement

(*i*)   Essay type tests are not comprehensive and some students may get good marks only because the questions have been set from the portion prepared by them.

(*ii*) These tests are not objective and the score of a student depends on various factors such as Examiners mood and whims etc.

(*iii*) This type of tests is not useful from the point of view of improvement. They fail to throw light on the defects of teaching-learning process or the defects of the curriculum.

From the above it can be concluded that essay type examination is not a correct method of evaluation in commerce. The improvement in system of evaluation is possible if following suggestions are given due consideration.

Task   What is analytical method of grading in essay type questions ?

Self Assessment

2. Fill in the blanks :

(*i*)   Essay type tests provide an opportunity to the child to show his initiative, _____, fertility of their imagination.

(*ii*) Essay types tests are quite Commonly used for _____ purposes.

(*iii*) Essay type tests are less _____ and so they lack of _____.

(*iv*) In essay type examination an improvement is possible if _____ are spread over the whole course.

30.8 Summary

- An essay test is an assessment technique that requires students to thoroughly respond to a question or prompt by developing, organizing, and writing an original composition. The purpose of an essay test is to assess students abilities to construct a logical, cohesive, and persuasive writing piece.

- Use ETQ where you must and avoid where you can.

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

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

- Pinpoint the specific assessment objectives, which becomes the basis of your question.

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

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

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

- Effectiveness of ETQ depends largely on how well they can be graded. A poorly developed ETQ cannot be salvaged by the most carefully devised method of grading. Improper grading, however can spoil the well-constructed essay question.

- It is also called the 'Point score method', in which a model answer is broken into specific, value points each of which is given proportional weightage in the allotment of marks.

- It is also called 'Holistic or rating method' Ideal answer is split into component parts. Students' responses are graded on the quality continuum, which has a number of standards or anchor points. Rater reads the response, forms a general impression and assigns a rating, using some standard on quality continuum. Response can be graded on a 3- or 5-point scale like excellent, good, average, satisfactory and poor.

- Such tests induce good study habits in the pupils.

- The guess work can be eliminated to large extent.

- Such tests are easier to prepare and administer.

- By this type of tests it is possible to measure all degrees of comprehensiveness and accuracy.

- Such tests can be used by all type of schools.

- At present new type of tests are taking the place of traditional or Essay type of tests. This trend is gaining importance because traditional type of examinations suffer from following defects.

- Defects from the Point of View of Students

- Defects of Essay-type Tests from the Point of View of the Teacher

- Defects from the Point of View of Achievement

30.9 Keywords

- Extend : To make an idea cover more areas or operate in more places
- Restrict : To control something with rules or laws
- Essay : a short piece of writing by a student as part of a course of study.

30.10 Review Questions

1. What is the meaning of Essay type Questions ?

2. What are the specific principles of constructing Essay type test ?

3. Give the methods of Grading essay type Questions.

4. What measures should be adopted to control the reliability of Essay type test ?

5. What are the advantages and limitations of Essay type tests of examinations ?

Answers : Self Assessment

1. (*i*)   a   (*ii*)   c   (*iii*)   b   (*iv*)   a

2. (*i*)   Originality of thought   (*ii*)   Evaluation

   (*iii*)   Objective, Validity   (*iv*)   Shorter Questions

30.11 Further Readings

Books

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 31 : Formative and Summative Evaluation

CONTENTS

Objectives

Introduction

31.1 Formative Evaluation

31.2 Summative Evaluation

31.3 Summary

31.4 Keywords

31.5 Review Questions

31.6 Further Readings

Objectives

The objectives of this unit can be summarised as below :

* To discuss about the Formative Evaluation
* To describe about the Summative Evaluation

Introduction

Formative assessment is designed to assist the learning process by providing feedback to the learner, which can be used to highlight areas for further study and hence improve future performance. Self and diagnostic assessment are types of formative assessment with specific purposes. Summative assessment is for progression and/or external purposes, given at the end of a course and designed to judge the students' overall performance.

Formative evaluation can serve an important developmental or capacity-building purpose, for the organisation as a whole and for individual members, where it is seen as a form of organisational learning.

Formative evaluation is time and labour intensive in comparison to most forms of summative evaluation. It relies primarily on qualitative methods that are heavy in their use of time and evaluation expertise, both at the data gathering stage as well as in analysis. Depending on the audience for the formative evaluation findings, the reliance on qualitative methods may fail to meet the expectations of some stakeholders for robust quantitative measures of progress.

31.1 Formative Evaluation

Formative evaluation seeks to strengthen or improve a programme or intervention by examining, amongst other things, the delivery of the programmer, the quality of its implementation and the organisational context, personnel, structures and procedures. As a change oriented evaluation approach, it is especially attuned to assessing in an ongoing way, any discrepancies between the expected direction and outputs of the programme and what is happening in reality, to analysing strengths and weaknesses, to uncovering obstacles, barriers or unexpected opportunities, and to generating understandings about how the programme could be implemented better. Formative evaluation is responsive to the dynamic context of a programme.

Formative evaluation pays special attention to the delivery and intervention system, but not exclusively. In formative evaluation, the evaluator also has to analyse the intervention logic, the outcomes, the results and impacts.

> *Notes* Formative evaluation activities include the collection and analysis of data over the lifecycle of the programme and timely feedback of evaluation findings to programme actors to inform ongoing decision-making and action (i.e. it is a form of operational intelligence). It requires an effective data collection strategy, often incorporating routinised monitoring data alongside more tailored evaluation activities.

Feedback is primarily designed to fine-tune the implementation of the programme although it may also contribute to policy-making at the margins through piecemeal adaptation.

Evaluators conducting a formative evaluation ask many different kinds of questions and use a variety of methods to address them. Questions are commonly open-ended and exploratory, aimed at uncovering the processes by which the programme takes shape, establishing has changed from the original design and why, or assessing soft organisational factors such as the extent of 'buy in' by practitioner staff to the programme's goals and intended outcomes.

Formative evaluation questions also investigate the relationship between inputs and outcomes, which can involve the formulation and measurement of early or short-term outcome measures. These often have a process flavour and serve as interim markers of more tangible longer term outcomes.

Methods which might be used include stakeholder analysis, concept mapping, focus groups, nominal group techniques, observational techniques and input-output analysis. Formative evaluation's concern with the efficiency and effectiveness of project management can be addressed through management - oriented methods like flow charting, PERT/CRM (Programme Evaluation and Review Technique and Critical Path Method) and project scheduling. The measurement of interim or short-term outcome measures, which capture steps in the theory of how change will be achieved over the long term, may involve construction of qualitative or process indicators and use of basic forms of quantitative measurement.

Formative evaluation may be planned and managed in a variety of ways. The prevailing practice has been to prioritise the information needs of central staff (policy makers, programme managers) as those primarily responsible for programme steerage, leaving unspecified the roles that local staff (local site managers, local practitioners) and clients can play in reshaping plans and strategies in response to feedback. Newer conceptions of formative evaluation (for example, the mutual catalytic model of formative evaluation outlined by Chacon-Moscoso et al, 2002) emphasise a more inclusive approach to the involvement of stakeholders, and as well seek to elicit their participation as collaborators in the evaluation process rather than simply as providers of information. The role of evaluator changes from one concerned with gathering data and communicating evaluation findings to one of engaging programme participants in a form of evaluative inquiry. Organisational actors are helped to generate their own data and feedback through collective learning processes.

> *Notes* Formative evaluation lends itself most readily to a case study approach, using a qualitative mode of inquiry. There is a preference for methods that are capable of picking up the subtleties of reforms and the complexities of the organisational context and wider policy environment.

31.1.1 Need for Formative Evaluation

Many commentators would argue that all Structural Fund initiatives operate in conditions of uncertainty, and that formative evaluation is a desirable corrective or steerage component of all

programmes. Formative evaluation is particularly relevant to programmes whose goals and objectives cannot be well specified in advance, are open to interpretation by actors at different levels of the system, or which seem likely to change over the lifetime of the programme. In many of the newer EU programmes, the objective is to introduce changes in the innovative behaviour of companies and regions and to launch a process of building up collective learning. Formative evaluation can be a driver of, and contributor to, the organic learning and knowledge creation processes that exist within regions and networks and should itself be understood as a developmental process.

Formative evaluation has most relevance at the *ex ante* and *mid-term* phases, and indeed some programmes evolve continuously, never reaching a stage of being finished or complete. Formative evaluation activities may be extended throughout the life of a programme to help guide this evolution. *Post-ante* evaluations may draw on evidence from formative evaluation although their primary focus is summative.

Formative evaluation is ideally built into the programme design as an ongoing activity rather than inserted into a particular phase. It may however take a particular form at different stages of the evaluation lifecycle.

At the needs assessment stage in an *ex ante evaluation*, formative evaluation can determine who needs the programme, how great the need is, and what might work to meet the need.

Formative evaluation can inform evaluability assessment. Working with funders, programme managers, staff and participants in the early stages of clarifying goals and strategies making them realistic and evaluable, establishing how much consensus there is among goals and interventions and where the differences lie constitutes the essential groundwork for a formative evaluation. Evaluability assessment becomes an improvement-oriented experience that leads to significant programme changes and shared understandings, rather than just being seen as a planning exercise preparing for summative evaluation.

Formative evaluation follows the lifecycle of the initiative through implementation, tracking the fidelity of the programme to goals and objectives, investigating the process of delivery, diagnosing the way the component parts of the programme come together and reinforce or weaken one another, and addressing problems as they emerge. Programme implementation is in large part about ongoing adaptation to local conditions. The methods used to study.

31.1.2 Purposes of the Technique

Large scale, medium to long term socio-economic programmes are often designed and implemented in dynamic, fluid contexts characterised by imperfect information, changing policy agendas and goal posts, unpredictable environmental conditions and moving target groups of intended beneficiaries. Formative evaluation is a strategy for dealing with a context of this kind. It starts from the premise that no matter how comprehensive and considered the programme design, it will invariably require steerage and possibly redirection, and will be considerably strengthened by opportunities for stakeholder reflection on what is working, what is not going to plan, and what kinds of changes need to be made. Formative evaluation is *prospective* in orientation, and conceived within a continuous cycle of information gathering and analysis, dialogue and reflection, and decision-making and action. It has commonalities with forms of evaluative inquiry that draw on organisational learning models and processes, giving it a strong developmental focus for the organisation as a whole and for organisational members.

Formative evaluations that are inclusionary and participative, involving local programme actors as active contributors and participants in the evaluation process, bring pragmatic benefits in addition to enhancing professional development and organisational capacity. Including staff as collaborators is likely to facilitate the collection not only of more reliable data, but of data that are actively used to improve daily programme activities at the local level.

Formative evaluation can help to strengthen horizontal structures and processes by creating and fostering feedback mechanisms and for a, enabling lessons to be shard. It can cultivate much thinker networks of professional and informal contacts between levels of decision-making through facilitating intra- and inter- organisational dialogue and learning.

Formative evaluation can also have important catalytic effects, mobilising staff around a course of action, and engaging management thinking about future options. Patton introduced the idea of 'process use' to describe the utility to stakeholders of being involved in the planning and implementation of an evaluation, irrespective of findings and recommendations that occur. The developmental and capacity building benefits accrue to staff as a side effect of a participative, formative evaluation.

Although formative evaluation is commonly contrasted with summative evaluation, the distinction is not always helpful or apposite.

That can inform summative assessment. In some programme contexts, a more fruitful approach would be to see both types of evaluation as part of the same exercise.

31.1.3

Formative evaluation includes several evaluation types :

*   **needs assessment** determines who needs the program, how great the need is, and what might work to meet the need

*   **evaluability assessment** determines whether an evaluation is feasible and how stakeholders can help shape its usefulness

*   **structured conceptualization** helps stakeholders define the program or technology, the target population, and the possible outcomes

*   **implementation evaluation** monitors the fidelity of the program or technology delivery

*   **process evaluation** investigates the process of delivering the program or technology, including alternative delivery procedures

> *Did u know?* The *process* of formative evaluation may be an important component in summative evaluation; formative evaluation can produce early outcome measures which serve as interim markers to programme effects; and by tracking changes and linkages between inputs, outputs and outcomes it can help to identify causal mechanism.

31.1.4 Process of Formative Evaluation

The main steps involved

Step 1 : A first step is gaining the commitment of key stakeholders and programme actors at all levels to a formative evelution as a collective learning and change-oriented process. This may require among other things negotiation about access and the use of information, clarification of roles and relationships, and agreement about what kinds of information will be relevant for which kinds of stakeholders.

