MODULE

Educational Assessment, Measurement and Evaluation

AET06207

Institute of Adult Education
Adult and Continuing Education Studies Department
Ordinary Diploma in Adult and Continuing Education – Through ODL

Copyright

Institute of Adult Education
Ordinary Diploma in Adult and Continuing Education – Through ODL

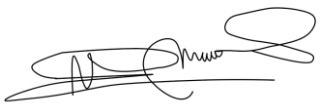
P. O. Box 20679
Dar es Salaam
Tanzania

Fax: +255 022 2150836
Email: info@iae.ac.tz
Website: www.iae.ac.tz

Acknowledgements

The Director of the Institute of Adult Education Dr. Michael Ng'umbi, wishes to extend his sincere thanks and gratitude to various individuals who contributed in one way or another in accomplishing this module. The following staffs of the IAE deserve special mention for tireless efforts and dedication in the production of this module:

Dr. Julius Chaligha:	Head, Adult and Continuing Education Studies Department
Mary Kimathy:	Institute of Adult Education (Writer)
Dr. Anathe Kimaro	Institute of Adult Education (Reviewer)
Eric Samba	Institute of Adult Education (Editor)



……

Dr. Michael W. Ng'umbi

Director

Institute of Adult Education

Contents

About this module

This module has been produced by the Institute of Adult Education. All modules produced by the Institute of Adult Education are structured in the same way, as outlined below:

How this module is structured

Module overview:

The module overview gives you a general introduction to the module. Information contained in the module overview will help you determine:

- If the module is suitable for you.
- What you already need to know.
- What you can expect from the module.
- How much time you will need to invest to complete the module.

The overview also provides guidance on:

- Study skills.
- Where to get help.
- Module assignments and assessments.
- Activity icons.
- Units.

We strongly recommend that you read the overview *carefully* before starting your learning.

Module content:

The module is broken down into units. Each unit comprises:

- An introduction to the unit content.
- Unit objectives
- Unit outcomes.
- New terminology.

- Core content of the unit with a variety of learning activities.
- A unit summary.
- Assignments and/or assessments, as applicable.

Resources:

For those interested in learning more on this subject, we provide you with a list of additional resources at the end of this module; these may be books, articles or web sites.

Your comments:

When reading this module, we would appreciate it if you would make a few comments to give us your feedback on any aspect of this module. Your feedback might include comments on:

- Module content and structure.
- Module reading materials and resources.
- Module assignments.
- Module assessments.
- Module duration.
- Module support (assigned tutors, technical help, and others).

Your constructive feedback will help us to improve and enhance this module.

Module overview

Welcome to this module

Dear learner, as prospective an adult education facilitator/expert, you will learn educational assessment, measurement and evaluation. The module will enable you to gain knowledge and skills on how to select and prepare test items which are vital for you to perform your duties as an adult education facilitator. This module is organized into four units. Unit one introduces key concepts in educational assessment, measurement and evaluation. Unit two is all about classroom tests. Unit three you will learn about instructional objectives. In the last unit we shall discuss about characteristics of a good assessment tool. I hope you will enjoy the module. Welcome!

General competence

After completing this module, you should be able to gain knowledge, analytical skills and understanding on educational assessment, measurement and evaluation. It enables you select and prepare various test items. More important, module will equip with knowledge and skills in asserting quality of a good assessment tools. This module will enable you to fulfill your duties and responsibilities as a learner and future adult education facilitator.

Study Skills

Essentially, you will be taking control of your learning environment. As a result, you will need to consider performance issues related to time management, goal setting, stress management, etc. Perhaps you will also need to acquaint yourself in areas such as essay planning, coping with exams and using the web as a learning resource.

Your most significant considerations will be *time* and *space* i.e. the time you dedicate to your learning and the environment in which you engage in that learning.

Need help?

Dear learner, in the course of your study, you may need help in various issues such as the location and how to get support from resource centres, clarification of various issues pertaining to your study materials i.e., modules, and so on. If this happens, you are advised to ask for the help from your centre coordinator or facilitator. You can also visit the website of the Institute of Adult Education, which is www.iae.ac.tz; or, call No. +255 22 2150838 and ask for help.

Module Assessment

After each unit, you will be required to attempt one unit assignment. This is not meant for submission, rather, for reflection on what you have learned in the whole module. You will also do tests and assignments for submission as guided by your module facilitator. Finally, you will sit for semester examinations to accomplish your assessment.

Icons used in this module

While working through this module, you will notice the frequent use of margin icons. these icons serve to "signpost" a particular piece of text, a new task or change in activity. they have been included to help you to find your way around the module.

A complete icon set is shown below. We suggest that you familiarize yourself with them and their meaning before starting your study.

 Reflection	 Assessment	 Assignment	 Help
 Learning Outcomes	 Module Outcome	 Help	 Reflection

Unit 1

Basic Concepts in Educational Assessment and Evaluation

Introduction

Dear learner, welcome to unit one which is all about basic concepts in educational assessment and evaluation. In this unit you will be able to define different concepts such as assessment, evaluation and measurement. Welcome!

Learning Outcomes

Upon completion of this unit you will be able to:

- Define assessment, evaluation and measurement;
- Explain measurement scale in relation to education;
- Explain types of educational evaluation and assessment;
- Explain uses of evaluation in adult education; and
- Differentiate between evaluation, assessment and measurement.

What is Assessment?

We begin by distinguishing among four related terms (Keeves, 1997; UNESCO, 2000b): measurement, testing, evaluation, and assessment. Measurement refers to the process by which a value, usually numerical, is assigned to the attributes or dimensions of some concept or physical object. For example, a thermometer is used to measure temperature while a test is used to measure ability or aptitude. Testing refers to the process of administering a test to measure one or more concepts, usually under standardized conditions. For example, tests are used to measure how much a student has learned in a course of mathematics. Evaluation refers to the process of arriving at judgments about abstract entities such as programs, curricula, organizations, and institutions. For example, systemic evaluations (e.g., national assessments) are conducted to ascertain how well an education system is functioning. In most education contexts, assessments are a vital component of any evaluation. Assessment is defined as “the process of obtaining information that is used to make educational decisions about

students, to give feedback to the student about his or her progress, strengths and weaknesses, to judge instructional effectiveness and curricular adequacy and to inform policy"(AFT, NCME, NEA,1990: 1).This process usually involves a range of different qualitative and quantitative techniques. For example, the language ability of learners can be assessed using standardized tests, oral exams, portfolios, and practical exercises. Assessment plays many roles in education and a single assessment can serve multiple, but quite distinct, roles. For example, results from a selection test can sometimes be used to guide instruction, while a portfolio of learner work culled from assessments conducted during a course of study can inform a decision about whether the learner should obtain a certificate of completion or a degree.4 Simplifying somewhat, we can posit that from a learner's perspective, there are three main roles for assessments: Choose, Learn, and Qualify. The data from an assessment can be used to choose a program of study or a particular course within a program. Other assessments provide information that can be used by the learner, teacher, or parents to track learner progress or diagnose strengths and weaknesses. Finally, assessments can determine whether learners obtain certificates or other qualifications that enable them to attain their goals. Assessment in the service of individual learning is sometimes referred to as "formative assessment, "in contrast to "summative assessment, "which is intended to guide decision-making (see Blackand Wiliam, 1998). From the perspective of the authorities, the three critical functions of assessment are: Select, Monitor, and Hold Accountable. One of the most important functions is to determine which learners are allowed to proceed to the next level of schooling. Assessment results, along with other measurement data (such as those obtained through periodic surveys), are also used to track the functioning of different components of the system (generally referred to as national assessments), and sometimes are used to hold accountable the individuals responsible for those components

I. Basic concepts

What is assessment?

The term ***assessment*** is generally used to refer to all activities teachers use to help students learn and to gauge student progress. Assessment is the process of collecting information about student learning for decision making. Assessment is the systematic, continuous process of monitoring the various pieces of learning to evaluate student achievement and instructional effectiveness. It includes tests, homework, assignments, class projects, class presentations, and participation and teacher observation. Assessment does not make value judgments to a process or programme; it just monitors the learning progress and nothing else.

What is evaluation? Why do we evaluate?

Dear learner, Evaluation simply is the process of determining the value or worth of something, process or programme in order to reach meaningful decision about that something, process or programme. It is the process that involves examining many components of a whole and making instructional decisions. At classroom level, evaluation can be defined as the systematic process of collecting, analyzing and interpreting information to determine the extent to which students are achieving instructional objective. The evaluation process focuses upon determining previously-established priorities and goals. Evaluation helps document the effectiveness of a course or program, identifies weakness and strengths and spots areas in need of revision. Generally, evaluation refers to a series of activities that are designed to measure the effectiveness of the programme or course. Evaluation involves making judgments about the decisions made in terms of set goals. Such judgment is based on a composite of different types of measurement obtained from tests, projects, assignments, laboratory work, reports, and examinations.

