UNIT ONE: THE CONCEPTS OF TEST, MEASUREMENT, ASSESSMENT AND EVALUATION IN EDUCATION

INTRODUCTION

During our discussion of curriculum development and general methods in education, we gave the importance of objectives in education. We also distinguished between Instructional and behavioural objectives. We observed that curriculum implementation and lesson delivery often culminate in ascertaining whether the objectives we set out to achieve were actually achieved. This is often called evaluation.

This unit introduces you to some important concepts associated with ascertaining whether objectives have been achieved or not. Basically, the unit takes you through the meanings of test, measurement assessment and evaluation in education.

Their functions are also discussed. You should understand the fine distinctions between these concepts and the purpose of each as you will have recourse to them later in this course and as a professional teacher.

OBJECTIVES:

By the end of this unit, you should be able to:

1.   distinguish clearly between test, measurement, assessment and evaluation;

2.   state the purposes of assessment and evaluation in education; and

3.   give the techniques of assessment in education.

THE CONCEPTS OF TEST, MEASUREMENT, ASSESSMENT AND EVALUATION IN EDUCATION

These concepts are often used interchangeably by practitioners and if they have the same meaning. This is not so. As a teacher, you should be able to distinguish one from the other and use any particular one at the appropriate time to discuss issues in the classroom.

Measurement

The process of measurement as it implies involves carrying out actual measurement in order to assign a quantitative meaning to a quality i.e. what is the length of the chalkboard? Determining this must be physically done.

Measurement is therefore a process of assigning numerals to objects, quantities or events in other to give quantitative meaning to such qualities.

In the classroom, to determine a child's performance, you need to obtain quantitative measures on the individual scores of the child. If the child scores 80 in Mathematics, there is no other interpretation you should give it. You cannot say he has passed or failed.

Measurement stops at ascribing the quantity but not making value judgement on the child's performance.

Assessment

Assessment is a fact finding activity that describes conditions that exists at a particular time. Assessment often involves measurement to gather data. However, it is the domain of assessment to organise the measurement data into interpretable forms on a number of variables.

Assessment in educational setting may describe the progress students have made towards a given educational goal at a point in time. However, it is not concerned with the explanation of the underlying reasons and does not proffer recommendations for action. Although, there may be some implied judgement as to the satisfactoriness or otherwise of the situation.

In the classroom, assessment refers to all the processes and products which are used to describe the nature and the extent of pupils' learning. This also takes cognisance of the degree of correspondence of such learning with the objectives of instruction.

Some educationists in contrasting assessment with evaluation opined that while evaluation is generally used when the subject is not persons or group of persons but the effectiveness or otherwise of a course or programme of teaching or method of teaching, assessment is used generally for measuring or determining personal attributes (totality of the student, the environment of learning and the student's accomplishments).

A number of instrument are often used to get measurement data from various sources. These include Tests, aptitude tests, inventories, questionnaires, observation schedules etc. All these sources give data which are organised to show evidence of change and the direction of that change. A test is thus one of the assessment instruments. It is used in getting quantitative data.

Evaluation

Evaluation adds the ingredient of value judgement to assessment. It is concerned with the application of its findings and implies some judgement of the effectiveness, social utility or desirability of a product, process or progress in terms of carefully defined and agreed upon objectives or values. Evaluation often includes recommendations for constructive action. Thus, evaluation is a qualitative measure of the prevailing situation. It calls for evidence of effectiveness, suitability, or goodness of the programme.

> *It is the estimation of the worth of a thing, process or programme in order to reach meaningful decisions about that thing, process or programme.*

The Purposes of Evaluation

According to Oguniyi (1984), educational evaluation is carried out from time to time for the following purposes:

(i)   to determine the relative effectiveness of the programme in terms of students' behavioural output;

(ii)   to make reliable decisions about educational planning;

(iii)   to ascertain the worth of time, energy and resources invested in a programme;

(iv)   to identify students' growth or lack of growth in acquiring desirable knowledge, skills, attitudes and societal values;

(v)   to help teachers determine the effectiveness of their teaching techniques and learning materials;

(vi)   to help motivate students to want to learn more as they discover their progress or lack of progress in given tasks;

(vii)   to encourage students to develop a sense of discipline and systematic study habits;

(viii)   to provide educational administrators with adequate information about teachers' effectiveness and school need;

(ix)   to acquaint parents or guardians with their children's performances;

(x)   to identify problems that might hinder or prevent the achievement of set goals;

(xi)   to predict the general trend in the development of the teaching-learning process;

(xii)   to ensure an economical and efficient management of scarce resources;

(xiii)   to provide an objective basis for determining the promotion of students from one class to another as well as the award of certificates;

(xiv)   to provide a just basis for determining at what level of education the possessor of a certificate should enter a career.

ACTIVITY 1

1.   Distinguish clearly between Assessment, Measurement and Evaluation.

Other definitions of evaluation as given by practitioners are:

1.   A systematic process of determining what the actual outcomes are but it also involves judgement of desirability of whatever outcomes are demonstrated. (Travers, 1955)

2.   The process of ascertaining the decision of concern, selecting appropriate information and collecting and analysing information in order to report summary data useful to decision makers in selecting among alternatives (Alkin, 1970).

3.   The process of delineating, obtaining and providing useful information for judging decision alternatives (Stufflebeam et al 1971)

In line with this fine distinction between assessment and evaluation, we shall briefly deliberate a little more here on evaluation and leave the discussion on assessment to the latter units.

ACTIVITY 1I

1.   Why is evaluation important to the classroom teacher?

TYPES OF EVALAUTION

There are two main levels of evaluation viz: programme level and student level. Each of the two levels can involve either of the two main types of evaluation – **formative** and **summative** at various stages. **Programme evaluation** has to do with the determination of whether a programme has been successfully implemented or not. **Student evaluation** determines how well a student is performing in a programme of study.

Formative Evaluation

The purpose of formative evaluation is to find out whether after a learning experience, students are able to do what they were previously unable to do. Its ultimate goal is usually to help students perform well at the end of a programme. Formative evaluation enables the teacher to:

1.   draw more reliable inference about his students than an external assessor, although he may not be as objective as the latter;

2.   identify the levels of cognitive process of his students;

3.   choose the most suitable teaching techniques and materials;

4.   determine the feasibility of a programme within the classroom setting;

5.   determine areas needing modifications or improvement in the teaching-learning process; and

6.   determine to a great extent the outcome of summative evaluation. (Ogunniyi, 1984)

Some of the questions often asked under this type of evaluation include:

1.   What is the objective of the lesson?

2.   What materials will be needed to teach this lesson?

3.   In what sequence will the different aspects of the topic be treated? How much time should be given to different aspects of the topic?

4.   What teaching techniques will be most suitable to transmit this knowledge or skill?

5.   What evaluation techniques would be used to assess student achievement? Will they be effective or not?

6.   What assignment or project should be given as part of or apart from class work?

7.   Has the objective been achieved?

8.   What progress are the students making? What difficulties are they encountering relative to the topic?

9. What additional facilities or resources would enhance the knowledge or skills gained by the students?

10. Are students' needs and interests being met? Are the students able to transfer their knowledge or skills to other areas?

Thus, Formative evaluation attempts to:

(i)   identify the content (i.e. knowledge or skill) which has not been mastered by the students;

(ii) appraise the level of cognitive abilities such as memorization, classification, comparison, analysis, explanation, quantification, application and so on; and

(iii) specify the relationships between content and levels of cognitive abilities.

In other words, formative evaluation provides the evaluator with useful information about the strength or weakness of the student within an instructional context.

SUMMATIVE EVALUATION

Summative evaluation often attempts to determine the extent the broad objectives of a programme have been achieved (i.e. **SSSCE, (NECO or WAEC), PROMOTION, GRADE TWO, NABTEB Exams** and other public examinations). It is concerned with purposes, progress and outcomes of the teaching-learning process.

Summative evaluation is judgemental in nature and often carries threat with it in that the student may have no knowledge of the evaluator and failure has a far reaching effect on the students. However, it is more objective than formative evaluation. Some of the underlying assumptions of summative evaluation are that:

1. the programme's objectives are achievable;

2. the teaching-learning process has been conducted efficiently;

3. the teacher-student-material interactions have been conducive to learning;

4. the teaching techniques, learning materials and audio-visual aids are adequate and have been judiciously dispensed; and

5. there is uniformity in classroom conditions for all learners.

ACTIVITY III

1. With suitable examples, distinguish between formative and summative evaluation,

---

FACTORS TO BE CONSIDERED FOR SUCCESSFUL EVALUATION

1. Sampling technique – Appropriate sampling procedure must be adopted.

2. Evaluation itself must be well organized.

    -       treatment

- conducive atmosphere

- intended and un-intended outcomes and their implications considered.

3. Objectivity of the instrument.

- Feasibility of the investigation

- Resolution of ethical issues

- Reliability of the test (accuracy of data in terms of stability, repeatability and precision)

- validity – test should measure what it is supposed to measure and the characteristics to be measured must be reflected.

4. Rationale of the evaluation instrument

5. It must be ensured that the disparity in students' performances are related to the content of the test rather than to the techniques used in administering the instrument.

6. The format used must be the most economical and efficient.

7. Teachers must have been adequately prepared. They must be qualified to teach the subjects allotted to them.

SUMMARY

In this unit, we have distinguished clearly between measurement, assessment and evaluation.

● Measurement is seen as a process of assigning numbers to objects, quantities or events in other to give quantitative meanings to such qualities.

● Assessment is the process of organizing measurement data into interpretable forms. It gives evidence of change and the direction of change without value judgement.

● Evaluation is the estimation of the worth of a thing, process or programmes in order to reach meaningful decisions about that thing, process or programme. It calls for evidence of effectiveness, suitability of goodness of the programme or process.

● Evaluation serves a number of purposes in education

● Evaluation could be formative or summative. The two serve different purposes in the classroom.

● A number of factors such as sampling techniques, organization, objectivity etc must be considered for successful evaluation.

ASSIGNMENT

1. Distinguish clearly between programme and student evaluation.

2. Describe how a normal programme can be evaluated.

REFERENCES

Ogunniyi, M. B. (1984) *Educational Measurement and Evaluation*: Longman Nig. Mc. Ibadan.

Okpalla P. M. et al (1999) Measurement and Evaluation in Education. Stiching – Horden Publishers (nig.) Ltd. Benin City.

UNIT TWO: ASSESSMENT AND CLASSROOM TESTS

INTRODUCTION

In the last unit, we distinguished between assessment, measurement and evaluation. We also discussed the importance of evaluation in education. In this unit, we shall discuss the purpose of assessment and tests in the classroom. You should pay particular attention here as you may have to construct special types of tests in the latter unit.

OBJECTIVES

At the end of this unit, you should be able to:

i.   give the purpose of assessment;

ii.   explain Bloom's taxonomy of educational objectives;

iii.   give stages in assessment practice;

iv.   compare old and modern assessment practices;

v.   explain what a test is; and

vi.   state the aims and objectives of classroom tests.