Step 2 : Building evaluation into programme design so that it is perceived as an essential tool for managing the programme and helping it to adapt to local conditions within a dynamic environment. This might include laying the basis for formative evaluation in the early stages of needs assessment and evaluability assessment, as well as embedding formative evaluation into ongoing organisational processes and structures. Successful formative evaluation depends on the early adoption of an effective data collection strategy and in many cases a management information database which allows evaluators and programme staff easy access to well organised programme information.

Step 3 : Creating an evaluation infrastructure to support formative evaluation as a learning, change-oriented, developmental activity. This includes working with programme staff on an ongoing basis to :

- create a culture that supports risk-taking, reduces fear of failure, and values lessons learned from mistakes

- establish channels of communication that support the dissemination of information and allow organisational members to learn from one another in ways that contribute to new insights and shared understandings

- create new opportunities for shared learning and knowledge creation

- modify systems and structures that inhibit organisational learning

Step 4 : A fourth step entails finding out about the decision-making cycle, the different stakeholder groups and their respective information needs and interests. These might include policy makers and programme makers at central level, local site programme managers, and operational staff. Each set of stakeholders will be asking different questions of the evaluation and have a preference for the way that findings are presented and or/communicated. Where there is a lack of appropriate mechanisms or opportunities for feedback, the evaluator will need to establish a structured way to provide relevant stakeholders with feedback.

Step 5 : Formative evaluation involves an ongoing cycle of data gathering and analysis. The choice of methods will be determined largely by the questions being addressed, and the methodological preferences of different stakeholders. Most formative evaluations use a variety of methods. Where a collaborative, participative approach is taken to formative evaluation, the methods are likely to include those which foster and support interaction, dialogue, learning and action.

Step 6 : There are different views as to whether the evaluator's responsibility stops with feeding back findings and facilitating processes of learning among programme actors, or whether she or he also has a role to play in follow-through action. Where the evaluator is external to the organisation, the role is likely to be limited to the former. Formative evaluators may however be internally located, especially where the preferred model of formative evaluation is influenced by organisational learning concepts and practices. In these circumstances, the formative evaluation cycle is likely to include shared responsibility for implementing the action plan and monitoring its progress.

31.1.5 Strengths and Limitations

Formative evaluation provides a rich picture of a programme as it unfolds. It is a source of valuable learning not just prospectively for the programme but for future programmes as well. Formative evaluation is highly complementary to summative evaluation and is essential for trying to understand why a programme succeeds or fails, and what complex factors are at work. Large scale programmes are often marked by a discrepancy between formal programme theory and what is implemented locally. Formative evaluation can help determine whether the substantive theory behind the programme is flawed, whether the evaluation was deficient, or if implementation failed to pass some causal threshold.

To be effective and achieve its purpose of programme improvement, formative evaluation requires strong support from the top as well as bottom-up support. It must be endorsed by programme decision-makers and others who will need to act on its findings. Support may be withdrawn, overtly or covertly, if the findings expose weaknesses in programme design or implementation, especially where the organisational culture is one of blame and discourages innovation or learning from mistakes. Research findings suggest that programme managers are more receptive to 'bad news' that is communicated by internally located evaluators ('one of us'), than by independent evaluators.

Formative evaluation can serve an important developmental or capacity-building purpose, for the organisation as a whole and for individual members, where it is seen as a form of organisational learning.

Formative evaluation is time and labour intensive in comparison to most forms of summative evaluation. It relies primarily on qualitative methods that are heavy in their use of time and evaluation expertise, both at the data gathering stage as well as in analysis. Depending on the audience for the formative evaluation findings, the reliance on qualitative methods may fail to meet the expectations of some stakeholders for robust quantitative measures of progress.

Task   What do you understand by the term "process use" ?

Self Assessment

1. Multiple Choice Questions :

Choose the correct option :

(*i*) _____ evaluation seeks to strengthen or improve a programme by examining, amongst other things, the delivery of the programme, structures and procedures.

    (a) Summative   (b) Formative

    (c) External   (d) Internal

(*ii*) _____ evaluation monitors the fidality of the program or technology delivery.

    (a) Implementation   (b) Process

    (c) Need   (d) Structured

(*iii*) _____ helps stakeholders define the program or technology, the target population, and the possible outcomes

    (a) need assessment   (b) Process evaluation

    (c) Structured Conceptulization   (d) External Evaluation

(*iv*) _____ assessment determines whether an evaluation is feasible and how stakeholders can help shape its usefulness.

    (a) evaluability   (b) Structured Conceptulization

    (c) Implementation   (d) process

31.2 Summative Evaluation

Summative evaluation looks at the impact of an intervention on the target group. This type of evaluation is arguably what is considered most often as 'evaluation' by project staff and funding bodies- that is, finding out what the project achieved.

Summative evaluation can take place during the project implementation, but is most often undertaken at the end of a project. As such, summative evaluation can also be referred to as *ex-post* evaluation (meaning after the event).

Summative evaluation is often associated with more objective, quantitative methods of data collection. Summative evaluation is linked to the evaluation drivers of accountability. It is recommended to use a balance of both quantitative and qualitative methods in order to get a better understanding of what your project has achieved, and how or why this has occurred. Using qualitative methods of data collection can also provide a good insight into unintended consequences and lessons for improvement.

Summative evaluation is not about stating that three workshops were held, with a total of fifty people attending (outputs), but rather the result of these workshops, such as increased knowledge or increased uptake of rainwater tanks (outcomes).

31.2.1 Need for a Summative Evalation

There are some key reasons for using a summative evaluation :

- Summative evaluation provides a means to find out whether your project has reached its goals/objectives/outcomes.

- Summative evaluation allows you to quantify the changes in resource use attributable to your project so that you can track how you are the impact of your project.

- Summative evaluation allows you to compare the impact of different projects and make results-based decisions on future spending allocations (taking into account unintended consequences).

- Summative evaluation allows you to develop a better understanding of the process of change, and finding out what works, what doesn't, and why. This allows you to gather the knowledge to learn and improve future project designs and implementation.

31.2.2 Characteristics

1. It is the procedure to assess or grade educators' level of learning in certain period of time.

2. It tends to use well defined evaluation designs (i.e. fixed time and content).

3. It provides descriptive analysis (i.e. in order to give a grade, all the activities done throughout the year are taken into account). 4. It tends to stress local effects.

5. It is unoppressive and not reactive as far as possible.

6. It is positive, tending to stress what students can do rather than what they cannot.

31.2.3 Summative evaluation can also be subdivided

- **outcome evaluations** investigate whether the program or technology caused demonstrable effects on specifically defined target outcomes

- **impact evaluation** is broader and assesses the overall or net effects - intended or unintended - of the program or technology as a whole

- **cost-effectiveness and cost-benefit analysis** address questions of efficiency by standardizing outcomes in terms of their dollar costs and values

- **secondary analysis** reexamines existing data to address new questions or use methods not previously employed

- **meta-analysis** integrates the outcome estimates from multiple studies to arrive at an overall or summary judgement on an evaluation question

Notes Summative evaluation is outcome-focused more than process focused. It is important to distinguish outcome from output.

Summative Evaluation is carried out at the end or conclusion of a programme of instruction. It is very useful to determine whether or not the broad aims, objectives or goals of the programme have been achieved. Summative evaluation is judgemental in nature and is offered as class tests, examinations set up like termly, yearly and external examinations like 'o' levels, 'A' levels. The examinations used to design course grades and certify the learners.

31.2.4 Advantages of Summative Evaluation

Advantages of summative evaluation is judgemental in nature, hence it shows the worth or the value of a programme.

It can be used for decision making so that learners can easily selected into placement streams such as pupils are put into classes according to their test results.

Summative evaluation is useful for communicating with other stakeholders like parents, NGOs or church organisation who need to assist in the education sector.

It is a useful tool for guidance and counselling. Above all it is done to evaluate the effectiveness of teachers.

31.2.5 Disadvantages of Summative Evaluation

Disadvantages of summative evaluation comes rather too late at the end when learning difficulties can no longer be rectified. It is only concerned with the end results.

Did u know? Formative assessment is most appropriate where the results are to be used internally by those involved in the learning process (students, lecturers, learning support etc.), whilst summative assessment is most appropriate to succinctly communicate students' abilities to external interested parties.

Self Assessment

2. Fill in the blanks :

(*i*) _____ is associated with more objective, quantitative methods of data collection.

(*ii*) _____ evaluation is broader and assesses the overall or not effects.

(*iii*) _____ analysis reexamines existing data to address new questions or use methods not previously employed.

(*iv*) Summative evaluation is _____ in nature and it is used for _____ .

31.3 Summary

- Formative evaluation seeks to strengthen or improve a programme or intervention by examining, amongst other things, the delivery of the programmer, the quality of its implementation and the organisational context, personnel, structures and procedures.

- Formative evaluation pays special attention to the delivery and intervention system, but not exclusively. In formative evaluation, the evaluator also has to analyse the intervention logic, the outcomes, the results and impacts.

- Formative evaluation may be planned and managed in a variety of ways. The prevailing practice has been to prioritise the information needs of central staff (policy makers, programme managers) as those primarily responsible for programme steerage, leaving unspecified the roles that local staff (local site managers, local practitioners) and clients can play in reshaping plans and strategies in response to feedbacks.

- Formative evaluation has most relevance at the *ex ante* and *mid-term* phases, and indeed some programmes evolve continuously, never reaching a stage of being finished or complete. Formative evaluation activities may be extended throughout the life of a programme to help guide this evolution.

- Formative evaluation is ideally built into the programme design as an ongoing activity rather than inserted into a particular phase.

- Formative evaluation can inform evaluability assessment. Working with funders, programme managers, staff and participants in the early stages of clarifying goals and strategies making them realistic and evaluable, establishing how much consensus there is among goals and interventions and where the differences lie constitutes the essential groundwork for a formative evaluation.

- Large scale, medium to long term socio-economic programmes are often designed and implemented in dynamic, fluid contexts characterised by imperfect information, changing policy agendas and goal posts, unpredictable environmental conditions and moving target groups of intended beneficiaries. Formative evaluation is a strategy for dealing with a context of this kind. It starts from the premise that no matter how comprehensive and considered the programme design, it will invariably require steerage and possibly redirection, and will be considerably strengthened by opportunities for stakeholder reflection on what is working, What is not going to plan, and what kinds of changes need to be made.

- Formative evaluation can help to strengthen horizontal structures and processes by creating and fostering feedback mechanisms and for a, enabling lessons to be shard. It can cultivate much thinker networks of professional and informal contacts between levels of decision-making through facilitating intra- and inter- organisational dialogue and learning.

- Formative evaluation includes several evaluation-types :

  - needs assessment determines who needs the program, how great the need is, and what might work to meet the need

  - evaluability assessment determines whether an evaluation is feasible and how stakeholders can help shape its usefulness

  - structured conceptualization helps stakeholders define the program or technology, the target population, and the possible outcomes

  - implementation evaluation monitors the fidelity of the program or technology delivery

  - process evaluation investigates the process of delivering the program or technology, including alternative delivery procedures

- The main steps involved

  - Step 1 : A first step is gaining the commitment of key stakeholders and programme actors at all levels to a formative evaluation as a collective learning and change-oriented process.

  - Step 2 : Building evaluation into programme design so that it is perceived as an essential tool for managing the programme and helping it to adapt to local conditions within a dynamic environment. This might include laying the basis for formative evaluation in the early stages of needs assessment and evaluability assessment, as well as embedding formative evaluation into ongoing organisational processes and structures.

  - Step 3 : Creating an evaluation infrastructure to support formative evaluation as a learning, change-oriented, developmental activity. This includes working with programme staff on an ongoing basis to :

  - Step 4 : A fourth step entails finding out about the decision-making cycle, the different stakeholder groups and their respective information needs and interests. These might include policy makers and programme makers at central level, local site programme managers, and operational staff.

  - Step 5 : Formative evaluation involves an ongoing cycle of data gathering and analysis. The choice of methods will be determined largely by the questions being addressed, and the methodological preferences of different stakeholders.

  - Step 6 : There are different views as to whether the evaluator's responsibility stops with feeding back findings and facilitating processes of learning among programme actors, or whether she or he also has a role to play in follow-through action.

- Formative evaluators may however be internally located, especially where the preferred model of formative evaluation is influenced by organisational learning concepts and practices.

- Formative evaluation provides a rich picture of a programme as it unfolds. It is a source of valuable learning not just prospectively for the programme but for future programmes as well.

- Formative evaluation is highly complementary to summative evaluation and is essential for trying to understand why a programme succeeds or fails, and what complex factors are at work. Large scale programmes are often marked by a discrepancy between formal programme theory and what is implemented locally.

