

ADVANCED DIPLOMA IN GUIDANCE AND COUNSELLING (DGC)

DGC 206: PSYCHOLOGICAL TESTING AND ASSESSMENT

UNIT 1: THE NATURE OF PSYCHOLOGICAL TESTING

INTRODUCTION

A test is an instrument used to determine the extent of presence of a particular trait in someone or group. This trait could be in the form of interest, intelligence, aptitude, knowledge or skill or ability. It may not be possible to determine how much of a trait is present in someone without presenting a task which the person (testee) will solve or tackle. In modern times, tests have been employed in a number of places outside the school setting to obtain information about persons. In the hospitals, psychological tests are employed to help in diagnosis and prognosis. In industries and the banking sector, tests are employed in the recruitment, placement and promotion of staff members. All these are pointers to the fact that tests are of great importance in almost all areas of human endeavour. In this unit, you will be introduced to the nature and uses of psychological tests.

OBJECTIVES

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

- i) explain the meaning of psychological tests;
- ii) state at least three uses of psychological tests; and
- iii) describe two types of psychological tests used by counsellors.

HOW TO STUDY THE UNIT

1. Skim through the whole unit to have an idea of what it is all about
2. Go back to the beginning of the unit and read through carefully by paying attention to every detail.
3. Work on the assignment given

4. Attempt to answer all questions asked without consulting the text.

WORD STUDY

Psychology: the scientific study of human behaviour

Testing: the presentation of task or series of tasks to a respondent (called testees) to which he or she is expected to react in order that the extent of presence of a trait or ability could be determined.

Validation: This is the process of ascertaining the qualities of a test. In other words, it is a way of ensuring that tests are of moderate level of difficulty and discrimination.

MEANING OF PSYCHOLOGICAL TESTS

As a student or classroom teacher, you must have, at one time or the other, taken or administered series of tests in a classroom situation. These tests were used to discover the best among equals. Across nations, continents and over time, tests have been used in schools to help in determining the level of achievement of individuals in school subjects. Thus, achievement tests have become indispensable instruments in the school system.

However, there are other tests which do not specifically measure achievement in the classroom. They are called psychological tests. Psychological tests are series of instruments containing statements or tasks (sometimes in form of questions) seeking information that could help in the determination of how much of a particular trait, attribute or quality is present in a person. The use of psychological tests dated back to the time of old Chinese Empire (Dubois, 1966 in Anastasi, 1992) when series of tests were developed for selection into the civil service. Today, in many occupations and industries, psychological tests are used in the process of staff recruitment and/or selection.

ACTIVITY I

1. In your own words, give a concise definition of psychological tests.

USES OF PSYCHOLOGICAL TESTS

Psychological tests are used for the following purposes:

- ✓ Selection and Placement of candidates in Institutions and work places.
- ✓ Prediction of behaviour and task performance.
- ✓ Discover inadequacies in curriculum.

Understanding of one's personality

Psychological tests that could be employed to investigate personality types have been developed. Series of these tests could help in the categorization of the 'brilliant' and 'less brilliant', amiable, sociable, extroverts, introverts, phlegmatic, melancholic and sanguine. Any individual who attempts any of the standardized personality tests will understand himself or herself.

Determination of Individual Differences

No two individuals possess exactly the same personality or psychological traits. Two identical twins will have some areas of differences. These could be in the area of temperament or academic achievement. Remember a saying that fingers are not equal. In the same manner, differences exist from one individual to the other. These differences are revealed when an individual or group of individuals respond to a psychological test,

such as one measuring intelligence, aptitude or interest. The results of psychological tests provide information that enables teachers, counsellors and psychologists to understand differences among their students, clients and subjects.

Classification and Guidance

Psychological tests are useful in determining who should or should not be in a class or group. Also, counsellors use the outcome of psychological tests to advise students on the subjects they are expected to offer at the Senior Secondary School. Psychological tests could help the counsellor to determine the appropriate therapy to be employed in helping a client overcome his or her problems.