Measurement

What is Measurement? What are we measuring?

Measurement is a broad term that refers to the systematic determination of outcomes or characteristics by means of some sort of assessment device.

It is a systematic process of obtaining the quantified degree to which a trait or an attribute is present in an individual or object. In other words it is a systematic assignment of numerical values or figures to a trait or an attribute in a person or object. For instance what is the height of door? What is the weight of the meat? What is the length of the classroom?

In education, the numerical value of JOSEPHS' ability, aptitude, achievement, etc can be measured and obtained using instruments such as paper and pencil test. It means that the values of the attribute are translated into numbers by measurement. Measurement is concerned with finding out how well students are performing in terms of specified objectives

Generally, Measurement involves the assigning of numbers to represent the amount of a characteristic possessed by an object, event or system.

I. Measurement Scales

Measurement exists in several levels depending on what is to be measured, the instrument to be employed, the degree of accuracy or precision desired

and the method of measurement. There are four levels of measuring scales. These are nominal, ordinal, interval and ratio scales.

Nominal Scale

Dear learner, sometimes, you are involved in the classification of objects, things, human beings, animals, etc. For instance, you can classify human beings as males and females; you can also classify living things as plants and animals. Have you ever assigned your students into classes or groups?

Nominal scale is the simplest of the measurement scales involves only the assignment of numbers to classes or groups and does not imply magnitude. The numbers do not refer to any quantity but are only used to label and classify. It is used when we are interested in knowing if certain objects belong to the same or different classes. We may decide to assign teachers in the school system into two groups of graduate teachers = I and non-graduate teacher = 0. You will note that these numbers are just codes or labels, as they do not indicate magnitude.

We may group students into classes A, B, C D, etc. These are also codes or labels. There is no order involved here, there is nothing like greater than or less than. Any letter, number or numerals used in a nominal scale have no quantitative significance; they are used for convenience only.

Ordinal Scale

This is the next scale in simplicity. Unlike the nominal scale, in this scale there is order. The number indicates rank order among objects or events with qualitative property. One real number may be greater than or equal to or less than another real number. For instance, the number 5 is greater than the number 3. There is classification, as well as indication of size and rank. For example of ordinal measurement is the ranking of students in a classroom according to their test scores say from the first, second, third to nth positions.

You will note that what is important here is the position of the individuals or things or events in a group. The ranks cannot be compared. For instance, you cannot say that a student who is 2nd in a test is twice as good as the student who is ranked 4th and half as good as the student ranks 1t, in spite of the fact that 2 is half of 4 and 1 is half of 2.

Let us illustrate this with an example. Take the scores of four students ranked 5th, 6th, 7th and 8th as 70%, 68%, 50%, and 49% respectively. You note that the difference between 5th and 6th positions is only 2%, while between 6th and 7th positions, we have, not 2% but 18%. And the difference between 7th and 8th positions is just 1%. It means therefore that equal intervals on the scale do not represent equal quantities. The distances between the objects or events are not necessary equal.

Interval Scale

This is the next scale after ordinal scale. This measurement scale contains the nominal and ordinal properties and it is also characterized by equal units or interval between score points. In this scale the equal intervals represent equal quantities. The amount of difference between adjacent intervals on the scale is equal. Take the calendar as an example, days on it, represent equal amounts of time. You will observe that equal amount of time separates 2, 4, 6, and 8 days.

In an examination, a student who scored 80% is 10% greater than the student who scored 70% and is 10% less than one who scored 90%. Because the data here are continuous, some arithmetic operations like addition, subtraction, can take place here.

In interval scale, the numbers are arbitrary units from an arbitrary origin. But there is no absolute zero in the interval scale. For example, the Fahrenheit or the centigrade scale: -10^{o}, -5^{o}, 0^{o}, 5^{o}, 10^{o}…100^{o}C. When we say arbitrary unit implies that the length or the distance between one unit and another is arbitrarily determined. And by arbitrary origin we mean that zero does not indicate the complete absence of property. For example, if you give a test in your class and a student scores zero, does it imply that the student does not know anything at all? Take the case of the calendar again as an example. Look at the calendars you have from different sources and places. Is there any one with zero point? Can there be any zero day or time? Another example, if you use a test to measure intelligence of students, any student who scores zero does not imply he/she has no intelligence at all.

Ratio scale

This is the highest level or scale of measurement. It has all the characteristics of the others. In addition there is an absolute zero here. The absolute zero indicates a total absence of property, the zero point is real. Let us use the meter rule as an example of ratio scale. This scale has equal intervals as in the intervals scale. It has a zero mark, which implies that at this point there is complete absence of what the meter rule measures. If it is used to measure height, it means that at zero, there is no height at all.

Let us use another example to drive this point home. If you use a weighing balance to measure the weights of objects, you will discover that at zero mark there is a complete lack of weight. There is no weight at all. Another example, if students get nothing in objective tests such as multiple choice items or matching items that means they scored zero mark in that test.

Activity

Differentiate between interval scale and ratio scale

Dear learner, before proceeding let us know difference between measurement, assessment and evaluation. Try to share with your colleague, then compare your answers with the following:

II. What is the difference between measurement, assessment and evaluation?

Dear learner, Measurement, assessment, and evaluation mean very different things, and yet most of students are unable to adequately explain the differences. Measurement refers to the process by which the attributes or dimensions of some physical object are determined. One exception seems to be in the use of the word measure in determining the IQ of a person. The phrase, "this test measures IQ" is commonly used. Measuring such things as attitudes or preferences also applies. However, when we measure, we generally use some standard instrument to determine how big, tall, heavy, voluminous, hot, cold, fast, or straight something actually is. Standard instruments refer to physical devices such as rulers, scales, thermometers, pressure gauges, etc. We measure to obtain information about what is. Such information may or may not be useful, depending on the accuracy of the instruments we use, and our skill at using them. In all of these examples, we are not assessing anything; we are simply collecting information relative to some established rule or standard. Assessment is therefore quite different from measurement, and has uses that suggest very different purposes.

Assessment is a process by which information is obtained relative to some known objective or goal.

The term ***assessment*** is generally used to refer to all activities teachers use to help students learn and to gauge student progress. Assessment is the process of collecting information about student learning for decision making. Assessment is the systematic, continuous process of monitoring the various pieces of learning to evaluate student achievement and instructional effectiveness. It includes tests, homework, assignments, class projects, class presentations, and participation and teacher observation. Assessment does not make value judgments to a process or programme; it just monitors the learning progress and nothing else.

Evaluation is perhaps the most complex and least understood of the terms. Inherent in the idea of evaluation is "value." When we evaluate, what we are doing is engaging in some process that is designed to provide information that will help us make a judgment about a given situation. Generally, any evaluation process requires information about the situation in question. A situation is an umbrella term that takes into account such ideas as objectives, goals, standards, procedures, and so on. When we evaluate, we are saying

that the process will yield information regarding the worthiness, appropriateness, goodness, validity, legality, etc., of something for which a reliable measurement or assessment has been made. We evaluate every day. Teachers, in particular, are constantly evaluating students, and such evaluations are usually done in the context of comparisons between what was intended (learning, progress, behaviour) and what was obtained.

Dear learner, have you notice the differences? Generally we can to sum up by saying, we measure distance, we assess learning, and we evaluate results in terms of some set of criteria. These three terms are certainly share some common attributes, but it is useful to think of them as separate but connected ideas and processes.

Types of Educational assessment and Evaluation

Concept of educational assessment

Dear learner let us familiarize with the concept of educational assessment. The term Educational assessment refers to the wide variety of methods that educators use to evaluate, measure and document the academic readiness, learning progress and skills acquisition of students. Educational assessment is the process of determining the extent to which learners have acquired specified knowledge, skills and attitudes. It can be viewed as the systematic, continuous process of monitoring the various pieces of learning to evaluate student achievement and instructional effectiveness. In general educational assessment is the systematic collection, review and use of information about educational program to improve student learning. In educational assessment, the information collected can be used for making decision about students, curriculum and programme and even educational policy. It focuses on what students know, what they are able to do and what values they have when they graduate.

Dear learner, educational assessment can be either **formal** or **informal.** Formal assessment usually implies a written document, such as a test, quiz, or paper. A formal assessment is given a numerical score or grade based on student performance. It is timed, taken under similar conditions or situations by the learners. It is seen as tests and is usually performed in a group. An informal assessment usually occurs in a more casual manner and may include observation, inventories, checklists, rating scales, rubrics, performance and portfolio assessments, participation, peer and self-evaluation, and discussion. It is untimed and taken under varying situations. Unlike formal assessment, informal assessment does not contribute to a student's final grade.