- Summative evaluation looks at the impact of an intervention on the target group. This type of evaluation is arguably what is considered most often as 'evaluation' by project staff and funding bodies- that is, finding out what the project achieved.

- Summative evaluation can take place during the project implementation, but is most often undertaken at the end of a project. As such, summative evaluation can also be referred to as *ex-post* evaluation (meaning after the event).

- Summative evaluation is often associated with more objective, quantitative methods of data collection.

- Summative evaluation provides a means to find out whether your project has reached its goals/objectives/outcomes.

- Summative evaluation allows you to quantify the changes in resource use attributable to your project so that you can track how you are the impact of your project.

- Summative evaluation allows you to compare the impact of different projects and make results-based decisions on future spending allocations (taking into account unintended consequences).

- Characteristics

  (*i*) It is the procedure to assess or grade educators' level of learning in certain period of time.

  (*ii*) It tends to use well defined evaluation designs (i.e. fixed time and content).

  (*iii*) It provides descriptive analysis (i.e. in order to give a grade, all the activities done throughout the year are taken into account).

  (*iv*) It tends to stress local effects.

- outcome evaluations investigate whether the program or technology caused demonstrable effects on specifically defined target outcomes

- impact evaluation is broader and assesses the overall or net effects - intended or unintended - of the program or technology as a whole

- cost-effectiveness and cost-benefit analysis address questions of efficiency by standardizing outcomes in terms of their dollar costs and values

- secondary analysis reexamines existing data to address new questions or use methods not previously employed

- meta-analysis integrates the outcome estimates from multiple studies to arrive at an overall or summary judgement on an evaluation question

- Summative Evaluation is carried out at the end or conclusion of a programme of instruction. It is very useful to determine whether or not the broad aims, objectives or goals of the programme have been achieved. Summative evaluation is judgemental in nature and is offered as class tests, examinations set up like termly, yearly and external examinations like 'o' levels. 'A' levels. The examinations used to design course grades and certify the learners.

Notes

- Advantages of summative evaluation is judgemental in nature, hence it shows the worth or the value of a programme.

- It can be used for decision making so that learners can easily selected into placement streams such as pupils are put into classes according to their test results.

- Summative evaluation is useful for communicating with other stakeholders like parents, NGOs or church organisation who need to assist in the education sector.

- It is a useful tool for guidance and counselling. Above all it is done to evaluate the effectiveness of teachers.

- Disadvantages of summative evaluation comes rather too late at the end when learning difficulties can no longer be rectified. It is only concerned with the end results.

31.4 Keywords

Formative   :   Having an important and lasting influence on the development of something or somebody's character.

Evaluation   :   The act to form an opinion of the amount value or quality of something.

Initiative   :   A new plan for dealing with a particular problem

31.5 Review Questions

1. What is formative evaluation ?

2. Explain the process of formative evaluation.

3. Give the different types of formative evaluation.

4. Discuss the advantages and disadvantages of formative evaluation.

5. Explain the Concept of Summative evaluation why is the need for this type of evaluation.

6. Give the types, advantages and disadvantage of summative evaluation.

Answers : Self Assessment

1. (*i*)   b   (*ii*)   a   (*iii*)   c   (*iv*)   a
2. (*i*)   Summative evaluation   (*ii*)   Impact
   (*iii*)   Secondary   (*iv*)   judgemental, decision making

31.6 Further Readings

Books

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*

Unit 32 : Measurement of Attitude, Aptitude, Personality and Intelligence

CONTENTS

Objectives

Introduction

32.1 Measurement of Attitude

32.2 Measurement of Aptitude

32.3 Measurement of Personality

32.4 Measurement of Intelligence

32.5 Summary

32.6 Keywords

32.7 Review Questions

32.8 Further Readings

Objectives

The objectives of this unit can be summarised as below :

*   To discuss about the measurement of attitude
*   To explain about the measurement of aptitude
*   To describe about the measurement of personality
*   To discuss the measurement of intelligence

Introduction

The mental feeling of a person towards an object, person, activity or thought is generally known as his attitude towards that object, person, activity or thought. Different persons have different attitudes towards different objects, persons, activities and thoughts. It is because of the fact that these attitudes among them are developed through their own experiences with those objects, person, activities and thoughts. With the change of experience regarding an object, person, activity and thought the attitude towards it is also changed. What a person experiences about an object, person, activity or thought, he develops similar attitude towards that and he reacts to it accordingly in the future.

The term *aptitude* is used in connection with the tests that predict future performance or status on the basis of an individual's current status, whereas a test of *ability* measures an individual's current status. Even intelligence tests, achievement tests as also the interest inventories may also be considered as aptitude tests when they are used to predict future behaviour. When factor analysis of abilities is undertaken, it produces clusters of special abilities that considerest central to performing task or a job. It is these clusters of special abilities that are called *aptitudes*.

Developments in the field of intelligence are the result of efforts made in the assessment of intelligence. Modern mental testing has its roots in the works of a Frenchman, J. Esquirol (1938), on mental disorders. Sir Galton, known as the father of mental tests, paved the way for developments in the measurement of intelligence. According to him, simpler, sensory, perceptual and motor responses were the key aspects of intelligence.

The first systematic attempt to develop intelligence tests was made by Alfred Binet. He attempted to identify children who did not benefit from regular class-room teaching and had to be placed in special schools. Binet and T. Simon developed the first test of intelligence in 1905. The core of intelligence was considered to be more complex mental processes such as memory, imagery, comprehension and judgment.

The term *personality* is derived from the Latin word *persona*, the mask used by actors performing on the stage in ancient Greece and Rome. Many professionals use the term as dynamic rather than descriptive. A dynamic orientation provides theoretical basis for a majority of the performance and projective tests of personality and self-concept. Theories of personality arose in the context of need of physicians for a rational procedure to deal with their patients having psychological problems. Personality and behaviour theories underwent rapid developments, leading to convergence of the two theoretical formulations.

32.1 Measurement of Attitude

32.1.1 Meaning and Definition of Attitude

It is clear that an attitude is not hereditary, a person develops it on the basis of his experiences. Another fact in this regard is that a person can form positive or negative feelings toward an object, person, activity or thought. Thus, an attitude towards an object, person, activity or thought may be positive or negative. **Thurston** has defined an attitude as a measure of positive and negative thought factor. In words :

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

— *Thurston*

Freeman has defined attitude the following words :

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

—*Freeman*

We can define attitude more clearly in the following words :

Attitudes are the mental feelings of a person towards an object, person, activity or thought which are developed by his own experiences regarding that object, person, activity or thought and which direct him to react to these accordingly.

In order to understand the concept of attitude clearly, it is necessary to differentiate it from other similar concepts.

32.1.2 Nature and Characteristics of Attitude

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

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

(2) Attitude can be either positive or negative.

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

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

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

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

(7) Attitudes influence the personality of a person.

32.1.3 Types of Attitudes

Attitudes are generally classified as follows :

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

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

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

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

32.1.4 Measurement

Measurement of attitude has been undertaken since ancient times in our country, and we have been dividing it into attitude of *sat* (goodness), attitude' of *raj* (passion), and attitude of *tam* (darkness). However, the history of measurement of attitude towards an object, person, activity or thought is not much long. Measuring attitude in this way was started by **Thurston** in 1927. Later several psychologists developed different techniques. These techniques are classified into two types— Behavioural techniques and Psychological techniques.

(I) Behavioural Techniques

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

(*i*) **Method of Direct Observation :** In this method measurement of attitudes of a person is done by observing his daily routine. How does a person behave towards an object, person, activity or thought, is observed by the tester. On the basis of this observation, his attitude towards that object, person, activity or thought is divided into suitable, unsuitable and indefinite categories; for example, if a person sings patriotic songs, sees patriotic movies and reads patriotic literature, then his attitude towards patriotism is expressed.

But this method has its own limitations. Obtaining results from this method is a time-consuming process. If the person comes to know that he is being observed, he can possibly hide his undesirable activities and can show off good activities. Besides, it becomes difficult, as well as impossible in some cases, to observe behaviour of all people.

(*ii*) **Method of Direct Questioning :** In this method, the measurement of attitudes of a person is done by asking questions to him directly. In it, the person is asked questions about an object, person, activity or thought and his advice is taken. This advice can be favourable, unfavourable or indefinite. The advice of a person on some national problem or planning is obtained by asking questions to him directly.

In this method, the attitude of a person can be measured correctly only when such person is not subject to any pressure and he is willing to respond truthfully. It happens in a few cases only, so the results obtained from this method are often not valid.

Notes   In this method of direct observation, the tester observes the behaviour and activities of a person directly and determines his attitude on its basis, so its results are quite valid.

(II) Psychological Techniques

In order to eradicate the shortcomings of the behavioural techniques of measuring attitudes, the psychologists have developed various other methods. Chief among them are—Thurston's Method of Paired Comparison, Thurston and Chave Method of Equal Appearing Intervals, Likert Method of Summated Rating, Suffir's Method of Successive Intervals, Guttman's Scalogram Method, and Scale Discrimination Technique. An introduction to them is presented here :

(*i*) **Thurston's Method of Paired Comparison :** In this method, some statements are prepared about a psychological object. These statements are such prepared as to express qualities from lowest level to the highest. These statements are in an ordered sequence for all possible values or positions of a psychological quality. This order may be—extremely like, much like, less like, indifferent, less dislike, much dislike, extremely dislike. Thus, each statement is scaled in a psychological order of continuum. After this, these statements are presented to the person in pairs, and he is asked whether he agrees or not, or likes or not. The number of total pairs can be known by the following mathematical formula :

$$\text{Number of figures} = \frac{n(n-1)}{2}$$

In which : n = number of total statements

For example, if there are a total of 10 statements in an attitude scale, then the number of comparative pairs will be :

$$\frac{n(n-1)}{2} = \frac{10 \times (10-1)}{2} = \frac{10 \times 9}{2} = 45$$

In this method, a person has to compare two statements at the same time, and has to give his decision about both of them.

(*ii*) **Method of Equal Appearing Intervals :** There are faced some problems in measuring attitude by method of paired comparison. The chief difficulty is that the number of statements is quite large and the possible number of pairs are many, and it is difficult to make F-matrices, P-matrices and Z-matrices by taking a person's advice on them. It consumes time and labour. Besides there is possibility of calculation errors. Considering these aspects, **Thurston** and **Chave** developed a simple method of measuring attitude. In this method, if the number of statements is large, then it can be used in attitude scale. The following steps are followed while constructing attitude scale :

(1) At first many attitude statements are prepared and then they are administered.

(2) These administered statements, with the help of specialists, are expressed into 11 grades. These 11 grades are situated on almost equal interval on the psychological continuum. We can present them as follows :

Unfavourable attitude---				Neutral attitude---			Favourable attitude---			
A	B	C	D	E	F	G	H	I	J	K

The most unfavourable statements are placed under A while the most favourable statements are placed under K. All the statements are classified into 11 classifications on the basis of the specialists' advice for construction of attitude.

(*iii*) **Method of Summated Rating :** In both the methods used by Thurston, the statements are scaled according to the advice of the specialists; then passing through a long process the attitude scale is constructed and then the attitude of a person is measure. **Lickert**,

developed a new method for the construction of attitude scale for measuring attitude. This method is based on a simple and clear process. The name of this method or technique is—Method of Summated Rating. In this method, the selection of statements for the construction of attitude scale is done on the basis of item analysis.

(1) At first, 50-60 favourable or unfavourable statements are constructed for the object, person, activity or thought for which the attitude has to be measured.

(2) Five responses for each statement are prepared to give marks to each favourable or unfavourable statement—Strongly agree, Agree, Undecided, Disagree and Strongly disagree. When a person chooses either of them, corresponding mark (out of 4, 3, 2, 1, and 0) is given for favourable statements as decided above. Now the mark weightage of the unfavourable statements is changed, that is, on an extremely favourable response for an unfavourable statement is made 0, 1 mark for agree response, and 2, 3 and 4 marks for undecided, disagree and extremely disagree responses respectively.

(3) The scale is administered on a large sample of those persons for which the attitude scale is being prepared. Each of them is asked to give his consent to one of the responses to each statement.

(4) The responses of the people are given marks as 0, 1, 2, 3, 4 for positive statements and 4, 3, 2, 1, 0 for negative statements, and the total of obtained scores (attitude score) is found out.

(5) The response sheets are arranged in descending order and the people falling within 27% from each of top and bottom are selected.