Measurement of Differences in Attitudes

Two individuals may not necessarily share the same opinion about an issue. Attitude of two good friends to the study of Mathematics may not be the same but this could not be ascertained until the administration of a test measuring the attitude to Mathematics.

Psychological Tests used by Counsellors

There are series of psychological tests employed by counsellors in the discharge of their duties to clients. These tests have been validated and their psychometric properties ensured before they were published. Depending on affordability/accessibility factor(s), counsellors make use of tests measuring scholastic aptitude, intelligence, achievement, vocational interest, personality variables, sociometry and socio-economic status.

ACTIVITY II

1. List and explain four uses to which psychological test could be put.

SUMMARY

- In this unit, you have been introduced to psychological tests which we described as series of instruments containing statements or tasks seeking information that could help in the determination of the extent to which a particular trait, attribute or quality is present in an individual. Examples of psychological tests include tests measuring intelligence, scholastic aptitude, achievement, vocational interest, personality variables, sociometry and socio-economic status.
- Psychological tests are used for the measurement of differences in attitudes of people; for classification and guidance; to determine individual differences and for the measurement and understanding of people's personality.

ASSIGNMENT

1. In not more than one page, trace the history of the development of psychological tests.

REFERENCES

Anastasi, A. (1982). **Psychological Testing**. New York: Macmillan Publishing Co. Inc.

Hopkins, K.D. (1998). **Educational and Psychological Measurement and Evaluation**.
Boston: Allyn & Bacon.

UNIT 2 ACHIEVEMENT TESTING

INTRODUCTION

The importance of tests in the school system cannot be over-emphasized. When a teacher has taught a topic or course to his or her pupils, it is always necessary to ascertain if the students have understood the subject matter. Achievement tests are a useful tool in the hand of the classroom teacher to ascertain whether or not the students have understood the subject matter. In this unit, you will learn the meaning of achievement tests, purpose, types and methods of classification.

OBJECTIVES

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

- i) define achievement tests in your own words;
- ii) explain the purposes of an achievement test; and
- iii) identify the similarities and differences between achievement and aptitude tests.

HOW TO STUDY THE UNIT

1. Read through this unit and ensure that you understand what is meant by achievement tests.
2. Make sure you understand other key concepts and terminologies used in the unit.
3. Use your dictionary to check the meaning of difficult expressions in the text.
1. Attempt all the activities and assignments included in the unit without referring to the text.

WORD STUDY

Achievement: The act of accomplishing or finishing a task, or an assignment successfully especially by means of studying, exertion, skills and perseverance.

Aptitude A natural or acquired disposition or capacity for a particular purpose, or tendency to a particular action.

Standardization: The act of checking or adjusting (by comparison with a standard) the accuracy of a test or a measuring instrument.

MEANING OF ACHIEVEMENT TESTS

Achievement tests are tests designed to find out the level of ability of a person in a given instructional task. It is common to give tests to students at the end of a school term or session. These tests are meant to ascertain the extent to which the candidate is proficient

in the instructions previously delivered to him. Even the apprentice motor-mechanic is subjected to testing situation when he is asked by his master to replace a bad brake-pad without help. Therefore, achievement tests are necessary in our daily life.

Achievement testing involves the presentation of tasks based on an exposure (in the present or past) to a content area. The main assumption of achievement testing is that testees have previously been exposed to instruction/teaching. The test is only a means to ascertaining the extent to which the body of knowledge or instruction was understood. Achievement tests reflect what a student has learned from the instructions given during classes, practicals and out-of-class sessions. An achievement test in Mathematics is expected to produce results that are assumed to be reflective of what the learners have learned from the content (and skills) taught them in their Mathematics class.

Achievement tests may also portray the future ability of the testees. A student with a first-class grade in the university degree examinations is considered to possess strong ability for postgraduate studies and research. It does not stop there, such a graduate is considered to be of high promise in other new areas with little training and supervision. Also a student who scores high marks in Junior Secondary School Integrated Science is adjudged to have the potential to offer science subjects in Senior Secondary Class.