Educational assessments come in many complex forms and they are used for a wide variety of purposes. The followings are few major forms of educational assessment.

Diagnostic assessments

This type of assessment is also referred to as pre-assessment or initial assessment. Diagnostic assessments are conducted prior to instruction or intervention to establish a baseline from which individual student growth can be measured. Diagnostic assessment measures a student's current knowledge and skills for the purpose of identifying a suitable program of learning. Self-assessment is a form of diagnostic assessment which involves students assessing themselves.

Formative assessments

Formative assessment is generally carried out throughout a course or project. In an educational setting, formative assessment might be a teacher or the learner, providing feedback on a student's work and would not necessarily be used for grading purposes. Formative assessments can take a variety of forms, from more formal quizzes and assignments to informal questioning techniques and discussions with students. Formative assessment is often referred to in a learning context as *assessment for learning* due to the fact that it is used by teachers to consider approaches to teaching and next steps for individual learners and the class.

Summative assessments

Summative assessment is generally carried out at the end of a course or project. In an educational setting, summative assessments are typically used to assign students a course grade. Summative assessments are evaluative. The summative assessments are made to summarize what the students have learned, to know if they understand well. This type of assessment is graded and often counts, it can be in form of tests, final exams, projects, etc. These assessments are important because they decide if the student passes or fails the class.

Dear learner, it is importance to be in mind that summative assessment is often referred to in a learning context as *assessment of learning.* It is called so because it is intended to measure learning outcomes and report those outcomes to students, parents and administrators. Assessment of learning generally occurs at the conclusion of a class, course, semester or academic year.

Criterion-referenced assessment

Dear learner, Criterion-referenced assessment as the name implies, occurs when candidates are measured against defined goal, objective, or a set standard criteria. Criterion-referenced assessment is often, but not always, used to establish a person's competence whether she /he can do something.

The best known example of criterion-referenced assessment is the driving test, when learner drivers are measured against a range of explicit criteria (such as "Not endangering other road users").

Norm-referenced assessment

This type of assessment is not measured against defined criteria. It is effectively a way of comparing student's performance against a national or other norm group. For example, Juma has got 80% in maths while Joseph got 40%. So when you compare the performance of one student to another student who doing the same assessment we are talking about norm-referenced assessment. Many entrance tests are norm-referenced, permitting a fixed proportion of students to pass. "Passing" in this context, means being accepted into the school or university rather than an explicit level of ability. This means that standards may vary from year to year, depending on the quality of the group. While criterion-referenced assessment does not vary from year to year unless the criteria change.

Before proceed, relax and share with you colleague about types educational assessment. Can you explain other types of educational assessment that you have shared with your friends?

Educational evaluation

Dear learner, educational evaluation is the process of determining the degree to which goals of a programmed have been achieved. It is the process by which teachers, schools/institutions and students are evaluated to indicate the effectiveness of education and the degree to which students integrate and understand information. It is concerned with the overall effectiveness of a programme.

Types of Evaluation

The different types of evaluation are: placement, formative, diagnostic and summative evaluations.

Placement Evaluation

This is a type of evaluations carried out in order to fix the students in the appropriate group or class. In some schools for instance, students are assigned to classes according to their subject combinations, such as science, Technical, arts, Commercial etc. before this is done an examination will be carried out. This is in form of pretest or aptitude test. It can also be a type of evaluation made by the teacher to find out the entry behaviour of his students before He/She starts teaching. This may help the teacher to adjust

his lesson plan. Tests like readiness tests, ability tests, aptitude tests and achievement tests can be used.

Formative Evaluation

This is a type of evaluation designed to help both the student and teacher to pinpoint areas where the student has failed to learn so that this failure may be rectified. It provides a feedback to the teacher and the student and thus estimating teaching success e.g. weekly tests, terminal examinations, etc.

Diagnostic Evaluation

This type of evaluation is carried out most of the time as a follow up evaluation to formative evaluation. As a teacher, you have used formative evaluation to identify some weaknesses in your students. You have also applied some corrective measures which have not showed success. What you will now do is to design a type of diagnostic test, which is applied during instruction to find out the underlying cause of students persistent learning difficulties. These diagnostic tests can be in the form of achievement tests, performance test, self-rating, interviews, observations, etc.

Activity

What are the major differences and similarities between formative evaluation and diagnostic evaluation?

Summative evaluation

This is the type of evaluation carried out at the end of the course of instruction to determine the extent to which the objectives have been achieved. It is called a summarizing evaluation because it looks at the entire course of instruction or programme and can pass judgment on both: the teacher and students, the curriculum and the entire system. It is used for certification. Think of the educational certificates you have acquired from examination bodies such as NECTA, etc. These were awarded to you after you had gone through some types of examination. This is an example of summative evaluation. Let us now discuss the roles of educational assessment and evaluation.

Uses of Assessment and Evaluation

Why assessment and evaluation is done?

Within education, curiosity and engagement are important, and essential to academic success. How to create those traits within an educational environment becomes an issue of great debate.

In adult education, the evaluation is relatively complex process which may start with assessing changes in individuals or with determining the

effectiveness of the total program. The major reason for evaluation is the improvement of program. The results of evaluation can be fed back continuously to stimulate and guide further improvement. By seeing what has been done and what is yet to be done, we gain insight into the factors necessary to help adults achieve the goals they desire. By taking a critical look, we can see deficiencies in method. We can discover unmet needs. we can review our goals and set more specific objectives for the immediate future we may find a need for more facts, more research, more thinking (Thatcher, 1963, p. 175).

Dear learner, educational evaluation and assessment are carried out from time to time for different purposes or uses. Among the purposes of educational assessment and evaluation include:

Promotion and classification of learners

Promotion of the students (to both adult learners and young learners) to the next grade or level is dependent on a learner's achievement. Assessment and evaluation results provide a basis for determining the promotion of students from one class to another. It also plays role of classification or streaming of students according to ability or subjects.

Reach out set goals

To determine how near the individual student, and the class as a whole, come to reaching the goal that they set out to attain.

To provide information which will be useful to the student the instructor, and the public at large.

Improving teaching/* or *modification of teaching methods.

To determine the effectiveness of specific teaching materials, methods and activities.

Evaluation results can help teachers to determine the effectiveness of teaching techniques and learning materials. They guide teachers on the approaches and methods to use for instruction as they can pinpoint learners individual differences.

For modification of the curriculum purposes

Evaluation and assessment results are used to indicate the effectiveness of curriculum implementation. They determine the relative effectiveness of the programme in terms of students behavioural output. The evaluation results are used to identify problems that might hinder or prevent the achievement of set goals.

Certification

The results are used to award certificates .They help to certify that a student has achieved a particular level of performance.

For guidance and counselling services

Assessment results can be used to identify areas in which learners need remedial teaching. Teachers may use those results to provide guidance and counselling services to students so as to improve their learning progress.

For reporting students' progress

Assessment and evaluation are used for providing feedback on the achievement to teachers, learners and other stakeholders

Stimulating learning

Assessment and evaluation results can be used to motivate learners to work harder, promote healthy competition and enhance learning. They help to motivate students to learn more as they discover their progress or lack of progress in given tasks. The results are used to encourage learners to develop a sense of discipline and systematic study habits.

For the purpose of selecting students for employment

The results can provide a basis for determining at what level of education possessor of a certificate (student) should enter a career.

Can you mention other purposes of education that you? If yes try. Within education, curiosity and engagement are important, and essential to academic success. How to create those traits within an educational environment becomes an issue of great debate. Mainstream educational institutions seem to rely on the use of rewards

Unit Reflection

Dear learner, this unit was about the meaning of the terms educational assessment, evaluation and measurement. In this unit you also learnt about the measurement scale, and types of assessment and evaluation as well as differences between measurement, assessment and evaluation. Lastly you learnt the purposes of the assessment and evaluation in teaching and learning process. Get prepared for unit assignment to assess your understanding of the unit.

Unit Assignment

1. What is assessment?
2. Normally terms assessment, evaluation and measurement are tend to be used interchangeably, but they mean differently. Differentiate these three terms
3. List 5 purposes of measurement and evaluation?

Unit 2

Carrying Out Evaluation in Adult Education

Introduction

Dear learner, welcome to unit two which is all about carrying out evaluation in adult education. In this unit you will be informed about methods of evaluation, design evaluation tools and administering evaluation tools.