(6) The value of *t* is found out on the basis of the difference between the discriminating powers of top and bottom groups with the help of items-analysis process. The medians and standard deviations of the obtained scores obtained by the persons of top and bottom groups are calculated to find out the value of *t*. The value of *t* is found out by the following formula for comparison of medians of the top and bottom groups :

$$t = \frac{M_H - M_L}{\sqrt{\dfrac{\sigma_H^2}{n_H} + \dfrac{\sigma_L^2}{n_L}}}$$

In which, M_H = Median of the top group for obtained scores of each statement

M_L = Median of the bottom group for obtained scores of each statement

σ_H = Standard deviation of the top group for each statement

σ_L = Standard deviation of the bottom group for each statement

n_H = Number of people in the top group

n_L = Number of people in the bottom group

n = Number of people in each group

If the number in both groups is equal, then $n_H = n_L = n$ can be used. In this context, the formula of *t* will be as follows :

$$t = \frac{M_H - M_L}{\sqrt{\dfrac{\sigma_H^2 + \sigma_L^2}{n}}}$$

(7) For meaningfulness of value of *t*, either of 0.05 or 0.01 meaningful levels are chosen. If the value of *t* is meaningful, the statement is accepted as a statement of discriminating power. If the value of *n* is large, that is more than 30, then it can be taken as meaningful at 0.05 level on its value being more than 1.65, and at 0.01 on it being more than 2.33. If the value of *n* is small, it would need to refer to the *t* values table for seeing its meaningfulness. In the following table number—4, the values of *t* have been shown for its meaningfulness at .05 and .01 levels. The value of degree of freedom is known by the following formula for comparison of two groups :

$$d_f = n_H + n_L - 2$$

In which, d_f = Degree of freedom

n_H = Number of people in the top group

n_L = Number of people in the bottom group

Table

t-Values for One tailed Test

dt	.05	.01
10	1.81	2.76
20	1.72	2.53
30	1.70	2.46
40	1.68	2.48
50	1.68	2.40
60	1.67	2.39
70	1.67	2.38
80	1.66	2.38
90	1.66	2.37
100	1.66	2.36
200	1.65	2.35
300	1.65	2.34
400	1.65	2.34
more than 500	1.65	2.33

(8) On the basis of meaningful value of *t*, 30-40 statements are selected. Care is taken in selection of statements to see that some of the statements are positive and some of them negative.

(9) The selected statements are arranged in a random order to prepare the attitude scale.

(10) Having prepared the attitude scale, the reliability and validity of the scale is found out using Test-retest method or Split-half method.

(11) Thus prepared attitude scale is administered on those people whose attitude has to be measured. Their scores on various statements of the scale are obtained according to the above mentioned marking process. The obtained scores of a person on all statements will indicate his attitude.

(12) For analysis of obtained scores, they can be changed into *t*-values or percentiles. The constructor of the attitude scale can also prepare a conversion table for it.

(*iv*) **Suffir's Method of Successive Difference :** The method of successive difference developed by **Suffir** is an improved form of Thurston's and Chave's Equal Appearing Intervals method. Its scale statements need the decisions of only one person. So, when the number of statements is large, this method is appropriate. In this method, the

psychological continuum is decided at first. Having ordered psychological continuum in 11 point, 9 point or 7 point in various degrees, the scaled values of the statements are found out with the help of cumulative ratios.

After this, attitude scale is prepared using the process of the Equal Appearing Method. This attitude scale is used for measuring attitude. The method of its administration and scoring is similar to that of the Equal Appearing Method.

(*v*) **Guttman's Scalogram Method :** This method is quite different from other methods. In the attitude scale prepared by this method, the statements concerned with the general subject are in pairs. The more are the marks in form of attitude, the more favourable will be the responses of a person in comparison to others. In fact, this method is not a method to prepare an attitude scale, but is a method to evaluate an attitude scale.

In this method, the subject field of attitude is selected. The statements are constructed on the selected subject. The statements are selected by item analysis or by other statistical methods. According to **Guttman**, if the number of statements is large, the statements can also be selected by experience or insight. The responses on the statements are obtained in two degrees—agree and disagree. Guttman has instructed to award 1 mark for 'agree' and 0 marks for 'disagree'. On a psychological object, 1 mark is given for more or favourable attitude and 0 for less or unfavourable attitude. The obtained scores on the attitude scale statements determine the attitude. The total sum of obtained scores on the scale statements indicates the attitude of a person.

(*vi*) **Scale Discrimination Technique :** This method was developed by **Edward** and **Kilpatric**. In it, the merits of all methods have been included. The following steps are followed in the preparation of attitude scale in this method :

(1) The statements, pertaining to a psychological object towards which attitude has to be measured, are prepared or gathered in a large number and are administered.

(2) Then the scaled values and Q-values of the statements are found out sing Equal Appearing Interval method. About 50% statements with Q-value more than quartile deviating are cancelled and rest of the 50% statements are retained for further analysis.

(3) After this, the 50% selected statements are presented according to the Lickert's Method. The statements are divided into six degrees instead of five—Strongly agree, Agree, Mildly agree, Mildly disagree, Disagree and Strongly disagree. Its marking technique is determined as 0, 1, 2, 3, 4, 5 as according to Lickert's Method. The obtained scale in the form of all statements is administered on a large group and discriminating power on each statement is obtained using Item Analysis method. Edward and Kilpatric have used Phi-coefficient for finding out discriminating power. To find out phi-coefficient, 27% persons from each of top and bottom are selected. The following formula is used to find out Phi-coefficient:

$$\varphi = \frac{bc - ad}{\sqrt{(a+b)(b+d)(a+c)(c+d)}}$$

Here for calculation of a, b, c and d, the persons of the top group obtaining 0, 1, 2 marks on each statement are considered to be high class. In the same way, the number of persons obtaining 0, 1, 2 marks on each statement are considered low and those obtaining 3, 4, 5 marks are considered high. Thus, obtained number of four classes is the value of a, b, c and d. The following table makes it all the more clear :

32.1.5 Administration and Scoring of Attitude Tests

The administration of all attitude tests is almost similar but their scoring is done differently. We are presenting the administration and scoring of a test.

32.1.6 Attitude Test

1. **Name of the Test :** XYZ School attitude scale

2. **Constructor of the Test :** ABC

3. **The Quality to be Measured :** Attitude towards school

4. **Utility of the Test :** School reform

5. **Items of the Test :** There are a total of 30 items (statements) in this test, among which some are positive and others are negative. In this test, the attitude towards the school has been divided into 5 areas. These areas are :

    (*i*) Attitude towards teachers

    (*ii*) Attitude toward school

    (*iii*) Attitude towards school programmes (curriculum, homework, and co-curricular activities, etc.)

    (*iv*) Attitude towards social acceptability (mutual relations with other students)

    (*v*) Attitude of the guardians as indicated by the students towards school.

    The test contains statements for expressing attitude in each field. The responses of the students are obtained in five degrees for each statement. These degrees are — Always, Most often, Frequently, Sometimes, and Never. On the positive statements, the responses are awarded as 5 for always, 4 for most often, 3 for frequently, 2 for sometimes and 1 for never. In the same way, the negative statements are awarded 1 for always, 2 for most often, 3 for frequently, 4 for sometimes, and 5 for never. The positive and negative statements are known by the following table :

    Table

Positive statements	1, 2, 7, 9,11, 13, 14, 15, 17, 18, 21, 22, 23, 25, 27
Negative statements	3, 4, 5, 6, 8, 10, 12,16, 19, 20, 24, 26, 28, 29, 30

6. **Administration of the Test :** This test is administered on the students of class VIII to XI. The test is administered in the following sequence :

    (*i*) The seating arrangement for the students is made at a place where there is no disturbance and noise. The space between any two students is kept as such that they are not able to talk to each other.

    (*ii*) Then rapport is established with them to bring them in normal mental condition.

    (*iii*) Then the objective of the test is made clear to them.

    (*iv*) Then the students are given instructions pertaining to the test.

    (*v*) Then the test papers and answer-sheets are distributed among them.

    (*vi*) Now they are asked to read the instructions. Any of their doubts is clarified with the help of the test manual.

    (*vii*) Now they are asked to write their answer. No time limit is given, but the time taken by each student is noted down.

7. **Scoring of the Test :** The scoring of each positive or negative response as indicated by each student in the test is done by the above decided method. On the basis of Obtained scores by each student, it can be said what is his attitude or advice towards the school.

8. **Analysis of Obtained Scores in the Test :** The analysis of obtained scores in this test is done on the basis of attitude scores. At first the attitude of the students is measured on the basis of total obtained scores. If the attitude of the student is positive, then it can be said that such a

school is suitable for him, he is satisfied with the school environment (physical and educational). If the attitude of the student is not suitable for school, his response on the different areas of the attitude scale is seen. On the basis of these responses reforms are made in the areas in which he has not given favourable response or has not shown positive attitude.

32.1.7 Use, Need and Importance of Attitude Test

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

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

(3) Favourable attitudes of the students are developed.

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

32.1.8 Limitations of Attitude Test

The attitudes of the students are subject to change, so the limitations of attitudes measured at one time cannot be applied at another time.

Task   What is scale discrimination technique

32.2 Measurement of Aptitude

32.2.1 Meaning and Definition of Aptitude

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

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

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

Aptitude is that hereditary ability, capacity or potential of a person which develops by environment, teaching and training, and which helps a person to get success in a particular field.

In order to understand the concept of aptitude clearly, it is necessary to differentiate it from other similar concepts.

32.2.2 Nature and Characteristics of Aptitude

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

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

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

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

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

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

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

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

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

32.2.3 Measurement

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

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

We have defined aptitude as hereditary abilities, capacities and potentialities. So the aptitude tests too should be defined in this form :

An aptitude test is such a test which measures a person's or a group of persons hereditary abilities, capacities or potentialities, and on its basis guides for creation of environment, teaching and training work in the future.

32.2.4 Types of Aptitude Test

The aptitude tests can be divided into three types keeping in view their nature

1. **General Aptitude Test :** These tests are used to measure general abilities; such as general intelligence, mental, ability, general creativity, learning ability, etc.

   Success of a person in a specific field can be predicted on the basis of these tests. And if specific training and teaching is arranged in such field, abilities and potentialities of a person can be developed as predicted.

   Besides, these tests have their own limitations. These tests only tell whether a person is intelligent or not, creative or not, quick learner or not; but they do not tell in which direction the intelligence of the person works faster, or in which field his creativity is inclined — in the field of art, science, literature or any other field.

2. **Differential Aptitude Test :** In these tests, there is a series of sub-tests within a test to measure some specific abilities. These are also called **'batterv-tvpe'tests.** These tests are used to measure specific abilities.

   Every sub-test which is included in these tests measure specific aptitude in some measure. Differentiation in the abilities of a person can be made on the basis of obtained scores in these tests. The differential aptitude tests measure different aspects of a general ability to express different aptitudes on the basis of obtained scores. On this basis, a specific ability of a student can be estimated and his skill can be enhanced by teaching and training. For example, take mental ability test of Thirston. It has seven sub-tests and each of these measures different aspects of mental ability. These are — spatial ability, perceptual ability, numerical ability, verbal ability, word fluency ability, memory ability and reasoning ability. On the basis of marks obtained in these sub-tests, it can be said about a student whether his power of reasoning is more than his verbal ability or not. In the same way, comparative analysis of other abilities can be made. Some of the examples of differential aptitude tests are — differential aptitude test (DAT), general aptitude test battery (GATB), aptitude survey (AS), aptitude classification test (ACT), and clerk aptitude test (CAT).

   These tests have their own limitations. Every person has several aptitudes. It is not possible to measure them all by anyone test. An ability can have various differentials. It is not possible to know them by these tests. Administration of these tests takes time. As such, the students may be bored up while giving tests and it can influence the obtained scores.

3. **Specified Aptitude Test :** These are the tests which are used to measure aptitude of a person or student in a specific field; for example, music aptitude test, teach in aptitude test, mechanical aptitude test, art aptitude test, etc. Various specified aptitude tests have been constructed in different fields. Such tests chiefly include foreign tests — Seashore Measures of Musical Talents, Wings Standardised Tests of Musical Intelligence, Musical Aptitude Test, Horn Art Aptitude Inventory, Meier Art Test, Groves Design Judgement Test, Minnesota Clerical Test, Test of Mechanical Comprehension. And the chief tests in the Indian tests are — Mechanical Aptitude Test, Clerical Aptitude Test, and Teaching Aptitude Test.

These tests help us know whether a student has aptitude in a specific field or not (such as in music, art, science or mechanical fields). And if he has, in what measure. On the basis of it, he is given educational and vocational guidance.