ACTIVITY I

1. In your own words, define achievement tests.
2. Give reasons why achievement tests may be considered futuristic in orientation.

ACHIEVEMENT VERSUS APTITUDE TESTS

Achievement tests and aptitude tests have one thing in common: they provide information on the testee's developed ability. The difference, however, is in the aspect of time. While achievement tests reflect current developed ability, aptitude tests reflect predicted ability (i.e. ability that is likely to be demonstrated at a later time). It could also be said that achievement tests rely on prior and present learning experiences but aptitude tests forecast learning or benefit/performance in an unknown situation in the future. Although both achievement and aptitude tests measure ability, the difference comes in the aspect of the test content. While achievement tests are content-specific, aptitude tests are construct-specific i.e. the items focus largely on psychological constructs such as scholastic aptitude, manual dexterity, etc. Also, items in achievement tests are drawn from previously-specified content areas which have or ought to have been taught.

You should however bear in mind that the difference between the two test types cannot be strictly established. Achievement tests can be employed to 'predict' future abilities and skills or benefits from a programme. In the same way, aptitude tests could also serve as a

reflection of prior learning and/or developed skills. The fact that all psychological tests measure current behaviour is enough to justify their similarity.

ACTIVITY II

1. State two major ways in which achievement tests are different from aptitude tests.

USES OF ACHIEVEMENT TESTS

Achievement tests aid learning. In the school system, the awareness of an impending test makes even the lazy student to read in order to pass. The use of tests in schools helps to improve the rate of learning of course/subject contents and consequently improve performance. For instance, in schools where Continuous Assessment (CA) tests are conducted regularly, there is the tendency for the students to perform better in their final examinations than students from schools where Continuous Assessment tests are not conducted on a regular basis.

Achievement tests are also used for classification of students/learners. Test results provide the basis for classifying students as high/low achievers, superior/inferior ability students, etc. In Nigeria, one popular standardized achievement test is the Senior Secondary Certificate Examination. Without the writing of tests, it would be difficult (and questionable) to classify a student as an “A” candidate and another as “F” candidate.

Closely related to this use is students’ placement. Achievement tests (especially in the secondary schools) are used to determine which students should be in which class. For example, the outcome of students’ entrance examination results (such as the National Common Entrance Examination) determines who secures placement in Federal Government Secondary Schools. In the same vein, the end-of-session examination results (developed by teachers) are conducted to determine who should be promoted to the next class.

Achievement tests facilitate teaching and learning. In a typical Secondary School in Nigeria, the Mock Senior School Certificate Examination (SSCE) is administered on students preparing for the Senior School Certificate Examination. This examination (though achievement by nature) exposes the weaknesses of individual candidates. The teacher then examines more closely those aspects of the subjects where each candidate (or the entire group) needs remediation. The terminal examinations conducted in schools are also directed at revealing to each student, teachers and parents those areas where candidates are doing well and the areas of their deficiency. This helps the subject-teachers in reviewing/appraising of their teaching strategy, objectives and methods. This ultimately leads to a general improvement in the teacher-learning process.

ACTIVITY III

1. List and explain four uses of achievement tests.

TYPES OF ACHIEVEMENT TESTS

Achievement tests can be divided into two main categories:

- (a) Teacher-made Achievement Tests
- (b) Standardized Achievement Tests.

Teacher-made Achievement Tests

Teacher-made achievement tests are those tests constructed by the classroom teachers, or better still, subject teachers for the assessment of their students. They are tests designed for students that have been taught the content of a subject or course. The items could be in essay or objective format. They are very common as teachers use them across schools and are more relevant to the objectives of the teachers and their learners. Teacher-made tests help the teacher in identifying students' weaknesses in order to plan for remediation and motivate them to develop positive study habits.

Teacher-made achievement tests are the most used in schools. They are employed for more frequent evaluations that the teacher has to make. Teacher-made achievement tests sample only content areas that have been taught by the teacher. They are usually designed for measuring the level of achievement of the classroom objectives set by the teacher and no established norm is employed in score interpretation.