Learning Outcomes

Upon completion of this unit, you will be able to:

- Identify evaluation methods;
- Design evaluation/assessment tools;
- Administer evaluation tools;
- Classify different types of tests;
- Explain the meaning of objective test and essay test;
- Describe the different types of objective test items;
- Describe the rules of constructing good essay questions and objective test items; and
- State the advantages and disadvantages of the objective test.

Methods of Evaluation

Methods of evaluation vary with the type of program to be evaluated. Some methods will yield objective data and other methods will rely on subjective. The choice of methods depends on the desired information and the cost of the method is weighed against the value of the information the method will produce. Evaluation can be classified into three levels: Informant evaluation, semi-formal evaluation, and formal evaluation which is also known as scientific research.

Informal Evaluation

The everyday judgment of one's work is informal evaluation. All adult educators do some of this judgements, even though they may not refer to it

as evaluation. Informal evaluation is less precise and less reliable than either of the other two classes, but it is useful in many teaching situations and one can greatly improve his ability to evaluate informally through study and practice. Some of the tools of information for informal evaluation are: observation of the work done by the learners each time the group meets, giving informal tests, talking to the learners outside of scheduled meetings, noting the changes in the learners' practices in their homes, in their places of work and in their community, and conferring with other teachers and facilitators.

Semi-formal evaluation

This is more practical in adult education. It is a more systematic and reliable type of evaluation than the informal type. It is the kind that the average teacher will use in determining the quality of his/her teaching and the achievement of his/her students. Much of the evidence in semiformal evaluation comes from these sources: reliable and valid tests, surveys to determine changes in practices, objective scales, and rating sheets, and the judgments of experts.

Formal evaluation

This also known as a research. It is a type of scientific research in which all of the pertinent facts are gathered, and the variables controlled by means of statistical procedures. This Kind of evaluation usually requires considerable time and much training in education field. It can be done by a specialist in the field of education. Information collected from an evaluation is also a base for assessing the effectiveness of instruction. The purpose of instruction is to facilitate learning, and the purpose of the evaluation is to determine the extent of learning. The quality of instruction is determined to a great extent by the result obtained from evaluations. An evaluation may also be used in making decisions about the practicality of a course or program.

Assessment Tools

Concept of Test as a Measuring tool

Dear learner, a test is a tool or device used for finding out the presence or absence of a particular phenomenon or trait possessed by an individual or group of individuals. A test is the major and most commonly used tool for the assessment of cognitive behaviors. In this context test simply means a set of questions that students are expected to answer. Their responses to the questions give a measure of their level of performance or achievement.

Usually, test is based on learned content of subject-specific area(s) and is directed to measure the learner's level of attainment of pre-specified objectives. Dear learner, do you know that to measure an attribute, a standard tool is needed? Therefore, assessments are done by describing the characteristics associated with such constructs in behavioral terms.

The expected behaviors (aptitude) such as ability to state, define, manipulate or perform experiment for instance in integrated science and similar activities are put down in form of test. The test score gives quantitative information about the existence of the construct (attribute) possessed by the testee. For this reason, the test items as measuring tool must be valid, reliable and usable in order to give dependable result.

Types of Tests

Dear learner, test may be classified in terms of what they measure, length and the ability they require. Based on this context, there are different types of tests. These types include: diagnostic tests, formative tests, achievement tests, placement tests, intelligence tests, aptitude tests, prognostic tests, performance tests, readiness tests, teacher- made tests, speed tests, norm-referenced tests, and criterion- referenced tests. Tests may also be classified in terms of format of test items. In this case we may have essay tests and objective tests.

Intelligence Test

Dear learner, intelligence is the ability to reason and learn from experience. Intelligence test provides an indication of an individual's general mental capacity. An Intelligence test usually includes a wide variety of tests so as to sample several aspects of cognitive function. It is thought to depend both on inherited ability (nature) and on surroundings in which a person is brought up (nurture). The first intelligence tests were devised by Alfred Binet in 1905 to give an Intelligence Quotient (IQ).

Aptitude Tests

When we talk about aptitude, we refer to the natural talent or ability especially specified. Thus, aptitude tests measure specialized abilities and the potential to learn or perform new tasks that may be relevant to later learning or performance in a specific area. They measure an individual's natural ability that makes it easy for them to do something. Hence they are future oriented, thus they used to predict the future performance of an individual in a specific task.

Achievement Test

Achievement tests are designed to measure the effects of a specific

programme of instruction or training which the learners attained usually by their effort. Generally, they represent a terminal evaluation of the learner's status on the completion of a course of study or training. That is, it is used to determine how much the learner has learned from specified content via systemic and controlled instructions. End of term examinations and classroom tests are mostly achievement tests.

Performance tests

These type of tests measure an ability of an individual to manipulate something. These are individual tests in which the learners are expected to manipulate objects, pictures, blocks or mechanic apparatus.

Readiness tests

These type of test measures a learner's maturity and previous knowledge before starting a new learning activity. They measure the readiness of the learners to start new a instruction. Or lesson.

Teacher- made tests

These tests also are called classroom tests. Teachers used these tests to assess students learning outcomes. The tests can be achievement or performance and /or any other type of test like practical tests, etc. prepared by the teacher for his/her specific class and purpose based on what he/she has taught. Dear learner let us now discuss the types of test items which are commonly used in the classroom to assess our learners.

Types of test items used in the classroom

Dear learner, there are two common test items that can be used in the classroom. These are: essay test items and objective test items. All these type of items are used by teachers to measure achievement and performance from classroom instruction.

Essay tests

In essay tests the students have the responsibility of thinking out the answers to the questions asked. They have the freedom to express or state the answers in their own words. It is a free answer kind of test. They are also referred as subjective test items.

There are two types of essay tests. These are: Extended – response type and Restricted response

Extended Response Type of Essay Test or long essay tests

In this type, questions are asked in such a way that the answers demand that the student is not limited to the extent to which he has to discuss the issues raised or question asked. The student has to plan and organize his thoughts in order to give his answer. Put his ideas across by expressing himself freely, precisely and clearly using his own words and his own writing.

Discuss the questions at length, giving various aspects of his knowledge on the question asked or issue raised.

Restricted – Response Type of essay test or short essay tests

In this type, the questions are so structured that the students are limited, the scope of the response is defined and restricted. The answers given are to some extent controlled. Student is required to treat the questions as briefly as possible.

Advantages of Essay Test

i. It measures complex learning outcomes that cannot be measured by other means. For instance, it has the ability to measure learner's communication skills. That is, the learner's ability to produce an answer, synthesize and organize ideas and present them readably in a logical and coherent form. This is the major advantage.

ii. It also enables the measurement of organizational and divergent thinking skills by laying emphasis on the integration and application of thinking and problem solving skills, creativity and originality.

iii. It is very applicable for measuring learning outcomes at the higher levels of educational objectives such as application, analysis, synthesis and evaluation of levels of the cognitive domain.

iv. It is easy and economical to administer.

v. Essay item is easy to construct and does not take much time. This fact has to be guarded seriously to avoid constructing questions that can be very misleading by not asking for specific behaviors emphasized in a particular set of learning outcomes.

vi. It can be used to measure in-depth knowledge especially in a restricted subject matter area. • It does not encourage guessing and cheating during testing.

Disadvantages of Essay Test

Despite the advantages already proffered for essay test, it does not satisfy the two most important qualities of a good measuring tool. Its disadvantages include:

i. It is inadequate in sampling subject matter content and course objectives since it provides limited sampling. The provision of few questions results in the invalid and narrow coverage of subject matter and instructional objectives.

ii. In addition to the invalidity of the measurement, evaluating the answers to carelessly developed questions tends to be confusing and time consuming task. This results in poor reliability in scoring. Studies have shown that answers to essay questions are scored

differently by different teachers and that even same teachers score the answers differently at different times.

iii. Sometimes an essay question implies many skills other than that which the item was intended to measure. The testee therefore perceives and reacts to the same questions differently. The differences in the perception of the questions encourage bluffing and hide differences in the knowledge of basic factual material and the learner's ability to use and organize such facts.

Principles of constructing essay questions

Dear learner, the following points are suggested as guide for construction of good essay test items that call for the desired behavior.

i) Restriction of the use of essay questions to only those learning outcomes that cannot be satisfactory measured by objective items.

ii) Formulation of questions that call forth behavior specified in the learning outcomes.

iii) Phrasing each question to clearly indicate the examinees task. An essay question has to specify precisely what is required of the examinee. Ensure that the testee's task is clearly indicated by delimiting the area covered by the item, using descriptive words to give specific direction towards the desired response.

iv) Indication of approximate time limit for each question. It is necessary to indicate time fixed to each question to enable the testees to pace their writing on each question.

v) Avoidance of the use of optional question.