However, construction and administration of such tests is a difficult task, it requires specialists.

32.2.5 Uses, Need and Importance of Aptitude Tests in Education

The psychologists have expressed the fact that a child becomes that for which he has aptitude. If a child is given opportunity to develop according to his aptitude, he progresses leaps and bounds, and if attempt is made to develop contrary aptitudes, the child does not progress in the desired direction and fails to obtain proficiency. The uses of aptitude tests in the field of education can be serialised as follows :

(1) The aptitudes of the students are known.

(2) They are given educational and vocational guidance on the basis of obtained results.

32.2.6 Limitations of Aptitude Test

The development of a person does not depend only upon hereditary aptitudes, but upon so many other facts.

(1) First, the development of a child's aptitude depends on environment, education and training. As such, he cannot be guided alone on the basis of obtained results of the aptitude test.

(2) Second a child's development is also dependent on his intelligence, interest, inclination, ability and capacity besides his aptitude. In such a condition, he cannot be given correct guidance on the basis of aptitude tests alone.

Self Assessment

1. *Multiple Choice Questions :*

Choose the correct option :

(*i*) _____ at first developed paired Comparison technique of measuring attitude

    (a) Binet   (b) Thurston

    (c) Guttman   (d) Lickert

(*ii*) _____ developed the method of summating rating at first.

    (a) Binet   (b) Thurston

    (c) Guttman   (d) Lickett

(*iii*) _____ developed at first the method of successive difference for measurement of attitude.

    (a) Thurston   (b) Guttman

    (c) Suffir   (d) Binet

(*iv*) _____ at first developed the method of equal appearing intervals

    (a) Thurston   (b) Guttman

    (c) Suffir   (d) Thurston and Chave

(*v*) _____ has clarified the nature of aptitude.

    (a)   Bingham   (b)   Freeman

    (c)   Binet   (d)   Cruz

(*vi*) Aptitude is _____.

    (a)   Hereditary   (b)   Administrative

    (c)   Reserved   (d)   Subjective

(*vii*) Differential aptitude test is also called _____ .

    (a)   Specified aptitude test   (b)   Battery type test

    (c)   Mechanical adaptibility   (d)   Mechanical knowledge test

32.3 Measurement of Personality

32.3.1 Meaning and Definition of Personality

Personality of a person is generally perceived as his physical get up, behavioural skill and the status that he enjoys in the society. However, it is something more than this. The form of personality has been analysed chiefly by philosophers, anthropologists, sociologists and psychologists. Ancient Indian scriptures Vedas have accepted human personality as the development of physique, mind and soul. In the Upanishads, personality is considered to be the sum total of physical self (Annamaya Kosa and Pranamaya Kosa), mental self (Manomaya Kosa and Vijnanamaya Kosa) and spiritual self (Anandamaya Kosa and Atman). In Yoga philosophy, personality is considered to be the sum total of body, mind, intellect, ego and soul. According to Indian philosophy, the more a person's body, mind, intellect, ego and soul have developed, the more developed his personality is. However, today we are neither able to understand the form of mind, intellect, ego and soul, nor are able to measure their development, so we are not able to understand the concept of personality in this form.

Did u know? Cattell first selected 4,500 traits out of the 18,000 traits given by Allport. After this, he removed the synonyms and selected 200 traits. At last, he selected only 35 traits after factor analysis. On this basis, his theory is also called the **Factorial System Theory.**

32.3.2 Nature and Characteristics of Personality

The psychologists are not in agreement regarding the concept of personality. How can then the nature of personality be described in a definite form ! However, the different views of different psychologists regarding personality and the facts that they have found out about it may be presented as the nature of personality or characteristics of personality.

1. **Personality is a Multi-Dimensional Concept :** A person's personality comprises of both his innate powers and acquired traits. The important among them are : physical and mental health, self-consciousness, will power, sociability, adjustability, and developmental continuity. The more factors a persons has the more developed his personality is.

2. **The Permanent Traits and Abilities of a Person are the Part of His Personality :** Permanent traits are those qualities which are always expressed in one's behavous. And permanent abilities mean those abilities which a person always makes use of. It has been seen that persons often assume some traits according to time and circumstances which they forsake with the change in time and circumstances. The psychologists do not consider such traits as part of their personality. The psychologists consider only those traits as part of a person's personality with which he is always attached and which are always expressed in his thought and behaviour.

3. **Personality is the Integrated Structure of Different Traits and Abilities :** The personality of **Notes**
    a person is not the sum total of his innate powers and acquired traits and abilities, rather it is
    a specific structure. As the sum total of bricks and cement does not make a house, rather a
    house is a specific structure made from them, in the same way, a person's personality is not
    the sum total of his traits and abilities, but is a specific structure made of them. And as the
    concept of house is in itself a complete organised structure, in the same way, the concept of
    personality is also complete and organised in itseif.

Did u know? The Indian yogic psychologists claim to increase intelligence by 30 to 45% by yogic
exercise, but doing such yogic exercise is a difficult task. Only a few persons can
perform the mediation, and the remaining persons have to depend on their hereditary
intelligence.

32.3.3 Types of Personality

Different disciplines have determined different types of personality. We shall describe only those
types which have been determined by the psychologists.

(I) *Classification on the Basis of Physical Constitution*

Some psychologists have classified personality on the basis of physical constitution. Of them,
we have already discussed the theory of physical constitution as propounded by W.H. Sheldon.
Now we present the classification done by psychologist **Kretschmer**.

1. **Asthenic :** Kretschmer has placed those persons in this group who are tall and weak
    and often like to live alone and enjoy criticising others, out do not tolerate their own
    criticism.

2. **Athletic :** Kretschmer has placed those persons in this group who are healthy sound
    and strong and have the capacity to adjust well.

3. **Pynic :** Kretschmer has placed those persons in this group who are short in height but
    healthy, easy going and entertaining.

(II) *Classification on the Basis of Nature*

Psychologist **Gallen** has classified persons on the basis of their nature into four classes :

1. **Chalenic :** Gallen has placed strong and offensive nature persons in this group. The
    persons of this group become angry soon.

2. **Melancholy :** Gallen has placed the persons suffering from anxiety in this group. Such
    persons are often melancholic, suppressed and pessimist.

3. **Phlegmatic :** Gallen has placed down hearted persons in this group. Such persons are
    often peace loving and lazy.

4. **Sanguine :** Gallen has placed the enthusiastic persons in this group. Such persons are
    optimist and active.

32.3.4 Measurement

In the present times, the analysis of personality and its measurement was started by the western
psychologists. They have developed several methods of measurement of personality. The
psychologists of our country are blindly following them. In the meantime, the methods that have
been developed for the measurement of personality, can be divided into two groups—Non-projective
methods and Projective methods. Some of the non-projective methods are subjective and someone
are objective. The chief methods under subjective methods are—Observation Method, Interview

Method, Questionnaire Method, Case History Method and Autobiographical Method. Under the objective methods, the chief methods are—Controlled Observation Method, Rating Scale, Physiological Tests and Situation Tests. The important projective methods include Association Technique, Construction Technique, Completion Technique, Ordering Technique, and Expression Technique. Under the expression Tests, the important are—Rorschach Ink Blot Test, Thematic Apperception Test (TAT) and Children Apperception Test (CAT).

(I) Subjective Methods of the Measurement of Personality

Subjective methods of measurement of personality are those methods whose measurement results are influenced by the views, likes and dislikes and personal norms of the measurer. The following are such methods :

1. **Observation Method :** Under this method, the measurer observes the behaviour to the subject and finds out the personality traits present in him.

   He also finds out the quantity or degree in which these traits are present in him-whether in excessive, normal or deficient measure. Psychologists Thorndike and Jersield used this method in the beginning.

   The most important characteristic of this method is that the measurer looks at all the characteristics (personality traits) of the subject by his own eyes and understand them by his own mind and he does not need any measuring tool.

   However, this method also has its shortcomings. First, it takes much time in the measurement of one's personality. Second, the result of measurement depends on the measurer and as such they are subjective and not objective.

2. **Interview Method :** Under this method, the measurer and subject are before each other, the measurer asks personality related questions and the subject answers them, This task can be performed in a formal way or in an informal way.

   The most important characteristic of this method is that the measurer establishes rapport with the subject, so the subject even tells the secret or confidential facts too. Second, some of the characteristics of the subject which cannot be known by other methods can be known by this method; for example, self-confidence and style of expression. This method can also be used to confirm the data gathered by other methods. This method is chiefly used to find out the problems of the subject for providing personal guidance.

   But this method consumes much time in the measurement of one personality because the measurement of the personality of one person is done at one time. It takes much time and energy. Moreover, it is a subjective method and its results are not reliable.

3. **Questionnaire Method :** It is also called **Inventory Method.** It is the formal and written form of interview method. Under this method, questions are formed on various aspects of a person's personality, such as intelligence, ability, emotions, ideals, values and sociability, etc., and they are listed.

   These lists are called personality inventory. These inventories are distributed to the subject or subjects. They are asked to answer on these inventories or on attached answer sheets.

   4. **Case History Method :** It is also called **Life History Method**. Under this method, general information about the subject from birth to the present time is collected; such as physical health, maturity, intelligence, family relations, social relations, education, personal problems, etc. The measurer gathers these informations by two means, one from the subject directly and the other from related person, such as parents, neighbours, class-fellows, etc.

   In the field of education, this method is chiefly used to find out the causes of child delinquency. In the field of medicine, the psychologists make use of this method.

But this method has shortcomings of its own. First, the subject and other related persons do not often give correct information. Second, it is a subjective method and its results depend on the personal likes, dislikes and norms. They may or may not be correct.

5. **Autobiography Method :** Autobiographical method has the following two forms :

    (*i*) **Structured Biography :** Under this method, the measurer presents a heading or draft before the subject, and asks the subject to express his views on it on the basis of his life. The condition is that the subject cannot delete something that he has written once. It is natural for this biography to express his personality traits. The measurer analyses this article to find out facts of his personality.

    (*ii*) **Unstructured Biography :** Under this method, the subject has to write his biography independently, he is not given any title, draft or guidance in this regard. The subject is not permitted to delete whatever he has written once. As a result, the subject expresses both, his conscious and unconscious. The measurer analyses his autobiography to find out facts regarding his personality.

    Whatever the method of getting the autobiography written, it brings out very special aspects of personality of the subject. From this biography the information of complex emotions of the subject can be found out and on the basis of it be can be given personal guidance.

    However, this method is not considered adequate for the measurement of personality of normal persons. It is used for those persons who have to be given personal guidance.

(II) Objective Methods of the Measurement of Personality

By objective method of measurement of personality is meant those methods the result of which are not influenced by the personal likes, dislikes or norms of the measurer. The following are the chief methods among them :

1. **Controlled Observation Method :** Under this method, a person's personality is measured under controlled conditions by observing his intellectual tasks, emotional development, interests, attitudes and habits. Only some reading are not used to infer conclusions, rather the subject is placed under different controlled situations for continuous observation in an objective manner, only then conclusions are drawn.

    The most important feature of this method is that the responses of a person are studied under different controlled situations which begets knowledge about the real facts of his personality.

    However, this task is a little difficult. Correct observation can be done by able and trained persons. Though this method is comparatively more objective than the general observation method, yet it is not completely objective.

2. **Rating Scale Mathod :** This method is a1so called **Determinant Scale Method** or **Order Determinant Method**. Under this method, several personality traits are presented before the subject or the persons related to the subject and they are asked to determine their order. At resent, several types of rating scales are used. The chief ones are as follows :

    (*i*) **Check List :** Check list is of two types. In one, some personality traits are given and the determiner is expected to tick (✓) on the traits which are present in the subjects and cross (×) on those traits which are not present in the subject; for example :

Courage	Yes/No

    In another type of check list, the quantity of the trait is written and the determiner is expected to tick (✓) on the applicable quantity; for example -Tick the quantity of the following trait :

Courage	High	Much	Normal	Low	Least

> *Did u know?* Though this method was first used by **Facner**, a psycho-physicist, yet the first publication of such a method was done by **Galton** in 1889 At present, this method is used the most.

(*ii*) **Numerical Scale :** It is also of two types. In one, some personality traits are given and the determiner is expected to mark 1 for the trait in highest quantity, and mark 2 for comparatively less quantity, and 3, 4 and 5 on the descending quantities. The other numerical scale contains 3, 5 or 7 personality related traits and the determiner is expected to mark 1 for the trait in highest quantity and mark 3, 5 or 7 for the least one; for example : give 1 mark for the trait in highest quantity, 2 marks for the trait less than that, and 3, 4 and 5 for others in descending order :

Truth   Non-violence   Non-thievery   Renunciation   Celibacy

(*iii*) **Graphic Scale :** In it, in place of awarding marks to a personality trait, it is expressed by marking a point on the horizontal line of the graph. This horizontal line is also called continuum. The measurer finds out the quantity of personality or trait on the basis of distance among points. This type of rating scale is used very rarely.