Standardized Achievement Tests

Standardized achievement tests are those constructed by test expert/specialists and they are administered, scored and interpreted under standard conditions. They are commercially-prepared tests usually administered on a large sample of a population usually by examining agencies such as the West African Examinations Council. Items on standardized achievement tests are of high technical quality as they would have been trial-tested on a sample and the psychometric properties ensured (through item analysis) before being presented to actual candidates.

Standardized achievement tests cover larger content areas of a subject when compared to teacher-made achievement tests. This may not necessarily mean that standardized achievement tests would have more items than teacher-made achievement tests. Standardized achievement tests have precise directions for administration and scoring. The processing of candidates' scores and subsequent award of grades follow a standard procedure which allows for the control of errors and omissions.

Another feature of standardized achievement tests is that alternative or equivalent forms are usually provided, so also is information relating to the extent of their equivalence. This is usually done as a precaution against leakage or any other problem that may necessitate the cancellation of the examination or non-administration of the examination. In such a situation, new date will be announced to candidates when an 'equivalent' version of the paper will be taken.

Information about norms are provided for standardized achievement tests. Norm refers to average performance (of the standardization sample) regarded as normal and upon which the performance of individuals could be compared. Norms are established in order to have bases for judging the performance of an individual in a subject or across subjects.

Another unique feature of standardized achievement test is that candidates are notified in advance of the date and venue of the test or examination. Unlike teacher-made achievement tests that could be administered at any time with or without prior notice to the testees.

The results of standardized achievement tests are published and sent to each candidate. If any of the candidates is not satisfied with his or her result, the candidate could lodge a protest and the examining body or agency will revisit such a case based on a pre-specified procedure.

ACTIVITY IV

1. Give two (2) differences between standardized and teacher-made achievement tests
2. List four agencies that conduct standardized achievement tests in Nigeria including the name of the examinations they conduct.

CLASSIFICATION OF STANDARDIZED ACHIEVEMENT TESTS

Standardized achievement tests could be grouped into three. These are: (i) diagnostic tests (ii) single subject-matter achievement test, and (iii) survey batteries.

Diagnostic tests are tests used in finding out the areas of weakness or strength of individuals. Diagnostic tests are designed to help the teacher, psychologist or counsellor in the identification of the particular aspect of a skill (or subject) where the candidate needs help or remediation. Diagnostic tests cover wide content areas and they contain several parts of the subject or skills being tested. For example, a diagnostic test in Mathematics usually contains several items on each of the operations being tested. Diagnostic tests provide several items on specific topics or skills in a subject. This gives room for the testee to attempt many of them and in the process, the areas of deficiency, and/or strength will be known. The quality of a diagnostic test depends on the quality of

its items in terms of analyzing knowledge of concepts into various sub-units. It also depends upon the ability to develop valid items that will measure the construct of interest.

In using standardized diagnostic tests, counsellors/teachers should ensure that the responses are objectively scored and carefully analysed. The counsellor's interest is not in the total score (as the case with single-subject matter tests) but in the specific areas of difficulty. Diagnostic tests are administered on candidates with difficulties only. Most standardized diagnostic tests are on Reading and Arithmetic. Examples include Stanford Diagnostic Reading Test (SDRT) and Stanford Diagnostic Mathematics Test (SDMT).

Single-subject-matter tests are used to measure the level of achievement in a particular content area. Single-subject-matter tests cover a particular subject matter with many items. Unlike standardized achievement tests, it contains items from many diverse areas such as reading skill, computation skill, manual dexterity, etc.

Survey test batteries contain items intended to measure candidates' knowledge, skills and developed abilities. The items on this type of test cover wide content areas and the measurement of course objectives is one of its target. Survey test does not contain many items on a single content area; rather, it samples representative units of the subject-matter covered.

ACTIVITY V

1. Identify the three classes of standardized achievement tests.
2. In what ways are diagnostic tests different from single subject-matter achievement tests?