Objective tests

In objective tests, the students are provided with a problem to which a limited numbers of choices are presented for them to select the wanted answer. It is so much structured that even where the student is to supply the answers, he/she is strictly limited to give specific and short answers. In objective tests students are given the opportunity to react to a large sample of questions which may cover the entire content area.

Dear learner, the objective tests can be classified into those that require the examinee to supply the answer to the test items (free-response type) and those that require the examinee to select the answer from a given number of alternatives (fixed response type). The free-response type consists of the short answer and completion items while the fixed response type is commonly further divided into true-false, arrangements, matching items and multiple-choice items

Multiple choice items

This is very popular and a widely used type of objective test items. It has two parts. The first is called the stem, which is the question. The stem may

be stated as a direct question or an incomplete statement, while the second part is the suggested answers called alternatives, options or choices. The correct option is the answer or the key while other options are called distractors or distracters. Before we give examples of multiple choice items, let us examine the following tips on how to make good multiple choice items.

Advantages

- The main advantage of multiple-choice test is its wide applicability in the measurement of various phases of achievement.
- It is the desirable of all the test formats being free of many of the disadvantages of other forms of objective items. For instance, it present a more well-defined problem than the short-answer item, avoids the need for homogenous material necessary for the matching item, reduces the clues and susceptibility to guessing characteristics of the true-false item and is relatively free from response sets.
- It is useful in diagnosis and it enables fine discrimination among the examinees on the basis of the amount of what is being measured possessed by them.
- It can be scored with a machine.

Disadvantages

- It measures problem-solving behavior at the verbal level only.
- It is inappropriate for measuring learning outcomes requiring the ability to recall, organize or present ideas because it requires selection of correct answer.
- It is very difficult and time consuming to construct.
- It requires more response time than any other type of objective item and may favour the test-wise examinees if not adequately and skillful constructed.

Principles of constructing multiple choice items

i. State the stems of the item in simple and clear language.
ii. All alternatives should be possible answers, related to the stems but with only one most correct or best answer to the question. In other words, all alternatives should be plausible and attractive.
iii. The distractors should be as attractive as the key and very effective in distracting testees who are not sure of the key.
iv. To avoid guessing the correct answers should not be positioned in a particular way so as to form a patter. It should be randomly

positioned.

v. Alternatives should be a minimum of four options for each item and maximum should be five or more.

vi. Avoid as much as possible the options 'none of the above', 'all of the above' or both A and B or a combination of options.

vii. State the stem of the item in positive form wherever possible. Avoid to state question negatively because negative tend to confuse students. If you cannot avoid a negative stem, then **bold, capitalize or underline** the negative word. Sometime instead of use NOT you may use EXCEPT.

Can you now construct good multiple choice items using these hints?

Matching items

The matching test items usually consist of two parallel columns. One column is called premises which contain a list of word, number, symbol or other to be matched to a word, sentence, phrase or other possible answer from the other column called responses. The examinee is directed to match the responses to the appropriate premises. Usually, the two lists have some sort of relationship. Although the basis for matching responses to premises is sometimes self-evident but more often it must be explained in the directions.

Advantages

i. The major advantage of matching exercise is that one matching item consists of many problems. This compact form makes it possible to measure a large amount of related factual material in a relatively short time.

ii. It enables the sampling of larger content, which results in relatively higher content validity.

iii. The guess factor can be controlled by skillfully constructing the items such that the correct response for each premise must also serve as a plausible response for the other premises.

iv. The scoring is simple and objective and can be done by machine.

Disadvantages

i. It is restricted to the measurement of factual information based on rote learning because the material tested lend themselves to the listing of a number of important and related concepts.

ii. Many topics are unique and cannot be conveniently grouped in homogenous matching clusters and it is sometimes difficult to get homogenous materials clusters of premises and responses that can sufficiently match even for contents that are adaptable for clustering.

iii. It requires extreme care during construction in order to avoid encouraging serial memorization rather than association and to avoid irrelevant clues to the correct answer.

Unit Summary

Dear learner, in this unit, you have learnt about evaluation methods, how to design assessment and evaluation tools as well as administration of assessment tools. Welcome to the assignment for assessing yourself.

Unit Assignment

1. Construct five matching item questions in any course of your choice.
2. Discuss four principles of essay test items items
3. Explain three differences between essay questions and objective questions.
4. State the advantages and disadvantages of objective test;

Unit 3

Describing Instructional Objectives

Introduction

Dear learner, this unit guide you how to state instructional objectives and how these educational objectives are classified according to Benjamin Bloom (1956). The unit also discusses in details the different levels of cognitive, affective and psychomotor domains in hierarchal order stating from simplest behaviour to most complex one. In this unit you are also going to learn how to plan a test.

Learning Outcomes

Upon completion of this unit, you will be able to:

- Define the term instructional objectives;
- Explain importance of instructional objectives;
- Identify levels of cognitive, affective and psychomotor domains;
- Construct test items which measure different levels of learning domains;
- Identify factors to consider during test planning; and
- Prepare a table of specification.

What are Instructional Objectives?

Dear learner, an instructional objective is statement that specifies in behavioral terms /measurable terms what a learner will be able to do as a result of instruction. The Instructional objectives are stated in terms of desirable behavior in the cognitive, affective and psychomotor domains. They are also called specific objectives or behavioural objectives since they are used to measure student performance terminal behaviours after a particular instruction

The roles of instructional objectives include:

- These objectives provide direction to teaching; and
- They aid the teacher in making decisions regarding the types of methods, activities, and materials he has to provide.

To be most useful, objectives must be stated in such a way they are attainable, observable, and measurable. When teachers prepare a lesson to teachhing they usually formulate general and specific objectives. The general objectives specify in general terms what the final desired learning

outcome is. But it is the specific objectives that usually serve as a blueprint for the construction of measuring instruments. Gronlund (1972) suggests that teachers should begin with a statement of general objectives that can be made more specific by listing examples of behaviors that reflect the general objective. Examples on how to state instructional objectives:

By the end of the lesson the students should be able to correctly name the main sources of revenue in Tanzania

By the end of the lesson (40 minutes), the students should be able to correctly define the terms educational assessment

Dear learner, instructional objectives should be SMART, thus should be Specific that describe who will perform the behaviour. Here we talk about the students/learners is expected to do. They should be Measurable that describe learning outcomes to be measured, thus measurable student terminal behaviour at the end of period of instruction. They should Attainable, Realistic and Time bonded.

Taxonomy of Educational Objectives

Dear learner, educational objectives or instructional objectives are classified by Bloom (1956) into cognitive, psychomotor, and affective domains. The taxonomy represents an educational system in that the categories correspond to a teacher's concern in developing curricula and selecting learning objectives. It represents a set of behavioral goals, as well as a system for developing learning goals. These frameworks are arranged in a hierarchy, with the lowest level of behavior believed to be the least complex and its achievement is presumed to be the key to successful achievement at the next higher level.

Cognitive Domain

Cognitive domain involves mental skills, intellectual capability. It deals with thinking, memory, reasoning, creativity, intellectual discourse and problem solving. According to Bloom, there are six areas of cognitive objectives on which tests can be based. The six areas or levels represent a hierarchy of learning that goes from simple behaviour to the most complex and from concrete to abstract. It should be known that the first one must be mastered before the next can take place. These levels are as follows:

i. **Knowledge**: It involves the recognition or recall of previous learned information. In order to help learners to recall what they learnt, facilitator may use action verbs or key words like: *write, list, label, name, state, define, outline and identify* in asking questions.

ii. **Comprehension**: it deals with understanding the meaning,

translation and interpretation of information based on prior learning. In this level requires the individual be able to paraphrase knowledge accurately, explain or summarize it in his own words, comprehend or interpret information. Examples of key words: *explain, summarize, paraphrase, describe, illustrate, re-arrange.*

iii. **Application**: it deals with the ability to select a given concept (idea, rule, procedure, or generalized method) appropriate to a new situation and to apply it correctly. In simple words, application deals the use of a concept or application of factual information, principles, methods and procedures in a new situation.