(*iv*) **Ordering Scale :** In it, in place of related questions or statements pertaining to a personality trait, several personality traits are presented together and they are ordered by the subject according to the weightage that he awards to them. For example : order the qualities of a teacher as per your preference :

Subject knowledge, Teaching skill, Sympathy, Sense of duty

(*v*) **Forced Choice Scale :** In it, two or more questions are given about a personality trait, and the subject or determiner has to choose one of them for his answer; for example :

What kind of novels do you like to read ?   Social ∕ Detective

The rating scales can be used to measure several personality traits, such as intelligence, politeness, honesty, discipline, courage, sense of duty, contemplation, etc., but only when the subject and determiner rates the trait correctly.

Though this method is extensively used in personality measurement, yet it has been seen that neither the subjects give correct information about themselves nor do the determiners determine the order correctly. Moreover, ordering is not a task meant for all.

3. **Situation Test :** Under this method, the subjects are presented a real life situation or an artificial situation resembling the real one. The subjects have to make responses toward these situations, and the measures measures personality on the basis of these responses. There are chiefly two types of situation tests : Sociometry and Psycho-drama.

(*i*) **Sociometric Method :** This method was constructed by **J.L. Moreno** in 1934. In this method, the persons of a particular group are asked to select or leave out certain persons on the basis of their traits. For example, to find out the most popular person in the group, the group members can be asked to do the following :

Write the names of any three persons in order of preference whom you want to make the group leader.

(*ii*) **Psycho-Drama :** In this method, dramas based on certain circumstances are organised. The person whose personality traits are to be ascertained, the drama is selected containing circumstances from his life, and this person is also asked to play the role of a certain character. The measurer pays special attention to the activities of this person and finds out the state of his mind on the basis of his emotions.

4. **Physiological Test :** The psychological and physiological activities are closely related. They influence one another. As such, a person's psychological activities can be ascertained by his physiological activities. In the meantime, the scientists have made several tools by which changes in the body can be measured; for example, sphygmograph, plenthysmograph, electro-cardiograph, pheumograph, electromyogram, psycho-galvanometer, electro-encephalogram, etc. These tools can be used to read the physiological changes occurring in a particular situation to study his psyche and his interactions. Thus, the hidden traits of one's personality are found out.

32.3.5 Use of Personality Test

At present, the personality tests are used in several fields. The important fields are : education, industry, commerce, psychology and armed forces.

Use in the Field of Education

(1) For selection of students for admission.

(2) For finding out interests, aptitudes, desires and needs of the students.

(3) For knowing the internal mental condition and social relations of students.

(4) For diagnosis and remedies of personal problems of the students.

(5) For providing individual, educational and vocational guidance to the students.

32.3.6 Limitations of Personality Test

The first thing about the personality tests is that the psychologists do not hold uniform views on the concept or personality. Second, the personality traits can not be measured correctly. Clarification is as follows :

(l) Most personality tests measure only limited personality traits, so no one test can be used to measure all personality traits correctly.

(2) The so far published projective tests for personality measurement are very difficult to administer and analyse.

(3) The data of personality testing are influenced by several factors, so correct measurement of a person's personality at any given time can not be done properly.

(4) All the personality tests that have been constructed so far, have a question mark on their validity.

(5) No two personality tests give the same results, so there is question mark on their validity.

(6) The administration and analysis of all personality tests depend more on the measurer, so they can not be called fully objective.

(7) The personality of a person is a dynamic organisation. The personality measured at any given time can not be considered his permanent or stable personality.

Task   What is psycho drama ?

32.4 Measurement of Intelligence

32.4.1 Concept of Intelligence

The psychologists do not hold uniform view on the concept of intelligence. Different psychologists have accepted it in the form of ability of different mental activities, and some psychologists have taken it in the form of capacity of different mental activities. It is evident that they accept intelligence

in the form of mental activities. All the methods which have so far been developed to measure intelligence, measure only the mental activities of a person and on the basis of the same the intelligence of a person is determined. In that sense, the mental activities will amount to be intellectual activities. The psychologists will have to understand the activities of brain, mind and intelligence separately, only then they will be able to discover the true concept of intelligence.

Notes According to Thurston, these seven mental abilities function in the form of intelligence. He further clarified that these abilities are independent in their own; a person develops according to the amount of these seven abilities.

32.4.2 Nature and Characteristics of Intelligence

All psychologists are not in agreement as regards the structure of intelligence. However, the facts that they have made out in respect of intelligence, may be considered as the nature and characteristics of intelligence.

1. **Intelligence is Hereditary Capacity :** Western psychologist Burt, Shields, Seal, Jenson and Schocaley, on the basis of their experiments reached the conclusion that human intelligence is hereditary. As the height of a person depends on the hereditary genes, in the same way, his intelligence is also depends on hereditary genes.

2. **Environment is Important for Proper Development of Intelligence :** Western psychologists Dwirsk, Mcland, Gaig and Common on the basis of their experiments, reached the conclusion that suitable environment is necessary for the development of hereditary intelligence. As the height of a person is determined by hereditary genes, but needs proper nutrition for its development, in the same way, development of intelligence requires suitable social and educational environment.

3. **Development of Human Intelligence is Completed by the Age of 20 :** Western psychologist Pintner made out this fact on the basis of his study that the development of hereditary intelligence is completed by the end of 16 years. According to some other psychologists, the development of intelligence is completed by the age of 20 years, after this there is no increase in the intelligence quotient (I.Q.).

4. **Yogic Exercise can Enhance Hereditary Intelligence :** As contrary to the conclusions of the western psychologists, the Indian yoga psychologists claim that hereditary intelligence of a person can be enhanced from 30 to 45% with the use of yogic exercise and this fact has now been accepted by the western psychologists too.

5. **Gender Causes no Difference in Intelligence :** Western psychologists Kole and Brush on the basis of their experiments concluded that there is no difference in intelligence on the basis of gender, and this fact is accepted by all.

32.4.3 Types of Intelligence

Psychologist **Thorndike** mentioned three types of intelligence : motor or mechanical intelligence, abstract intelligence and social intelligence. **Garette** named motor or mechanical intelligence as concrete intelligence. At present, these three are considered to be the types of intelligence.

1. **Concrete Intelligence :** The intelligence that assists the human beings in understanding the structure and act accordingly is termed concrete intelligence by Thorndike. It is concrete because it helps in understanding the concrete objects and performing concrete activities. Such intelligence is also called **motor intelligence** or **mechanical intelligence.** The children endowed with more of this intelligence, take special interest in breaking and joining of objects. They also

take interest in other physical activities, such as games, dance, etc. Such children become skilled artisans and engineers.

2. **Abstract Intelligence :** The intelligence that helps the human beings in acquiring bookish knowledge, knowledge about different facts, thinking and finding out solutions to problems, has been termed as abstract intelligence by Thorndike. It is abstract because it is helpful in abstract contemplation and problem-solving. The children endowed with more of this type of intelligence take more interest in book study and contemplation. Such children become good advocates, doctors, teachers, composers, artists and philosophers

3. **Social Intelligence :** The intelligence that helps the human beings in adjusting in the society and taking part in social activities, has been termed as social intelligence by Thorndike. The children endowed with more of this type of intelligence adjust with family members, friends and fellow-students and take interest in social activities. Such children become good businessmen, social workers and politicians.

32.4.4 Measurement of Intelligence

The history of intelligence measurement is very long. In the ancient times in our country, the estimate of a person's intelligence was taken from his ability to learn, memorise, reason and imagine. After the development of astrology, the horoscope was studied to tell whether a person will possess high, medium or low intelligence. The palmists tell this fact by reading the lines on palms. Some people are able to estimate a person's intelligence by looking at the structure of face, nose, ear and forehead. Measurement of children's intelligence by asking riddles is a common method. But all these methods only estimate which person is intelligent and who is not; but it cannot be told how much intelligent one is.

Binet and Simon considered mental activities such as decision, memory, reasoning and numerical abilities as the basis of intelligence measurement. They constructed several questions related to these activities and administered them on several children. Finally, they selected only 30 questions and arranged them in order of difficulty from easy to difficult and gave them the shape of an intelligence test. This test was published in 1905. After this in 1911, they published its amended form. In the Binet-Simon test, a child who answers all the questions meant for his age group correctly is of general intelligence, and the child who also answers some questions determined for children above his age, is of high intelligence; and the child who is not able to answer the questions determined for his age group is considered of low intelligence. At that time, William Stern had suggested the use of intelligence quotient (I.Q.). I.Q. is obtained by dividing child's mental age by his chronological age and multiplying it by 100.

$$\text{Intelligence Quotient (I.Q)} = \frac{\text{Mental Age (M.A.)}}{\text{Chronological Age (C.A.)}} \times 100$$

32.4.5 Chronological Age and Mental Age

Measurement of a child's or person's intelligence is done in the ratio of his mental age and chronological age. Therefore, let us understand what are they.

Chronological Age

In the context of intelligence test, the chronological age of a child or person is the age which is at the time of appearing in the intelligence test. Because, the development of a person's intelligence is complete by 16 years of age, the persons with more than 16 years of age are also considered to be of 16 years of choronological age.

The chronological age is calculated in years, months and days. For that convenience of calculation, 15 or more days are counted as one month. For example, If the chronological age of a child is 6 years 3 months and 4 days at the time of appearing in the intelligence test, it will be considered to be 6 years end 3 months; and if it is 6 years 3 months and 16 days, it will be considered to be 6 years and

4 months. Later, the months are also transformed into years. For example, 6 years and 4 months

Mean $6\frac{4}{12}$ years = $6\frac{1}{3}$ = $\frac{19}{3}$ years.

32.4.6 Mental Age

In the context of intelligence test, the mental age of a child or person shows the extent of intelligence development. If a child is able to perform such mental activities properly which are found suitable for his chronological age, it is considered that his mental age is equal to his chronological age. And if the child is able to perform some of the mental activities determined for the children above his age, correctly than his mental age will be more than his chronological age. On the other hand, if he is not able to perform the activities determined for his age, correctly, his mental age will be less than his chronological age. For example, if an 8-year old child is able to perform the activities determined for the common children of his age, correctly, his mental age will be 8 years. If he is also able to perform the mental activities meant for 10-year old children, his mental age will be considered to be 10 years. On the other hand, if he is not able to perform the mental activities meant for his own age, but is able to perform only the activities meant for 6-year old children, his mental age will be considered to be only 6 years.

The year for which a child performs all activities determined for that age, correctly is called the **basal year** in the context of intelligence test. After this, the number of months are credited to his mental age equivalent to the number of activities he is able to perform determined for above that age. In the context of intelligence test, it is called **credit age**. And the year for which he is not able to answer any of the questions, is called the **terminal year.** It is evident that the mental age of a person is obtained using an intelligence test, as follows :

Mental age = Basal year + Credit Age

Like chronological age, mental age is also calculated in years and months, and its calculation is done exactly in that form as is done for chronological age. For example, the meaning of mental age at 8

years 2 months will be $8\frac{2}{12}$ years = $8\frac{1}{6}$ = $\frac{49}{6}$ years.

32.4.7 Intelligence Quotient and Deviation Intelligence Quotient

Generally, the measurement of intelligence of a child or person is done in the form of intelligence quotient, but in some tests, it is also done in the form of deviation Intelligence quotient. Before proceeding further, it should he prudent to understand their meaning.

Intelligence Quotient

By intelligence quotient is meant the ratio of mental age and chronological age of a child or person. To remove the decimal point from this ratio, it is multiplied by 100. Thus,

$$\text{Intelligence quotient} = \frac{\text{Mental age}}{\text{Chronological age}} \times 100$$

It is written in brief as :

$$\text{I.Q.} = \frac{\text{M.A.}}{\text{C.A.}} \times 100$$

Deviation Intelligence Quotient

Intelligence quotient is found out as a ratio of mental age and chronological age. Because the development of intelligence is generally complete by the age of 16 years, so the chronological age for persons above 16 years of age is also considered to be 16 years. Therefore, is the development of a person's intelligence is completed before 16 years of age or it is completed after it by one, two or four years, error in measurement of intelligence quotient is natural. Again, it is now accepted that the intelligence of a person continues to develop under certain circumstances. As a result, the intelligence quotient as measured at a particular time cannot be considered to be the correct

measurement of one's intelligence. This is the reason that in some of the intelligence tests the deviation intelligence quotient is calculated in place of intelligence quotient. Deviation intelligence quotient is a kind of standard score whose mean (M) = 100, and standard deviation is (S.D., σ) = 16.