SUMMARY

- In this unit, achievement testing was considered. We have described achievement tests as those designed to determine the level of ability of learners in a given instructional task. While achievement test is present and past-oriented and content specific, aptitude test is future-oriented as it predicts ability that is likely to be demonstrated at a later time.
- Achievement tests aid teaching and learning; facilitates classification of students and help in taking placement decisions. Achievement tests could be teacher-made or standardized. The format for presenting achievement test could be essay or objective. Standardized achievement tests are classified into three: standardized diagnostic test, single-subject-matter achievement tests, and survey batteries.

ASSIGNMENT

1. Identify and discuss the features of the Nigeria Senior School Certificate Examination which in your opinion makes it a standardized examination.

REFERENCES

Hopkins, K.D. (1998). *Educational and Psychological Measurement and Evaluation*. Boston: Allyn & Bacon.

Mehrens, W.A. (1975). *Measurement and Evaluation in Education and Psychology*. New York: Holt, Rinehart and Winston.

Ughamadu, K.A., Onwuegbu, O.C. & Osunde, A.U. (1991). *Measurement and Evaluation in Education*. Benin City: World of Books Publishers.

UNIT 3 THE NATURE OF INTELLIGENCE

INTRODUCTION

In studying this unit, reflect on your secondary school days. Think of your classmates then. You will remember at least one person who had the requisite ability to perform well in school subjects; interact well with classmates; adapt well to new situations, demonstrate high reasoning ability and wisdom in speaking. Such an individual could simply be described as intelligent. It is intelligence that determines our abilities and readiness for learning and adjustment. In this unit, you will learn a lot about intelligence and how it is measured.

OBJECTIVES

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

- i) define the word “intelligence”;
- ii) explain what is meant by academic intelligence;
- iii) mention and explain two most-popular intelligence test measures.

HOW TO STUDY THE UNIT

- (1) Read through each of the sections carefully
- (2) Note the salient points in your notes and master them
- (3) Do the assignments and tasks listed under each of the activities.

WORD STUDY

Intelligence: The total capacity to think critically, reason and adjust to new situations in the environment.

Scale: A measuring instrument used to obtain information about an individual.

Gifted Children: Children whose level of intelligence are regarded as more than average for their age group.

THE NATURE OF INTELLIGENCE

Intelligence has been defined as “the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with the environment (Wechsler, 1958). The totality of an individual’s ability to learn, benefit from new experience and behave in a manner considered acceptable and normal are all embedded in the psychological construct called intelligence. It should be noted that the concept of intelligence is a complex one. Scholars and psychologists define it in different ways and

there appears to be a theoretical dispute among psychologists on the operational definition of intelligence. However, there is a consensus that intelligence is “that which an intelligence test measures”.

You will realize that no two individuals (identical twins inclusive) are the same in their abilities to learn a new concept or demonstrate mastery of a previously learned concept. The level of competence which people demonstrate in learning, and the ease with which people adjust to changing situations and environments are all dependent on their intelligence.

Intelligence testing is the term used to describe the totality of the processes involved in the administration of intelligence tests and subsequent measurement of intelligence. Thus, intelligence tests are tests containing series of tasks aimed at measuring the general intellectual functions in an individual.

ACTIVITY I

1. Identify and discuss three characteristics or attributes of a very intelligent child.

MEASURING INTELLIGENCE

The measurement of intelligence is a process that involves the administration of intelligence tests to an individual or a group, scoring of such test response, and administration of the score. For many years now, intelligence has been expressed in IQ (meaning Intelligence Quotient). Intelligence Quotient is a quantitative description/expression of a person’s or group of people’s level of intelligence based on established norm. The expression of intelligence using IQ began with the development of Stanford-Binet Intelligence Scale. The level of intelligence expressed in IQ is determined by finding the ratio of Mental Age (MA) to Chronological Age:

$$IQ = \frac{MA}{CA} \times 100$$

For example, if a student who is 15 years old (Chronological Age) attempts an intelligence test and obtains 20 marks, a score which is the average (norm score) for 18-year olds, it therefore means that the student’s mental age is 18 years although his actual (i.e. chronological) age is 15 years.