HOW TO STUDY THIS UNIT

1.   You should take note of the new words and their usage.

2.   Attempt all the activities.

3.   Try to visualize how you will carry out the new skills in your classroom

PURPOSE OF ASSESSMENT

Assessment involves deciding how well students have learnt a given content or how far the objective we earlier set out have been achieved quantitatively. The data so obtained can serve various educational functions in the school viz:

(a)   **Classroom function**

This includes

(i)   determination of level of achievement

(ii)   effectiveness of the teacher, teaching method, learning situation and instructional materials

(iii)   motivating the child by showing him his progress i.e. success breeds success.

(iv)   it can be used to predict students performance in novel situations.

(b)   **Guidance functions**

Assessment procedure can give the teacher diagnostic data about individual pupils in his class. These will show the pupils' strength, weaknesses and interests. It can also help to decide on which method to use or what remedial activities that are necessary. Parents and pupils can also be rightly guided in terms of career choice.

(c)   **Administrative functions**

(i)   Assessing can serve as communication of information when data collected are used in reports to parents

(ii)   It could form the basis upon which streaming, grading, selection and placement are based.

(iii)   Making appropriate decisions and recommendations on curricula packages and curricula activities.

For any form of assessment to be able to serve the above functions, it cannot be a one shot kind of assessment. It has to be an on-going exercise throughout the teaching and learning processes. This is why continuations assessment is advocated in the classroom.

THE CONCEPT OF CONTINUOUS ASSESSMENT

By continuous assessment, we mean assessing or weighing performance of students periodically to be able to determine progress made in teaching-learning activities. Continuous assessment tests are used to evaluate the progress of students periodically. Continuous assessment tests can be done daily, weekly, monthly, depending on the goals of teaching and learning.   Continuous assessment is defined in the Federal Ministry of Education handbook as:

"A mechanism whereby the final grading of a student in the cognitive, affective and psychomotor domains of behaviour takes account in a systematic way, of all his performances during a given period of schooling. Such an assessment involves the use of a great variety of models of evaluation for the purpose of finding and improving the learning and performance of the students."

Continuous assessment thus is a veritable tool in assessment in that it is comprehensive, systematic, cumulative and guidance oriented. Many schools in the country have since embarked on the implementation of continuous assessment. It is not surprising therefore to find teachers testing their pupils weekly, at the end of each unit or module etc. In recent times however, these tests have assumed disciplinary status to check noise making, absenteeism etc. At this juncture, Continuous Assessment in practice ceases to be a tool for aiding learning. One can only call it what it is – "Continuous Testing". I urge you to be aware of the practice of continuous testing in our school system which is injurious to learning as against Continuous Assessment that is being advocated. The following are the advantages of a continuous assessment:

→   It provides useful information about the academic progress of the learner;

→   It makes the learner to keep on working in a progressive manner;

$\rightarrow$ It informs the teacher about the teaching-learning effectiveness achieved;

$\rightarrow$ It gives a true picture of the student academic performance since it is a continuous process rather than one duration type of test which may be affected by many variables such as sickness, fatigue, stress, etc; and

$\rightarrow$ It makes learning an active rather than a passive process.

USING CONTINUOUS ASSESSMENT TO IMROVE TEACHING AND LEARNING

(a)   **Motivation**

The effectiveness of efforts to help people learn depends on the learner's activities and the achievement that results. Feedback regarding one's effectiveness is positively associated with perceived locus of causality, proficiency and intrinsic motivation (Deci, 1980).

When assessment is carried out systematically and in a purposive manner and the feedback of such is given immediately, it can go a long way in correcting any anomaly in the teaching-learning continuum. In the past, students often do hasty and last minute preparation towards final examinations. This neither helps them to have a thorough grasp of the learning experiences nor does it allow the teacher to apply remedial measures to the areas of deficiency or improve on his teaching methods.

However, using Continuous Assessment appropriately, students study more frequently and retain what they study for longer period of time. This generally improves their learning which goes a long way in motivating them to study further.

(b)   **Individual Differences**

The classroom is an admixture of the slow learners, average, gifted, extroverts, introverts, early bloomers etc. Each of these categories of students should be given a particular attention by the teacher.

Using Continuous Assessment, the teacher will be able to identify these differences and apply at the appropriate time, the necessary measure to improve not only his teaching but the learning of the students and hence their performances.

(c)   **Record-Keeping**

Continuous Assessment affords the teacher the opportunity to compile and accumulate student's record/performances over a given period of time. Such records are often essential not only in guidance and counselling but also in diagnosing any problem that may arise in future.

(d)   **Examination Malpractice**

This is an endemic problem at all levels of our educational system. In practice, continuous assessment had been able to minimize this to a tolerable level and the fear of using one single examination to judge performance of a wide range of course(s) is removed.

ACTIVITY 1

1.   Which type of test do you think the Nigerian education system support most: is it continuous assessment tests or one duration (e.g. 3-hour) examination that is all in all?

CHARACTERISTICS OF CONTINUOUS ASSESSMENT TESTS

i.   In most cases, continuous assessment tests are periodical, systematic, and well-planned. They should not be tests organized in a haphazard manner.

ii.   Continuous Assessment tests can be in any form. They may be oral, written, practical, announced, or unannounced, multiple choice objective, essay, or subjective and so on.

iii.   Continuous assessment tests are often based on what has been learnt within a particular period. Thus, they should be a series of tests.

iv.   In Nigerian educational system, continuous assessment tests are part of the scores used to compute the overall performance of students. In most cases, they are 40% of the final score. The final examination often carries 60%.

v.   Invariably, continuous assessment tests are designed and produced by the classroom teacher. Some continuous assessment tests are centrally organized for a collection of schools or for a particular state.

vi.   All continuous assessment tests should meet the criteria stated in Units three and five for a good test: validity, reliability, variety of tests items and procedure, etc.

ACTIVITY II

1.   What do you think are the disadvantages of continuous assessment tests designed and organized by a classroom teacher? List some of these on a piece of paper or in your exercise book.

If you have done Activity II very well, you might have put down the following disadvantages of continuous assessment tests organized by a classroom teacher. As often reported, continuous assessment tests have been abused by some dishonest teachers. This is done by:

•   Making the test extremely cheap so that undeserving students in their school can pass;

•   Inflating the marks of the continuous assessment tests so that undeserving students can pass the final examinations and be given certificates not worked for.

•   Conducting few (less than appropriate) continuous assessment tests and thus making the process not a continuous or progressive one;

•   Reducing the quality of the tests simply because the classes are too large for a teacher to examine thoroughly;

•   Exposing such tests to massive examination malpractices, e.g. giving the test to favoured students before-hand, inflating marks, or recording marks for continuous

assessment not conducted or splitting one continuous assessment test score to four or five to represent separate continuous assessment tests; etc

Indeed all these wrong application of continuous assessment tests make some public examination bodies to reject scores submitted for candidates in respects of such assessment. For continuous assessment tests to be credible, the teachers must be:

- honest and firm;

- be fair and just in their assessments;

- dedicated and disciplined; and

- shun all acts of favouritism, corruption, and other malpractice activities.

ACTIVITY III

1. Write **three** questions out on a piece of paper or in your exercise book to reflect a credible continuous assessment test in your field or subject area.

PROBLEMS OF CONTINUOUS ASSESSMENT

Plausible and important as the above discussion is, Continuous Assessment is not without its own problems in the classroom. However, real as the problems may be, they are not insurmountable. Some of these problems include:

1. Inadequacy of qualified teachers in the respective fields to cope with the large number of students in our classroom. Some time ago, a Minister of Education lamented the population of students in classrooms in some parts of the country.

2. The pressure to cover a large part of the curricula, probably owing to the demand of external examinations, often makes teachers concentrate more on teaching than Continuous Assessment. There is no doubt that such teachings are not likely to be very effective without any form of formative evaluation.

3. The differences in the quality of tests and scoring procedures used by different teachers may render the results of Continuous Assessment incomparable.

TAXANOMY OF EDUCATIONAL OBJECTIVES

Benjamin Bloom et al classified all educational objectives into three, namely: cognitive, affective and psychomotor domains.

Cognitive domain involves remembering previously learnt matter.

Affective domain relates to interests, appreciation, attitudes and values.

Psychomotor domain deals with motor and manipulative skills.

The focus of Assessment is on these three domains of educational objectives. However, researches have shown that the emphasis has been on the cognitive than the others. This may be because of the difficulty associated with writing of objectives in the other areas. This paper will also focus more on the cognitive domain because most of the Teacher-made Tests

will focus on this. However, the methods of assessing all the three domains will be discussed. For emphasis, the main areas of the cognitive domain are reproduced below:

Bloom's Cognitive Domain

1.0.0 Knowledge of specifics

1.1.1 Knowledge of terminology

1.1.2 Knowledge of specific facts.

1.2.0 Knowledge of Ways and Means of Dealing with specifics

1.2.2 Knowledge of trends and sequences.

1.2.3 Knowledge of classification and categories

1.2.4 Knowledge of criteria

1.2.5 Knowledge of methodology

1.3.0 Knowledge of universal and abstractions

1.3.1 Knowledge of Principles and Generalizations.

1.3.2 Knowledge of Theories and Structures.

2.0.0 Comprehension

2.1.0 Translation

2.2.0 Interpretation

2.3.0 Explanation

3.0 Application

4.0.0 Analysis

4.1.0 Analysis of Elements

4.2.0 Analysis of Relationships

4.3.0 Analysis of Organisational principles

5.0.0 Synthesis

5.1.0 Production of a unique communication.

5.2.0 Production of a plan or proposed set of operations.

6.0.0 Evaluation

6.1.0 Judgement in terms of internal evidence

6.2.0 Judgement in terms of External Criteria.

Stages in Assessment Practice

(i)   Understand and state the instructional outcomes you wish to assess.

(ii)   Formulate the specific behaviour you wish to assess

(iii)   Formulate and create situations which will permit such behaviour to be demonstrated by the pupils.

(iv)   Use appropriate device or instrument to assess the behaviour.

(v)   Take appropriate actions on the outcome of assessment carried out.

Stages in the Assessment of Cognitive Behaviours

A.   **Preparation:**

      i.   Break curriculum into contents (tasks) to be dealt with weekly.

      ii.   Break contents into content elements

      iii.   Specify the performance objectives

B.   **Practice**:

      i.   Give quality instruction

      ii.   Engage pupils in activities designed to achieve objectives or give them tasks to perform.

      iii.   Measure their performance and assess them in relation to set objectives.

C.   **Use of Outcome:**:

      i.   Take note of how effective the teaching has been; feedback to teacher and pupils.

      ii.   Record the result

      iii.   Cancel if necessary

      iv.   Result could lead to guidance and counselling and/or re-teaching.

STAGES IN ASSESSMENT OF PSYCHOMOTOR OUTCOMES

These learning outcomes cannot be assessed through achievement tests or class work. The learning outcomes stretch from handling of writing materials to activities in drama, practicals, laboratory activities, technical subjects, games and athletics. Some of the learning outcomes are subject based or non-subject based, e.g. subject based outcomes.

-        Drawing and painting from art

-        Fluency in speech from language

-        Saying prayers from religious studies

-        Measuring quantities and distance from mathematics.

- Laboratory activities in sciences

- Manipulative skills in subjects involving the use of apparatus and equipment.