32.4.8 Types of Intelligence Test

Intelligence tests are classified on two bases : number of subjects or examinees, and forms of presentation of the tests. Intelligence tests are divided into two categories on the basis of number of the subjects or examinees : individual intelligence tests and group intelligence tests. They are also classified into two types on the basis of their presentation : verbal intelligence tests and non-verbal intelligence tests. These two classifications of intelligence tests can be understood by the following chart collectively :
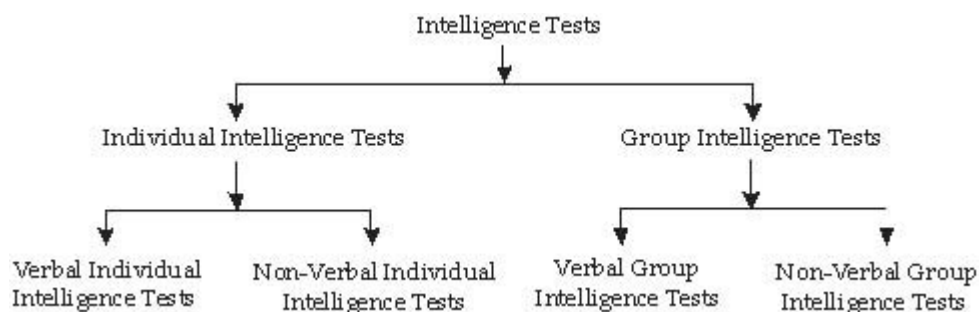
Here we present the general introduction of intelligence tests.

Individual and Group Intelligence Tests

The intelligence tests are divided into two types on the basis of number of subjects or examinees : individual intelligence tests and group intelligence tests.

1. **Individual Intelligence Test :** These are those intelligence tests which can be administered on only one subject or examinee at one time. In these tests, the examiner first establishes rapport with the subject and brings him into normal mental condition; he frees him from all kinds of fears and anxieties. After this, he gives directions about the test, and at last, asks him to answer the questions included in the test. Some of the important tests of this type are : Stanford-Binet test of Intelligence, Wechsler Intelligence Scale, Merril and Palmer Intelligence Scale, Pinter-Peterson Performance Scale, Merril-Palmer Block Building Test and Porteus Maze Test.

   The chief characteristic of these tests is that there is establishment of rapport between the examiner and the examinee in their administration, the examinee is free from all kinds of fears and anxieties, and he answers in normal state of mind.

2. **Group Intelligence Test :** These are those intelligence tests which can be administered one large number of subjects at a time. The administration of these tests does not require the examiner to establish rapport with the subjects. He distributes the intelligence tests himself or gets them distributed. The directions about the test are printed on the tests themselves, or they are printed separately for distribution. Some of such important tests are : Army Alpha Test, Burt's Group Intelligence Test, jalota's Intelligence Test, Raven's Progressive Matrix, Cattell's Culture Free Test and Army Beta Test.

   The chief characteristic of these tests is that they can be constructed and standardised easily. Second, their administration is easy, and does not require specialists. Third, their administration can be done on hundreds of subjects together at one time. As a result, their administration needs less time, energy and money; they are less expensive. Fourth and last, they are especially suitable for selection and classification of students and persons.

These tests have their own limitations and shortcomings also. First, the examiner and the subjects do not have direct contact, so the intellectual behaviour of the subjects cannot be deeply studied. Second, it is difficult to administer such tests on tiny children, dull children and problem persons. Third, their validity and reliability is less. Fourth, there are chances of copying in the administration.

32.4.9 Use of Intelligence Test

At present' the intelligence tests are used in several fields. The important fields include education, industry, commerce, defence services and research.

Utility in the Field of Education

The intelligence tests are most used in the field of education. They are chiefly used for the following tasks :

(1)   For selection of students for admission.

(2)   For classification of students after admission.

(3)   For identifying backward and superior students.

(4)   For arranging education for special children.

(5)   For diagnosis and remedy of problem children.

32.4.10 Limitation of Intelligence Test

All the intelligence tests that have been constructed so far are marked with a question mark as to how correctly they are able to measure intelligence. The first thing in this context is that the psychologists are not in agreement regarding the form of intelligence. Second, they have not been able to differentiate among activities of the mind, brain and intellect. Third, the factors on the basis of which they measure intelligence are not complete in themselves. Clarification is as under :

(1)   Most of the intelligence tests measure limited factors, so they cannot be used to measure intelligence of a person correctly.

(2)   Most of the intelligence tests measure general intelligence (G), so they cannot measure the intelligence or a person correctly.

(3)   The obtained marks in intelligence tests are affected by several factors, so they cannot measure the intelligence of a person correctly.

(4)   Intelligence tests are prepared on the basis of small samples, so they cannot be used to measure the intelligence of a large group correctly.

Task   What is intelligence quotient ?

Self Assessment

2. State whether the following statements are 'true' or 'false'.

   (*i*)   Introvert Category is a self-centered person placed.

  (*ii*)   Murray has classified personality on the basis of values.

 (*iii*)   Murray and Morgan has Constructed T.A.T. Test.

 (*iv*)   Sphygmograph tool is used to test mental activities.

  (*v*)   Intelligence is a sum total of several factors.

(*vi*) According to Hebb, intelligence is of two types : intelligence A and intelligence B.

(*vii*) CAVD test was prepared by psychologist Binet.

(*viii*) The unitary theory of intelligence was propounded by psychologist spearman.

32.5 Summary

- The mental feeling of a person towards an object, person, activity or thought is generally known as his attitude towards that object, person, activity or thought.

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

- Attitudes are the mental feelings of a person towards an object, person, activity or thought which are developed by his own experiences regarding that object, person, activity or thought and which direct him to react to these accordingly.

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

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

   (*ii*) Attitude can be either positive or negative.

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

   (*iv*) Emotions prove helpful in the development of attitude.

   (*v*) Development of attitudes occurs due to social relations.

- Types of Attitudes

- Attitudes are generally classified as follows :

   (*i*) Positive Attitude : When a person likes and accepts an object, person, activity or thought and is attracted towards it and tries to adjust himself according to that object, person, activity or thought. it is called his positive attitude towards that object, person, activity or thought.

   (*ii*) Negative Attitude : When a person dislikes and rejects an object, activity or thought and tries to make distance from that object, person, activity or thought, it is called his negative attitude towards such object, person, activity or thought.

   (*iii*) General Attitude : The attitude which is expressed in general or collective term towards an object, person, activity or thought is called general attitude, such as motherly affection towards women.

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

- Measurement of attitude has been undertaken since ancient times in our country.

- Behavioural Techniques

   Those techniques are included in behavioural techniques by which day-to-day behaviour of a person is observed or questions are asked directly from him to measure his attitude.

   (*i*) Method of Direct Observation : In this method measurement of attitudes of a person is done by observing his daily routine.

   (*ii*) Method of Direct Questioning : In this method, the measurement of attitudes of a person is done by asking questions to him directly.

- Psychological Techniques

   In order to eradicate the shortcomings of the behavioural techniques of measuring attitudes, the psychologists have developed various other methods.

    (*i*) Thurston's Method of Paired Comparison : In this method, some statements are prepared about a psychological object.

  (*ii*) There are faced some problems in measuring attitude by method of paired comparison.

 (*iii*) Lickert, developed a new method for the construction of attitude scale for measuring attitude. This method is based on a simple and clear process. The name of this method or technique is—Method of Summated Rating.

 (*iv*) The method of successive difference developed by Suffir is an improved form of Thurston's and Chave's Equal Appearing Intervals method.

  (*v*) This method is quite different from other methods. In the attitude scale prepared by this method, the statements concerned with the general subject are in pairs.

 (*vi*) Scale Discrimination Technique : This method was developed by Edward and Kilpatric. In it, the merits of all methods have been included.

- The administration of all attitude tests is almost similar but their scoring is done differently.

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

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

 (*iii*) Favourable attitudes of the students are developed.

 (*iv*) Attempt is made to diminish the unfavourable attitudes.

- The attitudes of the students are subject to change, so the limitations of attitudes measured at one time cannot be applied at another time.

- Aptitude is that hereditary ability, capacity or potential of a person which develops by environment, teaching and training, and which helps a person to get success in a particular field.

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

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

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

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

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

- The aptitude tests can be divided into three types keeping in view their nature :

    (*i*) General Aptitude Tests : These tests are used to measure general abilities; such as general intelligence, mental, ability, general creativity, learning ability, etc.

  (*ii*) Differential Aptitude Tests : In these tests, there is a series of sub-tests within a test to measure some specific abilities. These are also called 'batterv-type'tests. These tests are used to measure specific abilities.

 (*iii*) Specified Aptitude Tests : These are the tests which are used to measure aptitude of a person or student in a specific field; for example, music aptitude test, teach in aptitude test.

- Aptitude is measured by several methods. One of them is testing method.

    (*i*) At first, suitable seating arrangement for the subject/subiects is mad. Care is taken that there is light and air in the room in sufficient measure, and there is not any kind of disturbance or noise in the vicinity.

  (*ii*) Having made proper seating arrangement for the subjects, rapport is established with them and they are brought in normal mental condition.

(*iii*) After this, the subjects are given test related general instructions; such as, not to open the test unless ordered, etc.

- The psychologists have expressed the fact that a child becomes that for which he has aptitude. If a child is given opportunity to develop according to his aptitude, he progresses leaps and bounds, and if attempt is made to develop contrary aptitudes, the child does not progress in the desired direction and fails to obtain proficiency.

  - The aptitudes of the students are known.

- The development of a person does not depend only upon hereditary aptitudes, but upon so many other facts.

  (*i*) First, the development of a child's aptitude depends on environment, education and training. As such, he cannot be guided alone on the basis of obtained results of the aptitude test.

  (*ii*) Second a child's development is also dependent on his intelligence, interest, inclination, ability and capacity besides his aptitude.

- Personality is the dynamic organisation within the individual of those psycho-physical systems that determine his unique adjustment to his environment   *— G.W. Allport*

  - Psycho-Analytical Theory : This theory was propounded by psychologist Sigmund Freud. Freud analysed the mental activities of a person, and said on its basis that the behaviour of man is more dependent on his unconscious mind than conscious.

  - Type Theories : Some psychologists have analysed a person's personality on the basis of his biological characteristics, and some on his psychological characteristics. Here is presented an example of each these two theories.

    - Constitutional Theory : Psychologist W.H. Sheldon studied the relationship between the physical constitution of a man and his personality traits and concluded that physical constitution and traits of personality have deep relationship.

      (*i*) Edomorphic : This type of persons are easy going, entertaining, food loving and happy-go-lucky by nature.

      (*ii*) Mesomorphic : Such people are exciting, adventurous, influential and aim-centred.

      (*iii*) Ectomorphic : The people of this shape are peace loving and solitary. They sleep less and get tired easily. They are mostly cruel by nature.

    - Psychological Trait Theory : Psychologist Jung studied the mental nature of man and its relationship with his personality traits, and concluded that there is deep relationship between mental nature and the personality traits of a person. On the basis of psychological traits, he mentioned two types of personality : extrovert and introvert.

      (*i*) Extrovert : Such persons are social by nature.

      (*ii*) Introvert : Such people like to live alone do like less to meet others and fall into friendship with few people.

      Sometimes they are extrovert in their behaviour and sometimes introvert. They named such persons as ambivert.

- Trait Theories : The word 'trait' in the context of personality means a permanent trait of a person.

  (*i*) Allport's Trait Theory : According to Allport, personality is that behaviour of a person which is guided by certain traits present in him. Allport has divided the personality traits into two classes — common traits and personal traits.

  (*ii*) Cattell's Trait Theory : According to Cattell, personality is related with both overt and covert behavious of a persons.

- Need Theory : This theory was propounded by psychologist Henry A. Murray. Murray emphasised on this fact that man is an inspired creature which tries to reduce the stresses arising as a result of internal needs and external pressures.

- The psychologists are not in agreement regarding the concept of personality.

  (*i*) Personality is a Multi-Dimensional Concept : A person's personality comprises of both his innate powers and acquired traits. The important among them are : physical and mental health, self-consciousness, will power, sociability, adjustability, and developmental continuity.

  (*ii*) The Permanent Traits and Abilities of a Person are the Part of His Personality : Permanent traits are those qualities which are always expressed in one's behaviour. And permanent abilities mean those abilities which a person always makes use of.