Examples of key words: *use, compute, solve, demonstrate, apply, show, relate*, *prepare, manipulate*, calculate

iv. **Analysis:** it deals with the ability to separate or break down of materials or concepts into its component parts so that the organization structure may be understood. I.e. distinguish facts from inferences, recognize the hidden meaning. Examples of key words: *analyze, compare, separate, categorize, contrast, distinguish, differentiate, breakdown, relate.*

v. **Synthesis:** deals with building a structure or pattern from diverse elements i.e. ability to put together parts and elements to form a new whole. In other words, synthesis means ability to use old ideas to create new one. Student is expected to produce something which is unique and different. He/she may originate, integrate, and combine ideas into a new meaning or structure or product or plan. Examples of key words: *create, design, invert, develop, devise, rewrite, modify, combine, compose, compile, and organize.*

vi. **Evaluation**: it deals with ability to judge the value of materials or ideas by using criteria or standard data. Students may appraise, assess or critique on a basis of specific standard and criteria. Examples of key words: *judge, recommend, critique, appraise, conclude, criticize, defend, evaluate, interpret, justify, support, relate, account.*

Dear learner, I hope now you are familiar with levels of cognitive domain. To be conversant in this domain I advise you to find more information concerned about it. Let us discuss another domain which is affective. Please join the discussion that follow.

Affective Domain

Dear learner, this domain includes objectives related to attitudes, feelings, interests, appreciation, values, emotions, motivations. These things are intangibles. Things that are hard or difficult to touch, define, describe and measure. Affective domain has five hierarchical categories /levels namely: receiving, responding, valuing, organization and characterization.

i. Receiving

This is the lowest level of the learning outcomes in the affective domain. It deals with awareness, willingness and selected attention. It has three sub-levels:

***Awareness*:** This involves the conscious recognition of the existence of some problems, conditions, situations, events, phenomena. It means be aware of or attending something in the environment. Take for instance as a teacher, you come into your class while the students are making noise. You will notice that the atmosphere will change. This is because the students have become aware of your presence. They are merely aware.

Willingness: This is the next stage which involves the ability to acknowledge the object, event, problem instead of ignoring or avoiding it. It is the learner's willingness to attend to a particular stimulus or his being sensitive to the existence of a given problem, event, condition or situation. The students in your class kept quite because they noticed and acknowledged your presence. If they had ignored your presence they would continue to make noise in the class.

Controlled or selected attention*:* This involves the learner selecting or choosing to pay attention to the situation, problem, event or phenomenon. When you teach in the class, the learner is aware of your saying or the points you are making. In that case he will deliberately shut off messages or speeches or sounds as noises. Receiving in a classroom situation involves getting, holding and directing the attention of the learners to whatever the teacher has to say in the class.

ii. Responding

This involves ability to attend and react to a particular phenomenon by doing something. In this case the learner responds to the event by participating. For example, when students are actively participating in the learning process like discussion, give presentation, questions new concepts in order to understand them. Another example, If in your class you set a test for your students, first the students become aware of the test, they are willing to take the test, they now select to do it and they react by doing it.

iii. Valuing

This is concerned with the worth or value or benefit which a learner attaches to a particular object or behaviour or situation. The valuing of a person is ranging in degree from simple acceptance of value to a more complex level of commitment or an assumption of responsibility.

iv. Organization/organizing

In this level, the learner starts to bring together different values as an

organized system. He determines the interrelationships and establishes the order of priority by comparing, relating and synthesizing the values. He then builds a consistent value system by resolving any possible conflicts between them. If the learner tries to successfully internalize the value, he may encounter some situations which may demand more than one value. In this case, he has to organize the values into a system in order to decide which value to emphasis. Example, student can put together different value, information, ideas and accommodate them with his/her own schema/plan.

v. Characterization /characterizing

At this stage the value system is so internalized by the people or individuals so that they act consistently in accordance with such values, beliefs or ideals. This level has a value system that controls the behavior. A life-style which reflects these beliefs and philosophy are developed. The behavior of such individuals or groups can be said to be controlled by the value system.

Psychomotor Domain

The psychomotor domain includes physical movement, coordination and use of motor skill areas. It deals with muscular activities or motor skills which involve the use of the limbs (hand) or the whole of the body.

Can you think of such abilities or skills? Consider the skills in running, walking, swimming, jumping, eating, playing, throwing, etc.

Dear learner, skills in this domain describes the ability to physically manipulate something/ object. These skills are measured in terms of speed, distance, procedures or techniques in execution. The development of these skills requires practice.

The domain has several levels starting from the simplest behaviour to the most complex. These levels include: perception, set, guided response, mechanism, complex overt response, adaptation and origination.

i. Perception (awareness)

This level deals with the ability to use sensory clues or signs to guide motor activity.

ii. Set. (Readiness to act)

This level deals with the desire to learn a new process. It includes mental, physical and emotional sets. These three sets are disposition that determine a person's response to different situations. Sometimes are called mindsets.

iii. Guided responses

This is the early stages in learning complex skills. It includes imitation/practice i.e. observing and patterning behaviour after someone else. It also includes trial and error stage. Example, when student follows instruction to build a model, reproduce/respond.

iv. Mechanism

This is the intermediate stage in learning complex skills. In this stage the learnt responses have become habitual or routine and the movements can be performed with some confidence and skill. For instance, ability to demonstrate correct actions or procedures, like drive a car, use a computer

v. Complex overt responses (expert)

This category includes performing actions without hesitation and automatic performance. E.g. ability to demonstrate actions quickly and accurately like, operate a computer quickly and accurately.

vi. Adaptation

This level shows the ability to adjust to new situations. Respond effectively to unexpected experiences. In this level, skills are well developed and the individuals can modify movement patterns to fit special requirements.

vii. Origination

This level shows creativity/originality or innovation. In this level, an individual can create new movement patterns to fit a particular situation or specific problem.

Dear learner, before proceed to the next session, relax for a while then share with you colleagues the action verbs or key words that are used by teachers to construct test items which measure each level in cognitive, affective and psychomotor domains. Can you mention those action verbs? If not yet we shall discuss together during face to face session.

Dear learner, I hope you are familiar with Bloom taxonomy of educational objectives. Let us move the next session concerned with planning for test.

Plan for Classroom Test

In this session, you will learn how to plan a classroom test. You will learn what to consider in the planning stage, how to carry out content survey and to scrutinize the instructional objectives as relevant factors in the development of table of specification/test blue print. Thereafter you will learn how to develop the test blue print, moderate items generated and prepare the items for use.

Dear learner, the development of valid and reliable questions involves proper planning. The plan entails designing a framework that can guide the test developers in the items development process. This is necessary because classroom test is a key factor in the evaluation of learning outcomes. The validity, reliability and usability of such test depend on the care with which the test are planned and prepared.

Planning of test helps to ensure that the test covers the pre-specified instructional objectives and the subject matter (content) under

considerations. Hence, planning classroom test entails identifying the instructional objectives earlier stated, the subject matter (content) covered during the teaching/learning process. This leads to the preparation of table of specification (the test blue print) for the test while bearing in mind the type of test that would be relevant for the purpose of testing. The following serves as guide in planning a classroom test.

i. Determine the purpose of the test;
ii. Describe the instructional objectives and content to be measured.
iii. Determine the relative emphasis to be given to each learning outcome;
iv. Select the most appropriate item formats (essay or objective);
v. Develop the test blue print to guide the test construction (table of specification);
vi. Prepare test items that are relevant to the learning outcomes specified in the test plan;
vii. Decide on the pattern of scoring and the interpretation of result;
viii. Decide on the length and duration of the test, and
ix. Assemble the items into a test, prepare direction and administer the test.

What is a Table of Specification?

A table of specification is two- way chart showing the course of content in one direction and educational objectives in another direction. It describes the topics to be covered by a test and the number of items or points which will be associated with each topic. It is an essentially a blueprint for a test preparation. As the name implies, it specified the content of the test. Its basic purpose is to insure that all intended outcomes, and only the intended outcomes, are measured and that the test includes the appropriate number of items for each measured item. A table of specification is a two-way table with one axis being essentially a content outline and the other axis indicating the behaviours desired with respect to content. The following is an example of table of specification.

Table of specification

Topics/ content	Learning objectives						Total test number item
	Knowledge	Comprehension	application	Analysis	Synthesis	Evaluation	
Topic A	2	3	1			1	7
Topic B	2	2	1			1	6
Topic C	5	3		2	1	1	12
Total number of item	9	8	2	2	1	3	25
Total %	10%	10%	15%	20%	15%	30%	100%

What is the purpose of table of specification?

A table of specification is an important tool that helps teachers design tests that are well representative of the content and objectives for which they are intended to measure. It helps to improve the validity of tests. It also helps to ensure that there is a match between what is taught and what is tested. It also helps teachers to construct tests which focus on the key areas and weights those different areas based on their importance.