- Planting of crops/Experiments

However, elements of cognitive behaviour are also present in these activities because for one to do something well, one must know how since the activities are based on knowledge.

STAGES IN ASSESSING AFFECTIVE LEARNING OUTCOMES IN SCHOOLS

These learning outcomes include feelings, beliefs, attitudes, interests, social relationships etc. which, at times are referred to as personality traits. Some of these that can be assessed indirectly include:

i.   Honesty – truthfulness, trustworthiness, dependability, faithfulness etc.

ii.   Respect – tolerance, respect for parents, elders, teachings, constituted authority, peoples' feelings.

iii.   Obedience – to people and law

iv.   Self-control – temperamental stability, non-aggression, use of decent language etc.

v.   Social relationship – kindness, leadership and social skills.

The most appropriate instrument for assessment here is *observation*. Others like *self-reporting inventories; questionnaires, interviews; rating scales, projective technique* and *sociometric technique* may as well be used as the occasion demands. In assessing students' personality traits, it is necessary to assume that every student possesses good personality characteristics until the contrary is proved.

Note that the purpose of assessing students' personality traits in the school is to give feedback to the students to help them *adjust* in the right direction rather than the *assignment of grades*.

Other personality traits which can be assessed directly are:

i.   Attendance behaviour – regularity and punctuality to school and other activities.

ii.   Participatory behaviour in non-academic activities.

iii.   Appearance: personal cleanliness, in clothes, and for materials handled.

iv.   Conduct: based on reported observed behaviours or incidents involving a display of exemplary character or culpable behaviour.

v.   Cooperative behaviour in groups

OLD AND MODERN PRACTICES OF ASSESSMENT

A.   **Old Assessment Practices**.

Comprised mainly of tests and examinations administered periodically, either fortnightly or monthly. Terminal assessment were administered at the end of term, year or course; hence its being called a 'one-shot' assessment. This system of assessment had the following shortcomings:-

i.   It put so much into a single examination.

ii.   It was unable to cover all that was taught within the period the examination covered.

iii.   Schools depended so much on the result to determine the fate of pupils.

iv.   It caused a lot of emotional strains on the students.

v.   It was limited to students cognitive gains only.

vi.   Its administration periodically, tested knowledge only.

vii.   Learning and teaching were regarded as separate processes in which, only learning could be assessed.

viii.   It did not reveal students' weakness early enough to enable teacher to help students overcome them.

ix.   Achievements were as a result of comparing marks obtained.

x.   It created unhealthy competition which led to all forms of malpractices.

B.   **Modern Practice**

i.   This is a method of improving teaching and learning processes.

ii.   It forms the basis for guidance and counselling in the school.

iii.   Teaching and learning are mutually related.

iv.   Teaching is assessed when learning is and vice-versa.

v.   Assessment is an integral and indispensable part of the teaching-learning process.

vi.   The attainment of objectives of teaching and learning can be perceived and confirmed through continuous assessment.

vii.   It evaluates students in areas of learning other than the cognitive.

TESTS IN THE CLASSROOM

What is a Test?

To understand the concept of "test" you must recall the earlier definitions of "assessment" and "evaluation". Note that we said people use these terms interchangeably. But in the real sense, they are not the same. Tests are detailed or small scale task carried out to identify the candidate's level of performance and to find out how far the person has learnt what was taught or be able to do what he/she is expected to do after teaching. Tests are carried out in order to measure the efforts of the candidate and characterize the performance. Whenever you are tested, as you will be done later on in this course, it is to find out what you know, what you do not know, or even what you partially know. Test is therefore an instrument for assessment. Assessment is broader than tests, although the term is sometimes used to mean tests as in "I want to assess your performance in the course". Some even say they want to

assess students' scripts when they really mean they want to mark the scripts. Assessment and evaluation are closely related, although some fine distinctions had been made between the two terms. Evaluation may be said to be the broadest. It involves evaluation of a programme at the beginning, and during a course. This is called **formative evaluation**. It also involved evaluation of a programme or a course at the end of the course. This is called **summative evaluation.** Testing is part of assessment but assessment is more than testing.

ACTIVITY 1

1.   What is test?

2.   What is assessment?

3.   What is evaluation?

Tests involve measurement of candidates' performance, while evaluation is a systematic way of assessing the success or failure of a programme. Evaluation involves assessment but not all assessments are evaluation. Some are reappraisal of a thing, a person, life, etc.

Purpose of Tests

This section discusses the reasons for testing. Why do we have to test you? At the end of a course, why do examiners conduct tests? Some of the reasons are outlined in this section.

i.   We conduct tests to find out whether the objectives we set for a particular course, lesson or topic have been achieved or not. Tests measure the performance of a candidate in a course, lesson, or topic and thus, tell the teacher or course developer that the objectives of the course or lesson have been achieved or not. If the person taught performed badly, we may have to take a second look at the objectives of the course of lesson.

ii.   We test students in the class to determine the progress made by the students. We want to know whether or not the students are improving in the course, lesson, or topic. If progress is made, we reinforce the progress so that the students can learn more. If no progress is made, we intensity teaching to achieve progress. If progress is slow, we slow down the speed of our teaching.

iii.   We use tests to determine what students have learnt or not learnt in the class. Tests show the aspects of the course or lesson that the students have learnt. They also show areas where learning has not taken place. Thus, the teacher can re-teach for more effective learning.

iv.   Tests are used to place students/candidates into a particular class, school, level, or employment. Such tests are called placement tests. The assumption here is that an individual who performs creditably well at a level can be moved to another level after testing. Thus, we use tests to place a pupil into primary two, after he/she has passed the test set for primary one, and so on.

v.   Tests can reveal the problems or difficulty areas of a learner. Thus, we say we use tests to diagnose or find out the problems or difficulty areas of a student or pupil. A

test may reveal whether or not a learner, for example, has a problem with pronouncing a sound, solving a problem involving decimal, or constructing a basic shape, e.g. a triangle, etc.

vi.   Tests are used to predict outcomes. We use tests to predict whether or not a learner will be able to do a certain job, task, use language to study in a university or perform well in a particular school, college, or university. We assume that if Aliyu can pass this test or examination, he will be able to go to level 100 of a university and study engineering. This may not always be the case, though. There are other factors that can make a student do well other than high performance in a test.

ACTIVITY II

1.   Going by some of the purposes of conducting tests, what do you think could be the aims and objectives of classroom tests?

If you have done Activity II very well, it will not be difficult for you to determine the aims and objectives of classroom tests. Let's study this in the next section.

AIMS AND OBJECTIVES OF CLASSROOM TESTS

In this section, we will discuss aims and objectives of classroom tests. But before we do this, what do we mean by classroom tests? These can be tests designed by the teacher to determine or monitor the progress of the students or pupils in the classroom. It may also be extended to all examinations conducted in a classroom situation. Whichever interpretation given, classroom tests have the following aims and objectives:

i.   **Inform** teachers about the performance of the learners in their classes.

ii.   **Show** progress that the learners are making in the class.

iii.   **Compare** the performance of one learner with the other to know how to classify them either as weak learners who need more attention, average learners, and strong or high achievers that can be used to assist the weak learners.

iv.   **Promote** a pupil or student from one class to another.

v.   **Reshape** teaching items, especially where tests show that certain items are poorly learnt either because they are poorly taught or difficult for the learners to learn. Reshaping teaching items may involve resetting learning objectives, teaching objectives, sequencing of teaching items or grading of the items being taught for effective learning.

vi.   **For certification** – we test in order to certify that a learner has completed the course and can leave. After such tests or examinations, certificates are issued.

vii.   **Conduct a research** – sometimes we conduct class tests for research purposes. We want to experiment whether a particular method or technique or approach is effective or not. In this case, we test the students before (pre-test) using the technique. We then teach using the technique on one group of a comparative level, (i.e. **experimental group**) and do not use the technique but another in another group of a

comparative level, (i.e. **control group**). Later on, you compare outcomes (results) on the experimental and control groups to find out the effectiveness of the technique on the performance of the experimental group

---

ACTIVITY III

1. What do you think are the aims and objectives of the following tests/examinations?

    a.   Trade tests

    b.   Teacher-made tests

    c.   School Certificate Examinations

    d.   National Common Entrance Examination

    e.   Joint Admission and Matriculation Board Examination (JAMB)

SUMMARY

- In this unit, we have explained what assessment is and its purpose in education. Bloom's cognitive domain was briefly summarised and stages in assessment practice were discussed. We also compared old and modern assessment practices. Attempt was also made to define tests, show the purpose of testing, the aims and objectives of classroom tests. You will agree with me that tests and examinations are "necessary evils" that cannot be done without. They must still be in our educational system if we want to know progress made by learners, what has been learnt, what has not been learnt and how to improve learning and teaching.

- Testing is an important component of teaching-learning activities. It is an integral part of the curriculum. Through tests, the teacher measures learners' progress, learning outcomes, learning benefits, and areas where teaching should focus on for better learning.

REFERENCES

Bloom, B. S. – (1966) *Taxonomy of Educational Objectives, Handbook 1 – Cognitive Domain*, David Mckay Co. Inc.

Macintosh H. G. et al (1976), *Assessment and the Secondary School Teacher*: Routledge & Kegan Paul.

Martin Haberman – (1968) "*Behavioural Objectives*: Bandwagon or Breakthrough" The Journal of Teacher Education 19, No. 1:91 – 92.

Davies, A. (1984) – *Validating Three Tests of English Language Proficiency*: Language Testing I (1) 50 – 69

Deci. E. L. (1975) – *Intrinsic Motivation*. New York: Plenum Press.

Mitchell R. J. (1972), *Measurement in the Classroom*: A Worktest: Kendall/Hunt Publishing Co.

Obe E. O. (1981), *Educational Testing in West Africa*: Premier Press & Publishers.

Federal Ministry of Education Science and Technology: (1985) *A handbook on Continuous Assessment.*

Ogunniyi, M. B. (1984). *Educational measurement and evaluation,* Longman Nigeria. Plc.

UNIT THREE: TYPES OF TESTS, CHARACTERISTICS OF GOOD TESTS AND TEST CONSTRUCTION

INTRODUCTION

In unit 1, you were taught what a test is and why you should conduct tests. In this unit, you will learn types of tests. The unit is based on the premise that there are different kinds of tests that a teacher can use. There are also various reasons why tests are conducted. The purpose of testing determines the kind of test. Each test also has its own peculiar characteristics.

OBJECTIVES

By the end of this unit, you should be able to:

a.   list different kinds of tests;

b.   describe each type of test and construct each type of tests to be treated in the unit;

c.   identify different characteristics of a good test;

d.   describe each characteristics as full as possible; and

e.   apply the characteristics to writing a test of your own.

TYPES OF TESTS

Types of tests can be determined from different perspectives. You can look at types of tests in terms of whether they are discrete or integrative. Discrete point tests are expected to test one item or skills at a time, while integrative tests combine various items, structures, skills into one single test.

ACTIVITY I

1.   What is a discrete point test?
2.   What is an integrative test?

DISCRETE POINT TESTS

As we have defined above, a discrete point test, measures or tests or one item, structure, skill, or idea, at a time. There are many examples of a discrete point test. For language tests, a discrete point test may be testing the meaning of a particular word, a grammatical item, the production of a sound, e.g. long or short vowels, filling in a gap with a particular item, and so on. In a mathematics test, it may be testing the knowledge of a particular multiplication table. Let's give some concrete examples.