  (*iii*) Personality is the Integrated Structure of Different Traits and Abilities : The personality of a person is not the sum total of his innate powers and acquired traits and abilities, rather it is a specific structure.

  (*iv*) As the houses made from bricks, cement, iron and wood are different from one another, so are the personalities of different persons made from integration of different traits and abilities.

  (*v*) A person's personality comprises of his internal and external factors, and the development of these two depends on both heredity and environment.

  (*vi*) Both heredity and environment are essential for the development of a person's personality.

  (*vii*) A person is connected with several persons, groups and institutions. These persons, groups and institutions always keep changing, and so his mode of behaviour also keep changing.

- Classification on the Basis of Physical Constitution

  Some psychologists have classified personality on the basis of physical constitution.

  (*i*) Asthenic : Kretschmer has placed those person in this group who are tall and weak and often like to live alone and enjoy criticising others.

  (*ii*) Athletic : Kretschmer has placed those persons in this group who are healthy sound and strong and have the capacity to adjust well.

  (*iii*) Pynic : Kretschmer has placed those persons in this group who are short in height but healthy, easy going and entertaining.

- Classification on the Basis of Nature

  (*i*) Chalenic

  (*ii*) Melancholy

  (*iii*) Phlegmatic

  (*iv*) Sanguine

- Classification on the Basis of Values

  (*i*) Theoretical

  (*ii*) Economical

  (*iii*) Social

  (*iv*) Political

  (*v*) Religious

  (*vi*) Aesthetic

- Classification on the Basis of Social Interaction

  (*i*) Introvert

   (*ii*) Extrovert

  (*iii*) Ambivert

- Classification on the Basis of Adjustment

   (*i*) Well-Adjusted

  (*ii*) Maladjusted

- General introduction to these methods has already been given in under measurement tools.

- Subjective Methods of the Measurement of Personality

  Subjective methods of measurement of personality are those methods whose measurement results are influenced by the views, likes and dislikes and personal norms of the measurer. The following are such methods :

   (*i*) Observation Method : Under this method, the measurer observes the behaviour to the subject and finds out the personality traits present in him.

  (*ii*) Interview Method : Under this method, the measurer and subject are before each other, the measurer asks personality related questions and the subject answers them.

 (*iii*) Questionnaire Method : It is also called Inventory Method. It is the formal and written form of interview method. Under this method, questions are formed on various aspects of a person's personality.

- Closed Questionnaire : These are those questionnaires in which the subjects have answer the questions from within restictions; such as yes or no type questions. An example of such question is:

- Open Questionnaires : These are those questionnaires in which the subjects have the liberty to answer the questions in their own way.

- Pictorial Questionnaires : In these questionnaires, the questions are presented in the form of pictures or figures, and the subjects have to answer by ticking at the suitable picture or figure.

- Mixed Questionnaire : These questionnaires include questions of two or all three types of questions.

- Case History Method : It is also called Life History Method.

- Autobiography Method : Autobiographical method has the following two forms :

   (*i*) Structured Biography

  (*ii*) Unstructured Biography

- Objective Methods of the Measurement of Personality

- Controlled Observation Method : Under this method, a person's personality is measured under controlled conditions by observing his intellectual tasks, emotional development, interests, attitudes and habits.

- Rating Scale Mathod : This method is also called Determinant Scale Method or Order Determinant Method. Though this method was first used by Facner, a psycho-physicist, yet the first publication of such a method was done by Galton in 1889 At present, this method is used the most.

   (*i*) Check List

  (*ii*) Numerical Scale

 (*iii*) Graphic Scale

 (*iv*) Ordering Scale

  (*v*) Forced Choice Scale

Notes

- Situation Tests

    (*i*) Sociometric Method

    (*ii*) Psycho-Drama

- Physiological Tests

- Projective Methods of the Measurement of Personality

    (*i*) Association Technique : In this method, certain stimuli are presented before the subject, and he is asked to give his response to them.

    (*ii*) Construction Technique : In this method, the subject is presented with some stimuli and he is asked to construct something on their basis.

    (*iii*) Completion Technique : In this method, the subject is presented with an incomplete construction. It can be in the form of incomplete sentence, poem or story.

    (*iv*) Ordering or Choice Technique : In this method, the measurer presents certain objects, pictures, words, statements, values and principles before the subject and asks him to order them in order of his preference.

    (*v*) Expression Technique : In these methods, the subject is free to give his response to the stimuli presented before him.

- Theories of Test : Rorschach was a psycho-therapist.

- Use in the Field of Education

    (*i*) For selection of students for admission.

    (*ii*) For finding out interests, aptitudes, desires and needs of the students.

    (*iii*) For knowing the internal mental condition and social relations of students.

    (*iv*) For diagnosis and remedies of personal problems of the students.

    (*v*) For providing individual, educational and vocational guidance to the students.

- Limitations of Personality Tests

    (*i*) Most personality tests measure only limited personality traits, so no one test can be used to measure all personality traits correctly.

    (*ii*) The so far published projective tests for personality measurement are very difficult to administer and analyse.

    (*iii*) The data of personality testing are influenced by several factors, so correct measurement of a person's personality at any given time can not be done properly.

- Unitary or Monarchic Theory : Binet, Terman and Stern, on the basis of their experiments, reached the conclusion that intelligence is a unit factor, power or energy which affects all activities of a person. According to this theory, if the intelligence of a person is sharp.

- Two Factor or Bi-Factor Theory : British psychologist Spearman concluded from his experiments that intelligence is the sum total of two powers or abilities. He described the first factor as general mental ability (G) and the other, specific mental ability (S).

- Three Factor Theory : Spearman himself was able to understand the shortcoming of his Two Factor Theory, so he propounded the Three Factor Theory to do away with the shortcoming of Two Factor Theory.

- Multi-Factor or Anarchic Theory : This theory was propounded by American psychologist, Thorndike. According to him, there is no factor as general ability (G) in the intelligence. Intelligence is the sum total of several specific factors or elements, and several elements come together to perform any mental task.

- Group Factor Theory or PMA Model.   **Notes**

- Hierarchical Theory : British psychologists, Burt and Philip Vernon synthesised the two factor, three factor and group factor theories of intelligence and propounded a new theory which is called Hierarchical Theory.

- Three Dimensional Theory or SI Model : This theory was propounded by psychologist J.P. Guilford. Guilford presented the structure of intelligence in the form of an orderly three dimensional model. It is called the Structure of Intelligence Model (S.I. Model).

    According to Guilford, theory, there can be more than 120 mental abilities. The first notable thing in this regard is that it is difficult to understand and differentiate these 120 mental abilities. Secondly, it is difficult to understand the structure of intelligence by it.

- Fluid and Crystalised Theory

    (*i*) Fluid Intelligence (G$_f$)

    (*ii*) Crystalised Intelligence (Gc)

- Theory of Intelligence A and B

    (*i*) Intelligence A

    (*ii*) Intelligence B

- Nature and Characteristics of Intelligence

    (*i*) Intelligence is Hereditary Capacity

    (*ii*) Environment is Important for Proper Development of Intelligence

    (*iii*) Development of Human Intelligence is Completed by the Age of 20

    (*iv*) Yogic Exercise can Enhance Hereditary Intelligence

    (*v*) Gender Causes no Difference in Intelligence

    (*vi*) Intelligence is the Ability to Learn

    (*vii*) Intelligence is the Ability to Adjust with Environment

    (*viii*) Intelligence is the Ability of Abstract Thinking

    (*ix*) Intelligence is the Ability to Benefit from Prior Experiences and Acquired Knowledge

- Development of Intelligence

- Development of Intelligence during Infancy

- Development of Intelligence during Childhood

- Development of Intelligence during Adolescence

- Types of Intelligence

1. Concrete Intelligence

2. Abstract Intelligence

3. Social Intelligence

- Measurement of Intelligence

- Binet and Simon considered mental activities such as decision, memory, reasoning and numerical abilities as the basis of intelligence measurement. They constructed several questions related to these activities and administered them on several children.

    Intelligence Quotient (I.Q) = $\dfrac{\text{Mental Age (M.A.)}}{\text{Chronological Age (C.A.)}} \times 100$

Intelligence is measured on the basis of its product. The children or persons whose intelligence has to be measured, different kinds of questions and problems are presented before them keeping in view their age, education and circumstance. These are called tasks in the technical language of intelligence tests.

The intelligence tests that have so far been prepared, in them the following tasks are got performed:

- General Information Tasks

  (a) General Knowledge

  (b) Common Sense

  (c) Search for Absurdities

- Verbal Tasks

  (a) Word Span

  (b) Synonyms

  (c) Antonyms

  (d) Analogies

- Chronological Age

  In the context of intelligence test, the chronological age of a child or person is the age which is at the time of appearing in the intelligence test. Because, the development of a person's intelligence is complete by 16 years of age, the persons with more than 16 years of age are also considered to be of 16 years of choronological age.

  The chronological age is calculated in years, months and days. For the convenience of calculation, 15 or more days are counted as one month.

- The months are also transformed into years. For example, 6 years and 4 months Mean $6\frac{4}{12}$

  years $= 6\frac{1}{3} = \frac{19}{3}$ years.

- Mental Age

- In the context of intelligence test, the mental age of a child or person shows the extent of intelligence development. If a child is able to perform such mental activities properly which are found suitable for his chronological age, it is considered that his mental age is equal to his chronological age.

- By intelligence quotient is meant the ratio of mental age and chronological age of a child or person. To remove the decimal point from this ratio, it is multiplied by 100. Thus,

- Intelligence quotient $= \dfrac{\text{Mental age}}{\text{Chronological age}} \times 100$

  $$\text{I.Q.} = \dfrac{\text{M.A.}}{\text{C.A.}} \times 100$$

- Types of Intelligence Tests

- Intelligence tests are classified on two bases : number of subjects or examinees, and forms of presentation of the tests. Intelligence tests are divided into two categories on the basis of number of the subjects or examinees : individual intelligence tests and group intelligence tests.

- Individual and Group Intelligence Tests

- The intelligence tests are divided into two types on the basis of number of subjects or examinees: individual intelligence tests and group intelligence tests.

- Individual Intelligence Tests
- Group Intelligence Tests : These are those intelligence tests which can be administered on a large number of subjects at a time.
- The intelligence tests are most used in the field of education. They are chiefly used for the following tasks :

  (*i*) For selection of students for admission.

  (*ii*) For classification of students after admission.

  (*iii*) For identifying backward and superior students.

  (*iv*) For arranging education for special children.

  (*v*) For diagnosis and remedy of problem children.

- Clarification is as under :

  (*i*) Most of the intelligence tests measure limited factors, so they cannot be used to measure intelligence of a person correctly.

  (*ii*) Most of the intelligence tests measure general intelligence (G), so they cannot measure the intelligence or a person correctly.

  (*iii*) The obtained marks in intelligence tests are affected by several factors, so they cannot measure the intelligence of a person correctly.

  (*iv*) Intelligence tests are prepared on the basis of small samples, so they cannot be used to measure the intelligence of a large group correctly.

32.6 Keywords

Aptitude	:	Natural ability or skill at doing something
Attitude	:	The way that somebody think and feelabout something
Personality	:	The various aspects of a person's character that combine to make them different from other people.

32.7 Review Questions

1. What do you understand by attitude ? Mention its nature and characteristics ?
2. What is the difference between behavioural and psychological techniques of measuring attitude?
3. What is aptitude ? Give an introduction of its methods of measurement.
4. Discuss administration and scoring method of any aptitude test.
5. Mention the different theories of personality.
6. Explain the subjective and projective methods of personality.
7. What are the types of intelligence test ?
8. What is meant by the chronological age, mental age and intelligence quotient of a person in the Context of intelligence that ?

Answers : Self Assessment

1.	(*i*)	b	(*ii*)	d	(*iii*)	c	(*iv*)	d
	(*v*)	a	(*vi*)	a	(*vii*)	b		
2.	(*i*)	T	(*ii*)	F	(*iii*)	T	(*iv*)	F
	(*v*)	T	(*vi*)	T	(*vii*)	F	(*viii*)	T

32.8 Further Readings

1. Education Measurement and Evaluation : *J.Swarupa Rani, Discovery Publishing house.*

2. Measurement and Evaluation in Teaching : *Norman Edward Gronlund Macmillian*

3. Measurement and Assessment in Teaching : *Robert L. Linn Pearson Education India.*

4. Program Evaluation and performance measurement : *James C. Me. David, Laura R.L. Hawthorn, Sage Publications.*