Unit Summary

Dear learner, in this unit you learnt that instructional objectives are specific and stated in behavioural terms, specifying what the learners are expected to do at the end of the lesson, using action verbs like discuss, explain, justify, mention and define. In this unit, you have also learnt that bloom classified educational objectives of an intellectual nature or the cognitive domain into six groups, which form a hierarchy of mental skills from the lowest and easiest level, knowledge or memory to the highest and most difficult level, evaluation. You also learnt about affective and psychomotor domains. Teachers should use these domains when construct tests/examinations to assess their students in a particular course or subject. Lastly you have also learnt factors to consider when you plan for the test and how to design a table of specification.

Unit Assignment

1. What is a major difference between objectives and aim?
2. With the use of action verbs discuss levels of cognitive and affective domains.
3. Choose any subject of your interest and then prepare a table of specification.
4. Enumerate what you may consider as guide in planning a classroom test.
5. Why is it necessary to prepare a table of specification or test blue print before writing test items?

Unit 4

Discussing Characteristics of a Good Assessment

Introduction

Dear learner, good assessment tools are characterized by the following features: validity, reliability, objectivity, discriminating power, comprehensiveness, usefulness to students, practicability, fairness, clarity, relevance, simplicity, time. Among of these characteristics are discussed in a more detailed manner below.

Learning Outcomes

Upon completion of this unit you will be able to:

- Explain qualities of good assessment tools;
- examine factors that determine the reliability of test;
- Explain methods of estimating reliability of test;
- Identify types of test validity; and
- Analyse factors affect validity of test.

Characteristics of a Good Assessment

Test Objectivity

Test objectivity means that an individual's score is the same, or essentially the same, regardless of who is doing the scoring. A test is objective when instructor/facilitator opinion, bias, or individual judgment is not a major factor in scoring. Tests may be scored by more than one person, at different or the same time. In other words, objectivity means that if the test is marked by different people, the score will be the same. Marking process should not be affected by the marking person's personality.

Discrimination

A good assessment tools should have a discriminating power. The test should be constructed in such a manner that it will detect or measure small differences in achievement or attainment. This is essential if the test is to be used for ranking students on the basis of individual achievement or for assigning grades. The discriminating

power of a test is increased by concentrating on and improving each individual test item.

Comprehensiveness

A good assessment tool should include items from different areas of materials students learnt. For a test to be comprehensive, it should sample major lesson objectives. It is neither necessary nor practical to test every objective that is taught in a course, but a sufficient number of objectives should be included to provide a valid measure of student achievement in the complete course.

Validity

The most important characteristic of a good examination is validity. The term validity refers to the extent to which a test measures what it is intended to measure and nothing more. A test is considered as valid when it measures what it is supposed to measure. If a test does not hit the intended target, it is termed as invalid test.

Reliability

Reliability refers to the consistency of a measure. A test is considered reliable if we get the same result repeatedly. If the same test is given twice to the same students, it should produce almost the same results; hence we said that that test is reliable.

Simplicity

A good test should be written in a clear, correct and simple language. All questions and instructions should be clear so that to enable students to know exactly what the examiner wants them to do.

Practicability

A good assessment tool should be realistically practical in terms of their cost, time and ease of application. A practical test is easy to administer and to score without wasting too much time or effort.

Fairness

A good assessment tool should be fair to all students. An assessment must accurately reflect the range of expected behaviours as described by the curriculum.

Usefulness to students

Assessment tool should useful to students that it should contribute

to the effectiveness of their learning. It should help students to identify their strengths and weakness by providing them feedback on how they are progressing.

Other characteristics include;

Specification of conditions of administration

A good test must specify the conditions under which the test must be conducted. The conditions must give all students a fair chance to show what their performance. This will improve the reliability of the test. Standardized must come along with specification on the conditions in which students must perform.

Direction for scoring and interpreting test results

Good test must come with direction on how to score and interpret the results of a test. This is especially important for standardized tests, which are used by individuals other than those who developed the test in the first place. Such directions are also important for locally-developed tests, since other individuals may get involved in the process of scoring and interpreting the test.

Generally, a good assessment has both validity and reliability, plus the other quality attributes noted above for a specific context and purpose. In practice, an assessment is rarely totally valid or totally reliable. A ruler which is marked wrongly will always give the same (wrong) measurements. It is very reliable, but not very valid.

Dear learner after discussing the qualities of good assessment tools let us now discuss validity and reliability of test in details.

What Is Test Validity

The most important characteristic of a good examination is validity. The term validity refers to the extent to which a test measures what it is intended to measure. It is vital for a test to be valid in order for the results to be accurately applied and interpreted. Validity isn't determined by a single statistic, but by a body of research that demonstrates the relationship between the test and the behaviour it is intended to measure.

There are three types of validity

a) Content validity, when a test has content validity, the items on the test represent the entire range of possible items the test should cover. Does the content of the test measure stated objectives? Individual test questions may be drawn from a large pool of items that cover a broad range of topics.

b) Criterion-related Validity, a test is said to have criterion-related validity when the test is demonstrated to be effective in predicting criterion or indicators of a construct. There are two different types of criterion validity:

- *Concurrent Validity* occurs when the criterion measures are obtained at the same time as the test scores. This indicates the extent to which the test scores accurately estimate an individual's current state with regards to the criterion. For example, on a test that measures levels of depression, the test would be said to have concurrent validity if it measured the current levels of depression experienced by the test taker.
- *Predictive Validity* refers to how well the test predicts some future behavior. Examples of test with predictive validity are aptitude tests, which are helpful in determining who is likely to succeed or fail in certain subjects or occupations. Aptitude tests predict how well the test will do in some future situation.

c) Construct Validity, this refers to how accurately a given test actually describes an individual in terms of psychological characteristics or traits. It is the degree to which test measures the theoretical construct or trait that it is designed to measure. For example might want to describe a person's reading comprehension, reasoning ability or aptitude creativity or to describe a person as being highly intelligent. All these are hypothetical qualities or constructs that we assume exist in order to explain behaviour. A test has construct validity if it demonstrates an association between the test scores and the prediction of a theoretical trait. Intelligence tests are one example of measurement instruments that should have construct validity.

d) Face- related validity, in face validity, a test is judged to be valid or invalid on the basis of looking at the test critically. In face validity, it is easy to detect when the questions being asked do not relate to the subject under consideration. We need to look the way questions asked, instruction given, format, time allocated, organization of questions, etc.

Factors affecting the validity of the test

Dear learner, validity of test scores can be affected by a number of factors. Among of the factors include: test too short, test too difficult, time allocated to that test, unclear directions that confuse

learners, poorly constructed test items, poor arrangement of items, poor preparation of students for the test, ambiguity in statements and construction, etc.

Factors influencing reliability of a test

The reliability of classroom tests is affected by some factors. These factors can be controlled through adequate care during test construction. Therefore, the knowledge of the factors are necessary to classroom teachers to enable them control them through adequate care during test construction in order to build in more reliability in norm referenced classroom tests. Some of factors that affect reliability of test including the following:

i. Length of Test

The reliability of a test is affected by the length. The longer the length of a test is, the higher its reliability will be. This is because longer test will provide a more adequate sample of the behaviour being measured, and the scores are apt to be less distorted by chance factors such as guessing.

Increase in length of a test brings test scores to depend closer upon the characteristics of the person being measured and more accurate appraisal of the person is obtained. However, we all know that lengthen a test is limited by a number of practical considerations. The considerations are the amount of time available for testing, factors of fatigue and boredom on part of the testees, inability of classroom teachers to constructs more equally good test items. Nevertheless, reliability can be increased as needed by lengthening the test within these limits.

ii. Spread of Scores

The reliability coefficients of a test are directly influenced by the spread of scores in the group tested. The larger the spread of scores is, the higher the estimate of reliability will be if all other factors are kept constant. Larger reliability coefficients result when individuals tend to stay in same relative position in a group from one testing to another. It therefore follows that anything that reduces the possibility of shifting positions in the group also contributes to larger reliability coefficient. This means that greater differences between the scores of individuals reduce the possibility of shifting positions. Hence, errors of measurement have less influence on the relative position of individuals when the differences among group members are large when there is a wide spread of scores.

iii. Difficulty of Test

When norm-referenced test are too easy or too difficult for the group members taking it, it tends to produce scores of low reliability. This is so since both easy and difficult tests will result in a restricted spread of scores. In the case of easy test, the scores are closed together at the top of the scale while for the difficult test; the scores are grouped together at the bottom end of the scale. Thus for both easy and difficult tests, the differences among individuals are small and tend to be unreliable.

iv. Objectivity

This refers to the degree to which equally competent scorers obtain the same results in scoring a test. Objectivity in scoring contributes to reliability. Objective tests easily lend themselves to objectivity because they are usually constructed so that they can be accurately scored by students. In essay testing and various observational procedures where the results of testing depend to a large extent on the person doing the scoring. Sometimes even the same scorer may get different results at different times. This inconsistency in scoring has an adverse effect on the reliability of the measures obtained.