From the words lettered A-D, choose the word that has the same vowel sound as the one represented by the letters underlined.

<u>Milk</u>

a.   quarry

b.   exhibit

c.   excellent

d.   oblique

Of course, the answer is (d.) because it is only the sound /i/ in the word oblique that has the same sound as "**i**" in milk.

As you can see in this test, only one item or sound is tested at a time. Such a test is a discrete point test.

Let's have another example in English.

Fill in the gap with the correct verb.

John _____ to the market yesterday.

Indeed, only one item can fill the gap at a time. This may be went, hurried, strolled, etc. The gap can only be filled with one item.

In mathematics, when a teacher asks the pupil to fill in the blank space with the correct answer, the teacher is testing a discrete item. For example:

Fill the box with the correct answer of the multiplication stated below:

$2 * 7 = $ ☐

Only one item and that is '14' can fill the box. This is a discrete point test. All tests involving fill in blanks, matching, completion, etc are often discrete point tests.

INTEGRATIVE TESTS

As you have learnt earlier on, tests can be integrative, that is, testing many items together in an integrative manner. In integrative tests, various items, structures, discourse types, pragmatic forms, construction types, skills and so on, are tested simultaneously. Popular examples of integrative tests are essay tests, close tests, reading comprehension tests, working of a mathematical problem that requires the application of many skills, or construction types that require different skills and competencies.

A popular integrative test is the close test which deletes a particular nth word. By nth word we mean the fourth, (4th word of a passage), fifth word (5th word of a passage) or any number deleted in a regular or systematic fashion. For example, I may require you to fill in the words deleted in this passage.

Firstly, he has to understand the _____ as the speaker says _____.   He must not stop the _____ in order to look up a _____ or an unfamiliar sentence.

The tests requires many skills of the candidate to be able to fill in the gaps. The candidate needs to be able to read the passage, comprehend it, think of the appropriate vocabulary items

that will fill in the blanks, learn the grammatical forms, tense and aspects in which the passage is written. When you test these many skills at once, you are testing integratively.

ACTIVITY II

Fill in the gaps in the passage given as an example of a close integrative test.

What nth words was deleted in each case throughout the passage?

Other integrative tests are:

Essay Questions

Give five main characteristics of traditional grammar. Illustrate each characteristic with specific examples.

Reading Comprehension Test

Read the passage below and answer the following questions: (A passage on strike)

Answer the following questions:

a.   Which sentence in the passage suggests that the author supports strikes?

b.   Why is the refusal of students to attend lectures not regarded as strikes according to the passage?

In answering the above questions very well, you will observe that the reading comprehension questions above require the candidates' reading skills, comprehension skills, writing skills and grammatical skills in order for them to answer the questions.

The second perspective for identifying different kinds of tests is by the aim and objectives of the test. For example, if the test is for recording the continuous progress of the candidate, it is referred to as a continuous assessment test. Some of the tests that are for specific purposes are listed below. The purpose for which the test is constructed is also indicated.

i.   **Placement test**: for placing students at a particular level, school, or college.

ii.   **Achievement tests**: for measuring the achievement of a candidate in a particular course either during or at the end of the course.

iii.   **Diagnostic tests**: for determining the problems of a student in a particular area, task, course, or programme. Diagnostic tests also bring out areas of difficulty of a student for the purpose of remediation.

iv.   **Aptitude tests**: are designed to determine the aptitude of a student for a particular task, course, programme, job, etc.

v.   **Predictive test**s: designed to be able to predict the learning outcomes of the candidate. A predictive test is able to predict or forecast that if the candidate is able to pass a particular test, he/she will be able to carry out a particular task, skill, course, action, or programme.

vi.   **Standardized tests**: are any of the above mentioned tests that have been tried out with large groups of individuals, whose scores provide standard norms or reference points for interpreting any scores that anybody who writes the tests has attained. Standardized tests are to be administered in a standard manner under uniform positions. They are tested and re-tested and have been proved to produce valid or reliable scores.

vii.   **Continuous assessment tests** are designed to measure the progress of students in a continuous manner. Such tests are taken intermittently and students' progress measured regularly. The cumulative scores of students in continuous assessment often form part of the overall assessment of the students in the course or subject.

viii.   **Teacher-made tests** are tests produced by teachers for a particular classroom use. Such tests may not be used far-and-wide but are often designed to meet the particular learning needs of the students.

ACTIVITY III

Which type of test, out of the ones described in this unit are stated below:

i.   end of term examination;

ii.   test before the beginning of a course;

iii.   test during the end of a programme;

iv.   school certificate examination;

v.   common entrance examination;

vi.   tests for pilots to quality to fly;

vii.   joint admissions matriculation examination;

viii.   TOEEL test; and

ix.   IETS test.

CHARACTERISTICS OF A GOOD TEST

A test is not something that is done in a careless or haphazard manner. There are some qualities that are observed and analyzed in a good test. Some of these are discussed under the various headings in this section. Indeed, whether the test is diagnostic or achievement test, the characteristic features described here are basically the same.

i.   **A good test should be valid**: by this we mean it should measure what it is supposed to measure or be suitable for the purpose for which it is intended. Test validity will be discussed fully in unit 5.

ii.   **A good test should be reliable**: reliability simply means measuring what it purports to measure consistently. On a reliable test, you can be confident that someone will get more or less the same score on different occasions or when it is used by different people. Again unit 5 devoted to test reliability.

iii.   A good test must be capable of **accurate measurement** of the academic ability of the learner: a good test should give a true picture of the learner. It should point out clearly areas that are learnt and areas not learnt. All being equal, a good test should isolate the good from the bad. A good student should not fail a good test, while a poor student passes with flying colours.

ACTIVITY IV

Think of what can make a good student to fail a test that a poor student passes with flying colours.

---

iv.   A good test should **combine both discrete point and integrative test procedures** for a fuller representation of teaching-learning points. The test should focus on both discrete points of the subject area as well as the integrative aspects. A good test should integrate all various learners' needs, range of teaching-learning situations, objective and subjective items

v.   **A good test must represent teaching-learning objectives and goals**: the test should be conscious of the objectives of learning and objectives of testing. For example, if the objective of learning is to master a particular skill and apply the skill, testing should be directed towards the mastery and application of the skill.

ACTIVITY V

List three objectives of testing school certificate students in mathematics.   Are these objectives always followed in the 'O' level mathematics examinations?

---

vi.   **Test materials must be properly and systematically selected**: the test materials must be selected in such a way that they cover the syllabus, teaching course outlines or the subject area. The materials should be of mixed difficulty levels (not too easy or too difficult) which represent the specific targeted learners' needs that were identified at the beginning of the course.

vii.   **Variety is also a characteristic of a good test**. This includes a variety of test type: multiple choice tests, subjective tests and so on. It also includes variety of tasks and so on. It also includes variety of tasks within each test: writing, reading, speaking, listening, re-writing, transcoding, solving, organizing and presenting extended information, interpreting, black filling, matching, extracting points, distinguishing, identifying, constructing, producing, designing, etc. In most cases, both the tasks and the materials to be used in the tests should be real to the life situation of what the learner is being trained for.

ACTIVITY VI

Why do you think variety should be major characteristics of a test?

Crosscheck your answer with the following. Do not read my own reasons until you have attempted the activity. Variety in testing is crucial because:

- it allows tests to cover a large area;

- it makes tests authentic;

- variety brings out the total knowledge of the learner; and

- with a variety of tasks, the performance of the learner can be better assessed.

SUMMARY

In this unit, you have studied:

- Discrete point tests and integrative tests. Discrete point tests focus on just one item or skill, concept etc, while integrative tests focus on many items, skills and tasks.

- Different types of tests that are determined by the purpose or aim of the test construction. Some of the tests studied are placement achievement, diagnostic, aptitude, predictive, standardized and continuous assessment tests.

ASSIGNMENT

a. You are requested to construct a good test in your field. Your test must be reliable, valid and full of a variety of test procedure and test types or

b. Assess a particular test available to you in terms of how good and effective the test is. What areas of the test that you have assessed, you think improvements are most needed? Supply the necessary improvements.

c. "Test types are determined by the purpose and aim for which the test hopes to achieve?. Discuss this statement in the light of tests that are taught in this unit.

REFERENCES

Davis, A. (1984) – *Validating Three Tests of English Proficiency*, Language Testing 1 (1) 50 – 69

Hughes, A and D. Porter (eds) (1983*) Current Developments in Language Testing*. London: Academic Press.

Oller, J. w. Jr (1979) *Language Tests at School.* London: Longman.

Pophan, W. J. (1975). *Evaluation in Education*: Longman.

UNIT FOUR: BASIC PRINCIPLES OF TEST CONSTRUCTION

INTRODUCTION

This unit is a very important one. You need to know how to construct different kinds of tests. Indeed, tests are not just designed casually or in a haphazard manner. There are rules and regulations guiding this activity.   The unit gives you basic principles to follow when constructing tests.   Before you study this unit, quickly revise the previous unit on characteristics of good tests.

OBJECTIVES

By the end of this unit, you should be able to:

a.   recognize how different types of tests are being constructed;

b.   determine the basic principles to follow in constructing tests; and

c.   apply these principles in the practical construction of tests.

CONSTRUCTION OF TESTS IN THE CLASSROOM

Teacher-made tests are indispensable in evaluation as they are handy in assessing the degree of mastery of the specific units taught by the teacher. The principles behind the construction of the different categories of Tests mentioned above are essentially the same. These shall now be discussed.

Planning for the Test

Many teacher-made tests often suffer from inadequate and improper planning.   Many teachers often jump into the classroom to announce to the class that they are having a test or construct the test haphazardly.

It is at the planning stage that such questions as the ones listed below are resolved:

(i)   What is the intended function of this test? Is it to test the effectiveness of your method, level of competence of the pupils, or diagnose area of weakness before other topics are taught?

(ii)   What are the specific objectives of the content area you are trying to achieve?

(iii)   What content area has been taught? How much emphasis has been given to each topic?

(iv)   What type of test will be most suitable (in terms of effectiveness, cost and practicality) to achieve the intended objectives of the contents?

Defining Objectives

As a competent teacher, you should be able to develop instructional objectives that are behavioural, precise, realistic and at an appropriate level of generality that will serve as a useful guide to teaching and evaluation.

This job has been made easier as these are already stated in the various curriculum packages designed by the Federal Ministry of Education, which are available in schools.

However, when you write your behavioural objectives, use such action verbs like define, compare, contrast, draw, explain, describe, classify, summarize, apply, solve, express, state, list and give. You should avoid vague and global statements involving the use of verbs such as appreciate, understand, feel, grasp, think etc.

It is important that we state objectives in behavioural terms so as to determine the terminal behaviour of a student after having completed a learning task. Martin Haberman (1964) says the teacher receives the following benefits by using behavioural objectives:

1.   Teacher and students get clear purposes.

2.   Broad content is broken down to manageable and meaningful pieces.

3.   Organizing content into sequences and hierarchies is facilitated.

4.   Evaluation is simplified and becomes self-evident.

5.   Selecting of materials is clarified (The result of knowing precisely what youngsters are to do leads to control in the selection of materials, equipment and the management of resources generally).

Specifying the Content to be covered

You should determine the area of the content you want to test. It is through the content that you will know whether the objectives have been achieved or not.

Preparation of the Test Blueprint

Test blueprint is a table showing the number of items that will be asked under each topic of the content and the process objective. This is why it is often called **Specification Table.** Thus, there are two dimensions to the test blueprint, **the content** and the **process objectives**.

As mentioned earlier, **the content** consists of the series of topics from which the competence of the pupils is to be tested. These are usually listed on the left hand side of the table.

The **process objectives or mental processes** are usually listed on the top-row of the table. The process objectives are derived from the behavioural objectives stated for the course initially. They are the various mental processes involved in achieving each objective. Usually, there are about six of these as listed under the cognitive domain viz: Knowledge, Comprehension, Analysis, Synthesis, Application and Evaluation.

i)   **Knowledge or Remembering**

        This involves the ability of the pupils to recall specific facts, terms, vocabulary, principles, concepts and generalizations from memory. This may involve the teacher

asking pupils to give the date of a particular event, capital of a state or recite multiplication tables.

Examples:

Behavioural objectives: To determine whether students are able to define technical terms by giving their properties, relations or attributes.

Question:

Volt is a unit of

(a)   weight   (b) force (c)   distance   (d)   work (e)   volume

You can also use picture tests to test knowledge of classification and matching tests to test knowledge of relationships.

(ii)   **Comprehension and Understanding**

This is testing the ability of the pupils to translate, infer, compare, explain, interpret or extrapolate what is taught. The pupils should be able to identify similarities and differences among objects or concepts; predict or draw conclusions from given information; describe or define a given set of data i.e. what is democracy? Explain the role of chloroplast in photosynthesis.

(iii)   **Application**

Here you want to test the ability of the students to use principles; rule and generalizations in solving problems in novel situations, e.g. how would you recover table salt from water?

(iv)   **Analysis**

This is to analyze or break an idea into its parts and show that the student understands their relationships.

(v)   **Synthesis**

The student is expected to synthesize or put elements together to form a new matter and produce a unique communication, plan or set of abstract relations.

(vi)   **Evaluation**

The student is expected to make judgments based upon evidence.

Weighting of the Content and Process Objectives

The proportion of test items on each topic depends on the emphasis placed on it during teaching and the amount of time spent. Also, the proportion of items on each process objectives depends on how important you view the particular process skill to the level of students to be tested. However, it is important that you make the test a balanced one in terms

of the content and the process objectives you have been trying to achieve through your series of lessons.

Percentages are usually assigned to the topics of the content and the process objectives such that each dimension will add up to 100%. (see the table below).

After this, you should decide on the type of test you want to use and this will depend on the process objective to be measured, the content and your own skill in constructing the different types of tests.

Determination of the Total Number of Items

At this stage, you consider the time available for the test, types of test items to be used (essay or objective) and other factors like the age, ability level of the students and the type of process objectives to be measured.

When this decision is made, you then proceed to determine the total number of items for each topic and process objectives as follows:

(i)   To obtain the number of items per topic, you multiply the percentage of each by the total number of items to be constructed and divide by 100. This you will record in the column in front of each topic in the extreme right corner of the blueprint. In the table below, 25% was assigned to soil. The total number of items is 50 hence 12 items for the topic (25% of 50 items = 12 items).

(ii)   To obtain the number of items per process objective, we also multiply the percentage of each by the total number of items for test and divide by 100. These will be recorded in the bottom row of the blueprint under each process objective. In the table below:

(a)   the percentage assigned to comprehension is 30% of the total number of items which is 50. Hence, there will be 15 items for this objective (30% of 50 items).

Blue Print for Mid-Term Continuous Assessment Test (Objective Items)

Content Areas	*Process Objectives*						
	Knowledge Recognizes Terms & vocabularies 30%	**Comprehension** Identifies facts, Principles, Concepts and Generalizations 30%	**Analysis** Break idea into its parts 10%	**Synthesis** Put elements together to form new matter 10%	**Application** Applies knowledge in new situation 10%	**Evaluation** Judge the worth of Information 10%	**Number of items**
A Soil 25%	4	4	1	1	1	1	12
B Water 20%	3	3	1	1	1	1	10
C Weather 30%	4	4	2	1	1	2	15
D Food 25%	4	4	1	2	2	2	13
Number of Items	15	15	5	5	5	5	50

(b)   To decide the number of items in each cell of the blue print, you simply multiply the total number of items in a topic by the percentage assigned to the process objective in each row and divide by 100. This procedure is repeated for all the cells in the blue print. For example, to obtain the number of items on water under knowledge, you multiply 30% by 10 and divide by 100 i.e. 3.

In summary, planning for a test involves the following basic steps:

(1)   Outlining content and process objectives.

(2)   Choosing what will be covered under each combination of content and process objectives.

(3)   Assigning percentage of the total test by content area and by process objectives and getting an estimate of the total number of items.

(4)   Choosing the type of item format to be used and an estimate of the number of such items per cell of the test blue print.

DIFFERENT KINDS OF TESTS TO BE CONSTRUCTED

In unit 2, you learnt about the different kinds of tests. Can you still remember the test types? List them on a piece of paper and crosscheck your answer with what was supplied in unit 2. All the kinds of tests described earlier on can be grouped into three major parts. The first part is objective or multiple-choice test. The second is subjective or essay type of test. And lastly, are short-answer tests. The multiple choice tests are stated in form of questions, which are put in the stem. You are expected to choose answers called options usually from A, B, C, D and sometimes E that is correct or can fill the gap(s) underlined or omitted in the stem. The essay type asks questions and sub-questions (i.e. questions within a larger unit of questions) and requires you to respond to these by writing full answers in full sentences and paragraphs. Short-answer questions can take various forms. They can be in form of fill-in-gaps, completion, matching, re-ordering, etc. in this kind of test, you do not need to write full sentences for an answer.

ACTIVITY 1

Fill in the gaps

Test can be grouped into ------------------ major groups. These are ------------------ and---------

The ------------- test has alternatives called -----, which usually follow the -------------------- of the question.

Crosscheck your answers by re-reading the section on the different kinds of tests to be constructed.

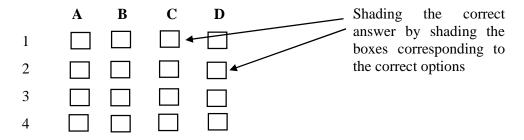## BASIC PRINCIPLES FOR CONSTRUCTING MULTIPLE-CHOICE QUESTIONS

Multiple-choice questions are said to be objective in two ways. First is that each student has an equal chance. He/she merely chooses the correct options from the list of alternatives. The candidates have no opportunity to express a different attitude or special opinion. Secondly,

the judgment and personality of the marker cannot influence the correction in any way. Indeed, many objective tests are scored by machines. This kind of test may be graded more quickly and objectively than the subjective or the easy type.

In constructing objective tests, the following basic principles must be borne in mind.

1. The instruction of what the candidate should do must be clear, unambiguous and precise. Do not confuse the candidates. Let them know whether they are to choose by ticking ($\sqrt{}$), by circling (o) or shading the box in the answer sheet.

    ANSWER SHEET

    An example of a fairly unambiguous instructions are stated below: read the instructions carefully.

    i.   Candidates are advised to spend only 45 minutes on **each** subject and attempt **all** questions.

    ii.   A multiple-choice answer sheet for the four subjects has been provided. Use the **appropriate section** of the answer sheet **for each subject**.

    iii. Check that the number of each question you answer tallies with the **number shaded** on your answer sheet.

    iv.   Use an **HB pencil** throughout.

ACTIVITY 111

1. Study the instruction above and put on a piece of paper or in your exercise book the characteristics of the instructions that were presented for the examination.

Crosscheck your answers with the ones below after you have attempted the activity.

As could be seen in the example just presented, instructions of a test must be:

*   unambiguous

*   clear

*   written in short sentences

*   numbered in sequence

- underlined or bold-faced to show the most important part of the instruction or call attention to areas of the instruction that must not be overlooked or forgotten.

2. The options (or alternatives) must be discriminating: some may be obviously wrong but there must be options that are closely competing with the correct option in terms of characteristics, related concept or component parts.

In question 1, choose the option **opposite** in meaning to the **underlined word**.

i. Albert thinks Mary is **antagonistic** because of her angry tone.

   a. noble

   b. hostile

   c. friendly

   d. harsh

   Answer the question above. State the options that are competing with each other. State the options that are obviously wrong.

   Compare your answer with the discussion below.

   If you have done Activity III very well, you will agree with me that option C is the correct answer and that options A and B are competing. A if the candidate remembers that the answer should be in the **opposite** and B, if he/she has forgotten this fact.

ii. The correct option should not be loner or shorter than the rest, i.e. the incorrect options. Differences in length of options may call the attention of the candidate. The stem of an item must clearly state the problem. The options should be brief.

iii. As much as possible, you should make alternatives difficult to guess. Guessing reduces the validity of the test and makes undeserved candidates pass with no academic effort. The distractions must be plausible, adequate and attractive. They should be related to the stem.

iv. Only one option must be correct. Do not set objective tests where two or more options are correct. You confuse a brilliant student and cause undeserved failure.

v. The objective tests should be based on the syllabus, what is taught, or expected to be taught. It must provoke deep reasoning, critical thinking, and value judgments.

vi. Avoid the use of negative statements in the stem of an item. When used, you should underline the negative word.

vii. Every item should be independent of other items.

viii. Avoid the use of phrases like "all of the above, all of these, none of these or none of the above"

ix.   The reading difficulty and vocabulary level must be as simple as possible.

Basic Principles for Constructing Short-Answer Tests

Some of the principles for constructing multiple choice tests are relevant to constructing short-answer tests.

i.   The instructions must be clear and unambiguous. Candidates should know what to do.

ii.   Enough space must be provided for filing in gaps or writing short answers.

iii.   As much as possible the questions must be set to elicit only short answers. Do not construct long answer-question in a short answer test.

iv.   The test format must be consistent. Do not require fill in gaps and matching in the same question.

v.   The questions should be related to what is taught, what is to be taught or what to be examined. Candidates must know before hand the requirements and demands of the test.

ACTIVITY IV

Answer the questions in this short answer test and bring out the characteristics of the test.

Fill in the gaps with the appropriate words or expressions.

1.   In multiple choice tests each student has an -----------------. Candidates have no opportunity to -------- a different --------- or special --------. But in short answer test candidates are allowed to write ----- by filling ------ or writing ---------- sentences.,

---

Go back to the relevant section to fish out answers that fill the gaps.