v. Marker's reliability / consistency of the cognitive process of the examiner

The inconsistency on the part of the marker can influence the reliability of the test. Suppose teacher awards one student 36 of out 50 in an English essay test. After a month the same teacher is given the same essay but this time teacher awards the same student 25 out of 50, then 40 out of 50 after another month. These marks vary quite widely. The inconsistency on the part of the marker raises the subject of reliability. How far we can depend on the markers to award consistency marks?

vi. Range of abilities of the group

If the group is too homogeneous such that either most of students pass or fail or are too bunched in the middle, the range of scores will be restricted, a usual small variances, hence low reliability of the test. In other side, the more the heterogeneous the group of students who take the test, the more reliable the measure will be.

vii. Time interval between the two administration tests

The time interval between the two administration tests can affect the reliability of test. The shorter the interval between two

administrations of a test the less likely that changes will occur hence the higher the reliability will be. In other side, the longer the time period between the testing will result in lower reliability due to the greater changes in the students.

viii. Variation with the testing situation/administration

Errors in the testing situation for examples, students misunderstanding or misreading test directions, noise level, distracters and sickness can cause test scores to vary.

ix. Instruction

The reliability of tests results will be affected if the instruction presented to a class tends to overemphasize the teaching points included in the examination. This is often known as "teaching the test" and is undesirable. When the instructor gives students obvious clues as to the test requirements, he not only affects the reliability of the test, but he insults the intelligence of his class.

Unit Summary

Dear learner, in this unit you learnt how to judge the quality of a classroom test/ examination. Specifically, you learnt factors to consider in order to assessing the goodness and badness of examination. In generally good examination/test should be valid and reliable including other factors. It should measure what it intends to measure and test scores should be consistency whenever that test is administered twice to the same group of students.

Unit Assignment

1. Discuss five factors that determine the reliability and valid of test.
2. Differentiate between content validity and criterion related validity
3. All valid tests are reliable tests but not necessary all reliable tests to be valid tests. Discuss.

Unit 5

Examining Report Writing in Educational Evaluation

Introduction

Report writing is an important part in educational evaluation. It communicates results in an educational evaluation. In this unit you will learn different parts of educational evaluation report includes criteria of good report, parts of report and uses of educational reports. Welcome.

Learning Outcomes

Upon completion of this unit you will be able to:

- Explain criteria of good report;
- Plan for a good evaluation report;
- Describe components of evaluation report; and
- Explain uses of educational evaluation reports.

What is Reporting in Educational Evaluation

Dear learner have you ever asked yourself what is a report? Have you written any report in your life? Reporting is the process of communicating results of assessment and evaluation to various audiences who require it such as: students, parents/guardians, teachers, administrators and other institutions. The purpose of reporting is to provide clear information in relation to student academic achievement to a teaching. Reporting involves communicating the summary and interpretation of information about student learning to various audiences.

Criteria of Good Reporting

i. **Relevant and Useful:** a good report should serve a specific purpose/use. It should avoid excessive, unnecessary

reporting.

ii. **Timely:** Any good report should adhere to specific time for its intended use. It is not supposed to be too late or infrequent.it should be done at a favourable or useful time

iii. **Complete: a good report** should provide sufficient information to its audience. It is supposed to have all necessary or appropriate parts.

iv. **Reliable:** A good report is supposed to be trusted, accurate representation of the facts. It should come from a reliable source of information.

v. **Simple and user friendly:** A good report should be appropriate for its audience. Clear language and format and easy to understood.

vi. **Consistent:** A good report should contain units and formats which allow comparison over time, should not contain any logical contradiction. It enable progress to be tracked.

vii. **Cost effective:** A good report should warrant the time and resources devoted to it. It should be able to produce good results without much costing.

Plan for Information Reporting and Utilization

1. **Plan for Reporting**

When writing a report one should consider the following:

Needs/Audience when reporting

	Internal Audience	**External Audience**
Primary Audience	Students, teachers, administrators, parents and personnel who make educational decisions	Government, policy makers, and other institutions
Primary Purpose	Improve students learning and measure effectiveness of teaching and learning activities	Quality assurance issues, Accountability, credibility, solicit funds, celebrates
Frequency	Regularly as per curriculum need	Less often
Content	Comprehensive	**Needs/audience when reporting** Concise, typically abstract, specific.
Format	Determined by curriculum and guidelines	Determined by external requirements

Plan for Reporting: Frequency

i. Data may be collected regularly, but not everything needs to be reported to everyone within the time, *for example*;
ii. Formative assessment/evaluation – daily basis report/weekly or monthly
iii. Summative assessment/evaluation – end of term/semester
iv. School/institution Self evaluation-as per prescription by regulatory authority

- It is critical to identify realistic reporting deadlines; feasible in relation to time, resources and capacity to produce
- Reporting frequency should be timed and based upon the information needs of the intended audience
- Reporting frequency will also be influenced by complexity and cost of data collection

Plan for Reporting: components of Evaluation Report

The components and contents of the report should base on the following:

i. The level and scope of content depends on to whom the report is intended (adhere to the requirements)
ii. Vary from written documents to video presentations
iii. Report Outline (Components may vary as per client need)

Components of Evaluation Report

The report should contain the following major parts

i. Title and preamble pages: includes table of contents, list of figure and acknowledgments
ii. Executive Summary: intended for time-constraint readers but must be attractive to make people curious so that they want to read the entire report.
iii. Introduction; designed to deal with the background of the project, the need for the evaluation, and the entire activity in a nutshell. Explain the purpose of the report, criteria used, participants and other relevant information.

iv Findings: accommodate the results about the efficiency, effectiveness, and impact thereof that have emerged.

v. Conclusion: relate to the evaluation objectives and provide answers to the evaluation questions. It include a discussion of the reasons for successes and failures, especially the constraints and enabling factors.

vi. Recommendation: Provide suggestions, timelines and cost estimates (where relevant) for implementation/ decision making. Should be SMART

vii. References: Provide list documents reviewed (reports, publications, comparison of examination results for previous years)

viii. Annex/ Appendices: Attach Terms of References (TORs) for the evaluation, list of candidates in different levels data collection instruments.

Uses of Evaluation Reports

i. Advocacy and awareness creation

- *example,* contributing to resource mobilization

ii. **For improvement of teaching and learning methods**

Assessment/evaluation reports helps to improve teaching and learning methods

iii. For higher learning entrance

Assessment /evaluation reports are used in placement of students into higher classes. However, more informally they do serve some means of evaluating the school performance with schools being considered acceptable of weak according to their examination result reports

iv. **School classification criteria**

Assessment /evaluation reports help in school classification where school rank is categorized according to number of students, sex, and teacher student ratio. This create competition among schools hence assists in quality achievement

v. **Inform curriculum implementation**

It inform institution management whether curriculum outcomes have been achieved or not

vi. Inform personnel who make educational decisions (diagnostic, instructional, placement, promotion, graduation, curriculum planning and implementation, programme development) about students.

Information Dissemination

- Key mediums of information dissemination (sharing);

a) Print materials distributed through mail or in person
b) Internet communication e.g. emails, websites, blog, etc.
c) Radio communication
d) Television and filmed presentations
e) Telephone communication includes text messaging and calls
f) Live presentations such as meetings and conferences

- Selection of reporting medium should be guided by what is most efficient in time and resources and suitable for the audience.
- Decision Making
- Form the heart of evaluation report utilization
- No matter how well the report is prepared or disseminated, it will ultimately be up to the user to decide when and how to use.
- Key considerations that can aid the use of information in decision making;
    a) Stakeholder dialogue: discussion and feedback via meetings, seminars & workshops, web-based forums, teleconferences or follow –up procedures
    b) Management response

Unit Summary

Dear learner, in this unit you explore on how to write a good educational report, criteria for a good report, uses of evaluation reports as well as dissemination of reports. I hope you enjoyed to study this module which was very practical.

Unit Assignment

1. Explain criteria of good report.
2. Plan for a good evaluation report.
3. Describe components of evaluation report.
4. Explain uses of educational evaluation reports.

References

Gronlund, N. E. (1985). Measurement and Evaluation in Teaching. New York:Macmillan Publishing Company.

Kerlinger, F. N. (1973). Foundations of Behavioural Research. Holt, Rinehart and Winston Inc. New York:

Ogunniyi, M. B. (1984). Educational Measurement and Evaluation.

Thorndike, R. L. and Hagen, E. P. (1977). Measurement and Evaluation in Psychology and Education. New York: John Wiley and Sons; New York.