BASIC PRINCIPLES FOR CONSTRUCTING ESSAY TESTS

Essay or subjective type of test is considered to be subjective because you are able to express your own opinions freely and interpret information in anyway you like, provided it is logical, relevant, and crucial to the topic. In the same way, your teacher is able to evaluate the quality and quantity of your opinions and interpretations as well as your organization and logic of your presentation. The following are the basic principles guiding the setting of essay question:

i.   Instructions of what to do should be clear, unambiguous and precise.

ii.   Your essay questions should be in layers. The first layer tests the concept, fact, its definition and characteristics. The second layer tests the interpretation of and inferences from the concept, fact or topic, concept, structure, etc to real life situation. In the third layer, you may be required to construct, consolidate, design, or produce your own structure, concept, fact, scenario or issue.

iii.   Essays should not merely require registration of facts learnt in the class. They should not also be satisfied with only the examples given in class.

iv.   Some of the words that can be used in an essay type of test are: compare and contrast, criticize, critically examine, discuss, describe, outline, enumerate, define, state, relate, illustrate, explain, summarize, construct, produce, design, etc. Remember, some of the words are mere words that require regurgitation of facts, while others require application of facts.

SUMMARY

● In this unit, you have been exposed to the basic principles for constructing multiple-choice, short answer and essay types of tests. In all tests, instructions must be clear, unambiguous, precise, and goal-oriented. All tests must be relevant to what is learnt or expected to be learnt. They must meet the learning needs and demands of the candidates. Tests should not be too easy or difficult.

ASSIGNMENT

Construct three multiple-choice, short answers and essay-tests each. Use each test constructed to analyze the basic principles of testing.

REFERENCES

Carroll, J. B. (1983), **Psychometric Theory and Language Testing**. Rowley, Mass: Newbury House.

UNIT FIVE: TEST CONTINGENCIES

INTRODUCTION

In Unit 4, you were exposed to the basic principles of test construction. Can you bring out these principles? Write them on a piece of paper or on your exercise book. There are some factors that affect tests, which are referred to as test contingencies.

In this unit, you will learn what is meant by validity and reliability of tests. As you already know through your study of unit 3, validity and reliability are essential components of a good test.

OBJECTIVES

By the end of this unit, you are expected to be able to:

a.   define and illustrate validity and reliability as test and measurement terms;

b.   describe validity and reliability of tests; and

c.   construct valid and reliable tests.

TEST CONTINGENCIES

As discussed above, we have seen how we can construct and administer the various types of tests based on the objectives of the different aspects of the syllabus. However, a number of factors can affect the outcome of the test in the classroom. These factors may be student, teacher, environmental or learning materials related:

(a)   **Student Factors**

Socio-economic background

Health

Anxiety

Interest

Mood etc

(b)   **Teacher Factors**

Teacher characteristics

Instructional Techniques

Teachers' qualifications/knowledge

(c)   **Learning Materials**

The nature

Appropriateness etc.

(d)   **Environmental**

     Time of day

     Weather condition

     Arrangement

     Invigilation etc.

All these factors no doubt do affect the performance of students to a very significant extent. There are other factors that do affect tests negatively, which are inherent in the design of the test itself: These include:

-     Appropriateness of the objective of the test.

-     Appropriateness of the test format

-     Relevance and adequacy of the test content to what was taught.

TEST VALIDITY

As you learnt in unit 3, validity of tests means that a test measures what it is supposed to measure or a test is suitable for the purposes for which it is intended. There are different kinds of validity that you can look for in a test. Some of these are: content validity, face validity, criterion-referenced validity and predictive validity. These kinds of validity will be discussed now.

1.   **Content Validity**

     This validity suggests the degree to which a test adequately and sufficiently measures the particular skills, subject components, items function or behavior it sets out to measure. To ensure content validity of a test, the content of what the test is to cover must be placed side-by-side with the test itself to see correlation or relationship. The test should reflect aspects that are to be covered in the appropriate order of importance and in the right quantity.

ACTIITY 1

Take a unit of a course in your subject area. List all the things that are covered in the unit. Construct a test to cover the unit. In constructing a test, you should list the items covered in the particular course and make sure the test covers the items in the right quantity.

---

2.   **Face Validity**

     This is a validity that depends on the judgment of the external observer of the test. It is the degree to which a test appears to measure the knowledge and ability based on the judgment of the external observer. Usually, face validity entails how clear the instructions are, how well-structured the items of the test are, how consistent the numbering, sections and sub-section etc are:

3.   **Criterion-Referenced Validity**

This validity involves specifying the ability domain of the learner and defining the end points so as to provide absolute scale. In order to achieve this goal, the test that is constructed is compared or correlated with an outside criterion, measure or judgment. If the comparison takes place the same time, we call this concurrent validity. For example, the English test may be compared with the JAMB English test. If the correlation is high, i.e. $r = 0.5$ and above, we say the English test meets the criterion-referenced, i.e. the JAMB test, validity. For criterion-referenced validity to satisfy the requirement of comparability, they must share common scale or characteristics.

4.   **Predictive Validity**

Predictive validity suggests the degree to which a test accurately predicts future performance. For example, if we assume that a student who does well in a particular mathematics aptitude test should be able to undergo a physics course successfully, predictive validity is achieved if the student does well in the course.

ACTIVITY II

----------------, --------------, --------------------,------------------- and ------------------------- are different types of validity. -----------------Validity measures ------------------ of the course while ------------------ validity measures clear --------------- and ----------------- in numbering, sections and subsections. ----------------- Predicts ------------------- of the subject.

Construct Validity

This refers to how accurately a given test actually describes an individual in terms of a stated psychological trait.

A test designed to test feminity should show women performing better than males in tasks usually associated with women. If this is not so, then the assumptions on which the test was constructed are not valid.

FACTORS AFFECTING VALIDITY

-       cultural beliefs

-       Attitudes of testees

-       Values – students often relax when much emphasis is not placed on education

-       Maturity – students perform poorly when given tasks above their mental age.

-       Atmosphere – Examinations must be taken under conducive atmosphere

-       Absenteeism – Absentee students often perform poorly

Reliability of Tests

In unit 3, we said a good test must be reliable. By this, we mean, measuring what it purports to measure consistently. If candidates get similar scores on parallel forms of tests, this suggests that test is reliable. This kind of reliability is called parallel form of reliability or alternate form of reliability. Split-half is an estimate of reliability based on coefficient of correlation between two halves of a test. It may be between odd and even scores or between first and second half of the items of the test. In order to estimate the reliability of a full test rather than the separate halves, the Spearman-Brown Formula is applied. Test and re-test scores are correlated. If the correlation referred to as **r** is equal to 0.5 and above, the test is said to be of moderate or high correlation, depending on the value of r along the scale (i.e. 0.5 – 0.9) – 1 is a perfect correlation which is rare.

ACTIVITY III

1.   How can reliability of a test be obtained? Describe two possible ways.

Internal consistency reliability is a measure of the degree to which different examiners or test raters agree in their evaluation of the candidates' ability. Inter-rater (two or more different raters of a test) reliability is said to be high when the degree of agreement between the raters is high or very close. Intra-rater (one rater rating scripts at different points in time or at different intervals is the degree to which a marker making a subjective rating of, say, an essay or a procedure or construction gives the same evaluation on two or more different occasions.

METHODS OF ESTIMATING RELIAILITY

Some of the methods used for estimating reliability include:

(a)   **Test-retest method**

An identical test is administered to the same group of students on different occasions.

(b)   **Alternate – Form method**

Two equivalent tests of different contents are given to the same group of students on different occasions. However, it is often difficult to construct two equivalent tests.

(c)   **Split-half method**

A test is split into two equivalent sub tests using odd and even numbered items. However, the equivalence of this is often difficult to establish.

Reliability of tests is often expressed in terms of correlation coefficients. Correlation concerns the similarity between two persons, events or things. Correlation coefficient is a statistics that helps to describe with numbers, the degree of relationship between two sets or pairs of scores.

Positive correlations are between 0.00 and + 1.00. While negative correlations are between 0.00 and – 1.00. Correlation at or close to zero shows no reliability;

correlation between 0.00 and + 1.00, some reliability; correlation at + 1.00 perfect reliability.

Some of the procedures for computing correlation coefficient include:

Product – moment correlation method which uses the derivations of students' scores in two subjects being compared.

$$R = \frac{\Sigma\left[(x - \bar{x})(y - \bar{y})\right]}{\sqrt{\left[\Sigma(x - \bar{x})^2 \Sigma(y - \bar{y})^2\right]}} = \frac{\Sigma d \times dy}{\sqrt{(d^2 \times \Sigma dy^2)}}$$

Pearson product – moment Correlation coefficient

$$R = \frac{N(\Sigma Xy) - (\Sigma x) \bullet (\Sigma y)}{\sqrt{\left[N(\Sigma x^2) - (X)^2\right]\left[N(\Sigma Y^2) - (Y)^2\right]}}$$

Where for the two equations

Σ	=	Sum of
x	=	a raw score in test A
	=	the mean score in test A
y	=	a raw score in test B
	=	the mean score in test B
d	=	deviation from the mean.
N	=	total number of scores.

Spearman's Rank Difference Method

However, a more simpler formular used to calculate correlation coefficient is the **Spearman's Rank-Difference** which is based on the formular:

$$R = 1 - \frac{6\Sigma D^2}{N(N^2 - 1)}$$

Where

 = the sum of

D^2 = squared differences between the rank orders assigned to individual scores.

Sample Calculation

Correlation between two sets of measurements (x and y) of the same individuals, ungroup data, product – moment coefficient of correlation.

Cases	X	Y	$X - \overline{X}$	$Y - \overline{Y}$	x^2	y^2	xy
1	13	11	+5.5	+3	30.25	9	+16.5
2	12	14	+4.5	+6	20.25	36	+27.0
3	10	11	+2.5	+3	6.25	9	+7.5
4	10	7	+2.5	-1	6.25	1	-2.5
5	8	9	+0.5	+1	0.25	1	+0.5
6	6	11	-1.5	+3	2.25	9	-4.5
7	6	3	-1.5	-5	2.25	25	+7.5
8	5	7	-2.5	-1	6.5	1	+2.5
9	3	6	-4.5	-2	20.25	4	+9.0
10	2	1	-5.6	-7	30.25	49	+38.5
Sum Σ	75	80	0	0	124.50	144	102.0

ACTIVITY I

(i)   Calculate the mean for X, cases and Y cases.

$$\frac{\Sigma X}{N} = ? \quad \frac{\Sigma Y}{N} = ?$$

Do you get 7.5 and 8.0 respectively? If not check your calculations again.

(ii)   Calculate the products (XY) for each of the cases and sum up

$\Sigma XY = ?$ Use the last column on the table to check your answer. You should get 102.0.

(iii)   Calculate the sum of $x^2 = (X - \overline{X})^2$

(iv)   Calculate the sum of $y^2 = (Y - \overline{Y})^2$

Your answers in (iii) and (iv) should be 124.50 and 144 respectively. Check.

(v)   Find the square root as follows:

$$\sqrt{(124.50)(144)}$$

(vi)   Then divide your answer in question (ii) by your answer in question (v).

75

Thus, you have applied the formula for calculating (R), the product-moment correlation between the two sets of scores (x) and (y).

ACTIVITY II

Try the following formula for the same problem.

$$(R = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

In this formula, you do not need to calculate the mean and find deviations. You just work with pairs of the scores following the steps:-

Step 1:   Square all x and y scores.

Step 2:   Find the xy product for every pair

Step 3:   Sum all the $x's$, the $y's$, the x^2s, the y^2s and the $xy's$.

Then apply the formula

ACTIVITY III

Try your hand on the ungrouped data below using your calculator.

Cases	X	Y	X^2	Y^2	XY
1	13	7	169	49	91
2	12	11	144	121	132
3	10	3	100	9	30
4	8	7	64	49	56
5	7	2	49	4	14
6	6	12	36	144	72
7	6	6	36	36	36
8	4	2	16	4	8
9	3	9	9	81	27
10	1	6	1	36	6
Sum $\sum$	70	65	624	533	472

Your answer should be + 0.14 Check.

FACTORS AFFECTING RELIABILITY

Some of the factors that affect reliability include:

- the relationship between the objective of the tester and that of the students.

- The clarity and specificity of the items of the test.

- The significance of the test to the students.

- Familiarity of the tested with the subject matter.

- Interest and disposition of the tested.

- Level of difficulty of items.

- Socio-cultural variables.

- Practice and fatigue effects.

SUMMARY

- In this unit, you have been exposed to the concept of validity and reliability of tests. Any good test must achieve these two characteristics. A test is said to be valid if it measures what it is supposed to measure. A test is reliable if it measures what it is supposed to measure consistently.

ASSIGNMENT

i. Take any test designed either by you or by somebody else and assess the face and content validity of the test.

ii. Construct a test of three items. Assess the reliability of the test by administering it to three persons at different points or intervals. Compute the coefficient of correlation of the test.

REFERENCES

Davies, A (19840) Validating Three Tests of English Language Proficiency. **Language Testing I** (1) 50 – 69.

Oller, J. N. (1979**) Language Tests at School**. London: Longman.

Carroll, J. B. (1983), **Psychometric Theory and Language Testing**.   Rowley, Mass: Newbury House.

UNIT SIX: TEST SCORING AND INTERPRETATION OF TEST SCORES

INTRODUCTION

The effort through this Module is to show what test is, why test is important, how tests are constructed and what precautions are taken to ensure validity of tests. The module will round off by explaining how tests are scored and interpreted. In order to enjoy the study of the unit, you should have other units by your side and cross-check aspects relevant to this unit that were discussed in the previous units.

OBJECTIVES

By the end of this unit, you should be able to:

a.   score and interpret tests in general and continuous assessment in particular;

b.   analyse test items;

c.   compute some measures of general tendency and variability; and

d.   compute Z – score and the Percentile.

SCORING OF TESTS

This section introduces to you the pattern of scoring of tests, be they continuous assessment tests or other forms of tests. The following guidelines are suggested for scoring of tests:

i.   You must remember that multiple choice tests are difficult to design, difficult to administer, especially in a large class, but easy to score. In some cases, they are scored by machines. The reasons for easy scorability of multiple-choice tests are because they usually have one correct answer which must be accepted across the board.

ii.   Essay or subject types of tests are relatively easy to set and administer, especially in a large class. They are, however, difficult to mark or assess. The reason is because easy questions require a lot of writing of sentences and paragraphs. The examiner must read all these.

iii.   Whether an objective or subjective tests, all tests must have marking schemes. Marking schemes are the guide for marking any test. They consist the points, demands and issues that must be raised before the candidate can be said to have responded satisfactorily to the test. Marking schemes should be drawn before testing not after the test has been taken. All marking schemes should carry mark allocation. They should also indicate scoring points and how the scores are totaled up to represent the total score for the question or the test.

iv.   Scoring or marking on impression is dangerous. Some students are very good at impressing examiners with flowery language without real academic substance. If you mark on impression, you may be carried away by the language and not the relevant

facts. Again, mood may change impression; your impression can be changed by joy, sadness, tiredness, time of the day and son on. That is why you must always insist on a comprehensive marking scheme.

v.   Scoring can be done question-by-question or all questions at a time. The best way is to score or mark one question across the board for all students. Sometimes, this may be feasible and tedious, especially in a large class.

vi.   Scores can be interpreted into grades, A, B, C, D, E and F. They may be interpreted in terms of percentages: 10%, 20%, 50% etc. Scores may be presented in a comparative way in terms of 1st position, 2nd position, and 3rd position to the last. Scores can be coded in what is called BAND. In band system, certain criteria are used to determine those who will be in Excellent, Very Good categories, etc. An example of a band system is the one given by the International English Testing Services (IETS) and the one by Teaching English as a Foreign Language (TOEFL) test.

OBJECTIVE SCORING: Correction Formular

As stated in the main paper, objective test is very easy to score. All other advantages of objective test are well known to all by now. However, the chances of guessing the correct answer are high.

To discourage guessing, some objective tests give instructions to candidates that they may be penalized for guessing. In such situation, the **correction formular** is applied after scoring. This is given as:

No. of questions marked right (R)   - $\dfrac{\text{No of questions marked wrong (W)}}{\text{No. of options per item (N) - I}}$

If in an objective test of 50 questions where guessing is prohibited, a candidate attempted all questions and gets 40 of them correctly, then the actual score after correction is.

$$S = 40 = \quad \text{(assuming the options per item is 5.)}$$

$$= 40 = 2.5 \quad = \quad 37.5 = 38 \text{ out of } 50.$$

Find the corrected scores of two candidates. A and B who both scored 35 in an objective test of 50, if a attempted 38 questions while B attempted all the questions.

$$S_A = 35 - ¾ = 34 \text{ and } S_B = 35 – 15/4 = 31)$$

Note that under rights only, each of the students gets 35 out of 50.

USING TEST RESULTS

As earlier mentioned, conducting tests is not an end in itself. However, before tests could be used for those purposes, the teacher needs to know how well designed the test is in terms of difficulty level and discrimination power, then he should be able to compare a child's performance with those of his peers in the class. Occasionally, he may like to compare the child's performance in one subject area with another.

To do this, he carries out the following activities at various times:

i.   Item analysis.

ii.   Drawing of frequency Distribution Tables.

iii.   Finding measures of central tendency (Mean, Mode, Median)

iv.   Finding measures of Variability and Derived Scores.

v.   Assigning grades.

As you are aware of these procedures, some will be discussed in passing while greater attention will be paid to the others for emphasis.

Item Analysis

Item analysis helps to decide whether a test is good or poor in two ways:

i.   It gives information about the difficulty level of a question.

ii.   It indicates how well each question shows the difference (discriminate) between the bright and dull students. In essence, item analysis is used for reviewing and refining a test.

Difficulty Level

By difficulty level we mean the number of candidates that got a particular item right in any given test. For example, if in a class of 45 students, 30 of the students got a question correctly, then the difficulty level is 67% or 0.67. The proportion usually ranges from 0 to 1 or 0 to 100%.

An item with an index of 0 is too difficult hence everybody missed it while that of 1 is too easy as everybody got it right. Items with index of 0.5 are usually suitable for inclusion in a test.

Though the items with indices of 0 and 1 may not really contribute to an achievement test, they are good for the teacher in determining how well the students are doing in that particular area of the content being tested. Hence, such items could be included. However, the mean difficult level of the whole test should be 0.5 or 50%.

Usually, the formular for their difficulty is $p = \dfrac{n \; x \; 100}{N}$ where

P   =   item difficult

n   =   the no of students who got the item correct.

N   =   the number of students involved in the test.

However, in the classroom setting, it is better to use the upper $\dfrac{1}{3}$ of the students that got the item right (U) and the lower $\dfrac{1}{3}$ of the students that got it right (L).

Hence difficulty level is given by

$$\frac{U + L}{N}$$

Where N is the number of students actually involved in the item analysis (upper $\frac{1}{3}$ + lower 1/3 of the tests).

Consider a class of two arms with a population of 60 each. If 36 candidates of the upper $\frac{1}{3}$ population and 20 of the lower 1/3 got question number 2 correctly, what is the index of difficulty (difficulty level) of the question?

Index of difficulty   = P =

N   = $40 = 40 \, (\frac{1}{3}$ upper + $\frac{1}{3}$ lower)

U   =   36

L   =   20

P   = $\frac{(36 + 20)}{80}$ = $\frac{56}{80}$

=   0.7

i.e.   P =   70%

If P = 0 (0%) or 1 (100%) then the test is said to be either too difficulty or too simple respectively. As much as possible, teachers should avoid administering test items with 0 or 1 difficulty levels.

Item Discrimination

The discrimination index shows how a test item discriminates between the bright and the dull students. A test with many poor questions will give a false impression of the learning situation. Usually, a discrimination index of 0.4 and above are acceptable. Items which discriminate negatively are bad. This may be because of wrong keys, vagueness or extreme difficulty. The formular for discrimination index is: $\dfrac{U - L}{\frac{1}{2}N}$ or $\dfrac{U - L}{0.5N}$

Where

U   =   the number of students that got it right in upper group.

L   =   the number of students that got it right in the lower group.

N   =   the number of students usually involved in the item analysis.

In summary, to carry out item analysis:

81

i)   arrange the scored papers in order of merit – highest and lowest

ii)   select the upper 33%

iii)   select the lower 33%)

Note that the number of students in the lower and upper groups must be equal.

iv)   Item by item, calculate the number of students that got each item correct in each group.

v)   Estimate

(a)   item difficulty $= \dfrac{U + L}{N}$

(b)   item discrimination index $= \dfrac{U - L}{\frac{1}{2}N}$

ACTIVITY

In the table below, determine the P and D for items 2, 3 & 4. Item 1 has been calculated as an example. Total population of Testee is 60.

No. that of the item Right				
Item	*1/3 upper group U = 20*	*1/3 lower group L = 20*	*Difficulty U + L/N*	*Discrimination U –L/1/2N*
1	15	10	25/40 - 62.5%	5/20 = 0.25
2	18	15	-------(?)	-------(?)
3	5	12	-------(?)	-------(?)
4	12	12	-------(?)	-------(?)

DISTRIBUTION AND MEASURES OF CENTRAL TENDENCY

We shall not dwell so much on the drawing of frequency distribution tables and calculating measures of central tendency i.e. mode, median and mean here. This will be taken care of elsewhere. However, we will mention the following about them:

Mode

The mode is the most frequent or popular score in the population. This is usually evident during the drawing of frequency tables. It is not frequently used as the median and mean in the classroom because it can fall anywhere along the distribution of scores (top, middle or bottom) and a distribution may have more than one mode.

Median

This is the middle score after all the scores have been arranged in order of magnitude i.e. 50% of the score are on either side of it. Median is very good where there are deviant or extreme scores in a distribution, however, it does not take the relative size of all the scores into consideration. Also, it cannot be used for further statistical computations.

The Mean

This is the average of all the scores and it is obtained by adding the scores together and dividing the sum by the number of scores.

$$\text{M or} = \overline{X} = \frac{\text{Sum of all Scores}}{\text{Number of Scores}}$$

Though, the mean is influenced by deviant scores, it is very important in that it takes into cognizance the relative size of each score in the distribution and it is also useful for other statistical calculations.

ACTIVITY

The mean score is the same as the average score i.e. Sum of all scores/the number of scores. This is the most common statistical instrument used in our classroom

If in a class of 9, the scores are 29, 85, 78, 73, 40, 35, 20, 10 and 5. Find the mean.

MEASURES OF VARIABILITY

Measure of variability indicates the spread of the scores. The usual measures of variability are Range, Quartile Deviation and Standard Deviation. Their computations are as illustrated below.

Range

The range is usually taken as the difference between the highest and the lowest scores in a set of distribution. It is completely dependent on the extreme scores and may give a wrong picture of the variability in the distribution. It Is the simplest measure of variability. Example: 7, 2, 5, 4, 6, 3, 1, 2, 4, 7, 9, 8, 10. Lowest score = 1, Highest = 10. Range = 10 -1 = 9

Quartile Deviation

Note that Quartiles are points on the distribution which divide it into "quartiles", thus, we have 1st, 2nd and 3rd quartiles.

Inter-quartile range is the difference between Q3 and Q1 i.e. Q3 = Q1. This is often used than the Range as it cuts off the extreme score. Semi inter-quartile Range is thus half of inter-quartile range.

This is also known as the semi-inter quartile range. It is half the difference between the upper quartile (Q3) and the lower quartile (Q1) of the set of scores.

$$\frac{Q3 - Q1}{2}$$

Where Q3 = P75 = point in the distribution below which lie 75% of the scores.

Q1 = P25 = Point in the frequency distribution below which lies 25% of the scores.

In cases where there are many deviant scores, the quartile deviation is the best measure of variability.

Standard Deviation

This is the square root of the mean of the squared deviations. The mean of the squared deviations is called the variance (S^2). The deviation is the difference between each score and the mean.

$$SD (M) \quad = \quad \sqrt{\frac{\sum x^2}{N}}$$

x $\quad = \quad$ X - $\overline{X}$ - deviation of each score from the mean

N $\quad = \quad$ number of scores.

The SD is the most reliable of all measures of variability and lend itself for use in other statistical calculations.

Deviation is the difference between each score (X) and the mean (M). To calculate the standard deviation:

(i) find the mean (m)

(ii) find the deviation (x-m) and square each.

(iii) sum up the squares and divide by the number of the population (N)

(iv) find the positive square root.

Calculation of Standard Deviation

Students	Marks obtained	Deviations $X - m$ Take m = 54	Squared deviation $(X - m)^2 = x^2$
A	68	14	196
B	58	4	16
C	47	-7	49
D	45	-9	81
E	54	0	0
F	50	4	16
G	62	8	64
H	59	5	25
I	48	-6	36
J	52	-2	4
			487

$$N = 10$$

$$SD (M) = \frac{\sum x^2}{N} = \frac{(X - M)^2}{N} = \frac{487}{10} = 6.97$$

Activity

Find the mean and standard deviation for the following marks.

20, 45. 39, 40, 42, 48, 30, 46 and 41.

DERIVED SCORES

In practice, we report on our students after examinations by adding together their scores in the various subjects and thereafter calculate the average or percentage as the case may be. This does not give a fair and reliable assessment. Instead of using raw scores, it is better to use derived scores". A derived score usually expresses every raw score in terms of other raw score on the test. The commonly used ones in the class room are the Z-Scores, T-Score and Percentiles. The computation of each of these will be demonstrated.

STANDARD SCORE OR Z-SCORE

Standard score is the deviation of the raw score from the mean divided by the standard deviation i.e.

$$Z = \frac{\overline{X} - X}{SD}$$

Where Z = Z - score

X = any raw score

$\overline{X}$ = the mean

SD = Standard Deviation

Raw scores above the mean usually have positive Z-scores while those below the mean have negative Z-scores. Z-scores can be used to compare a child's performance with his peers in a test or his performance in one subject with another.

T-Score

This is another derived score often used in conjunction with the Z-score. It is defined by the equation.

$$T = 50 + 10Z$$

Where z is the standard score.

It is also used in the same way as the Z-score except that the negative signs are eliminated in T-Scores.

Student	Score in English	Score in Maths	Total	Rank
A	68	20	88	(8)
B	58	45	103	(1)
C	47	39	85	(9)
D	45	40	85	(10)
E	54	42	96	(3)
F	50	48	98	(2)
G	62	30	92	(7)
H	59	36	95	(4)
I	48	46	94	(5)
J	52	41	93	(6)

Consider the maximum scores obtained in English and Mathematics in the table above. We cannot easily guarantee which of the subject was more tasking and in which the examiner was more generous. Hence, for justice and fair play, it is advisable to convert the scores in the two subjects into common scores (Standard scores) before they are ranked. Z – and T – score are often used.

The Z – score is given by

$$Z \text{ - Score} = \frac{\text{Raw Score} - \text{mean}}{\text{Standard deviaiton}} = \frac{X - M}{SD}$$

Also the T – Score is given as

$$T = 50 + 10 Z$$

The T-score helps to eliminate the negative or fractional scores arising from Z-scores.

Activity

Calculate the Z- and T-scores for students A,B,C and D in the table above.

Percentile

This expresses a given score in terms of the percentage scores below it i.e. in a class of 30, Ibrahim scored 60 and there are 24 pupils scoring below him. The percentage of score below 60 is therefore:

$$\frac{24}{30} \times \frac{100}{1} = 80\%$$

Ibrahim therefore has a percentile of 80 written P80. This means Ibrahim surpassed 80% of his colleagues while only 20% were better than him. The formula for the percentile rank is given by:

86

$$PR = \frac{100}{N} \times (b + \frac{F}{2}) \text{ where}$$

PR   =   Percentile rank of a given score

b   =   Number of scores below the score

F   =   Frequency of the score

N   =   Number of all scores in the test.

COURSE CREDIT SYSTEM AND GRADE POINTS

Perhaps the most precious and valuable records after evaluation are the marked scripts and the transcripts of a student. At the end of every examination e.g. semester examination, the marked scripts are submitted through the head of department or faculty to the Examination Officer. Occasionally, the Examination Officer can round off the marks carrying decimal, either up or down depending on whether or not the decimal number is greater or less than 0.5

The marks so received are thereafter translated/interpreted using the Grade Point (GP), Weighted Grade Point (WGP), Grade Point Average (GPA) or Cumulative Grade Point Average (CGPA).

CREDIT UNITS

Courses are often weighed according to their credit units in the course credit system. Credit units of courses often range from 1 to 4. This is calculated according to the number of contact hours as follows:

1 credit unit   =   15 hours of teaching.

2 credit units   =   15 x 2 or 30 hours

3 credit units   =   15 x 3 or 45 hours

4 credits units   =   15 x 4 or 60 hours

Number of hours spent on practicals are usually taken into consideration in calculating credit loads.

GRADE POINT (GP)

This is a point system which has replaced the A to F Grading System as shown in the summary table below.

WEIGHTED GRADE POINT (WGP)

This is the product of the Grade Point and the number of Credit Units carried by the course i.e. WGP = GP x No of Credit Units.

GRADE POINT AVERAGE (GPA)

This is obtained by multiplying the Grade Point attained in each course by the number of Credit Units assigned to that course, and then summing these up and dividing by the total number of credit units taken for that semester (total registered for).

$$\text{GPA} = \frac{\text{Total Points Scored}}{\text{Total Credit Units registered}}$$

$$= \frac{\text{Total WGP}}{\text{Total Credit Units registered}}$$

CUMMULATIVE GRADE POINT AVERAGE (CGPA)

This is the up-to-date mean of the Grade Points earned by the student. It shows the student's over all performance at any point in the programme.

$$\text{CGPA} = \frac{\text{Total Points so far Scored}}{\text{Total Credit Units so far taken}^A \text{or registered}}$$

SUMMARY OF SCORING AND GRADING SYSTEM

Credit Units	Percentage Scores	Lower Grades	Grade Points (GP)	Grade Points Average (GPA)	Cumulative Grade Point Average (CGPA)	Level of Pass in Subject
(I)	(II)	(III)	(IV)	(V)	(VI	(VII)
Vary according to contact hours	70 - 100	A	5	Derived by multiplying column (I)	4-50 – 5.00	Distinction
assigned to each	60 - 69	B	4	Gradaunts with GP and divided by	3.50 – 4.49	Credit
course per week	50 - 59	CC	3	Total Credit	2.40 – 3.49	Merit
per term and						
according to work				Units		
Load carried by	45 - 49	D	2		1.50 – 2.39	Pass
Student						
	40 - 44	E	1		1.00 – 1.49	Lower Pass
	0.39	F	0		0.99	Fair

(The scores and their letter grading may vary from programme to programme or Institution to Institution)

For example, a score of 65 marks has a GP of 4 and a Weighted Grade Point of 4 x 3 if the mark was scored in a 3 unit course. The WGP is therefore 12. If there are five of such courses with course units 4, 3, 2, 2 and 1 respectively. The Grade Point Average is the sum of the five weighted Grade Points divided by the total number of credit units i.e. (4 + 3 + 2 + 2 + 1)

ACTIVITY

Below is a sample of an examination transcript for a student

a.   Determine for each course the

    (i)   GP and

    (ii)   WGP.

b.   Find the GPA.

Course Codes	Credit Unit	Score	Grade	G.P.	W.G.P
EDU 121	2	40	E		
EDU 122	3	70	A		
EDU 123	1	50	C		
GSE 105	4	70	A		
GSE 106	1	60	B		
GSE 107	3	42	E		
PES 131	1	39	F		
PES 132	4	10	F		
PES 133	2	45	D		
TOTAL					

NOTE:

$$\text{GPA} \quad = \quad \frac{\text{WGD}}{\text{Total Credit taken}}$$

SUMMARY

- In this unit, we have discussed the basic principles guiding scoring of tests and test interpretations.

- The use of frequency distribution, mean, mode and mean in interpreting test scores were also explained.

- The methods by which test results can be interpreted to be meaningful for classroom practices were also vividly illustrated.

ASSIGNMENT

1.   State the various types of Tests and explain what each measure are?

2.   Pick a topic of your choice and prepare a blue-print table for 25 objective items.

3.   Explain why:

     (a)   we use percentile to describe student's performance and

     (b)   Z-scores to describe in a distribution.

4.   Give four factors each that can affect the reliability and validity of a test.

5.   Use the criteria and basic principles for constructing continuous assessment tests discussed in this unit to develop a 1 hour continuous assessment test in your subject area. By citing specific examples from the test you have constructed, show how you have used the testing concepts learnt to construct the test. You should bring out from your test at least ten testing concepts used in the construction of the test.

REFERENCES

Canale, M and Swain (1980) **Theoretical Basis of Communicative Approaches to Second Language Teaching and Testing**. Applied Linguistics 1 (I): 1 – 47.

Carroll, J. B. (1983**) Psychometric Theory and Language testing** in Oller, J. W (ed) Issues in Language Testing Research Rowley, Mass: Newbury House.

Cook, T. d and C. s. Reichardt (eds) (1979) **Qualitative and Quantitative Methods in Evaluation Research**. Beverly Hills, Calif… Sage.

Lado, R. (1961) Language **Testing: The Construction and Use of Foreign Language Tests.** London: Longman.

Licingston, S. A. and M. J. Zeiky (1982**) Passing Scores: A Manual for setting standards of Performance on Educational and Occupational Tests**.   Princeton N. J: Educational Testing Services.