EARLY WORKS OF ALFRED BINET (1857 – 1911)

Alfred Binet was regarded as the father of intelligence testing. As early as 1896, Binet had published a number of proposals for measuring the intelligence of young ones

between 3 and 11 years old. In 1904, Binet and another psychologist, Theodore Simon, were appointed by the French Government to come up with a strategy for identifying and educating the mentally retarded children. Thus, Binet and Simon came up with a scale known as Binet-Simon Scale in 1905. According to Binet, components of intelligence include judgement, comprehension and reasoning. Therefore, a child's intelligence will be determined by his scores on the cognitive tasks requiring the ability to make sound judgement, recognize familiar objects and reason critically. The scale usually called 1905 scale contained 30 tasks arranged in ascending order of difficulty.

The Binet-Simon Scale was revised in 1908 and 1911, leading to series of amendments to some of the items on the scale. The 1908 version was based on the performance of 300 normal children of between 3 and 11 years old. In 1916, a psychologist in Stanford University revised and adapted for use the Binet-Simon Scale.

THE STANFORD-BINET SCALE

As mentioned in the previous section, Lewis M. Terman of the Stanford University, U.S.A. published an amended version of the Binet-Simon Scale. The new scale known as Stanford-Binet Scale was published in 1916 with a manual titled "The Measurement of Intelligence". The scale provides an overall Intelligence Quotient score for individuals. The scale was later revised in 1937, 1960, 1972 and 1985. The Stanford-Binet Scale (1985 edition) consists of four sections – verbal reasoning, quantitative reasoning, visual/abstract reasoning and short-term memory. The Stanford-Binet Scale is individualistic in orientation. Series of studies have confirmed its high validity and reliability.

THE WECHSLER INTELLIGENCE TEST

David Wechsler in 1939 came up with an intelligence scale known as Wechsler-Bellevue Intelligence Scale for adults. The reason for calling it an adult scale was based on the facts that the existing intelligence scale was designed for children. Thus, the Wechsler Scale consists of an extension to performance testing from verbal testing on which the Binet Scale was based. It was revised as Wechsler Adult Intelligence Scale (WAIS) in 1955 and Wechsler Adult Intelligence Scale-Revised (WAIS-R) in 1981. In order that the scale could be used to measure children's intelligence, it was adapted for use on children in 1949. It was named Wechsler Intelligence Scale for Children (WISC). This was revised in 1974 (as WISC-R) and in 1991 (as WISC-III) for use on children between the ages of 6 and 16 years.

In measuring the intelligence of children between 3 and 7 years old, the Wechsler Scale was reviewed and published as Wechsler Preschool and Primary Scale of Intelligence (WPPSI). It therefore means that the Wechsler Scale is in three formats:

- (i) Wechsler Preschool and Primary Scale of Intelligence (WPPSI) for ages 3 – 7 years
- (ii) Wechsler Intelligence Scale for Children (WISC) for ages 6 – 16 years

- (iii) Wechsler Adult Intelligence Scale (WAIS) for 17 years and above.

The three forms of the Wechsler Scale have uniform sections and sub-sections. All the versions have adequate degree of validity and reliability. The WISC-III is perhaps the most frequently used intelligence test for children. Both the Stanford-Binet Scale and the Wechsler Scale are good measures of academic/scholastic aptitude.

ACTIVITY II

1. Mention three psychologists and the names of intelligence tests developed by them.
2. Differentiate between a child's chronological age and mental age.

OTHER MEASURES OF INTELLIGENCE

Apart from the Stanford-Binet and Wechsler scales, there are other measures of intelligence. One of them is the Kaufman-Assessment Battery for Children (K-ABC). This is a measure that is being touted as the successor to Stanford-Binet and Wechsler scales. Though the K-ABC has high reliability and validity scores, psychometrists feel that it is not representative of the normative sample used in the validation process.

There is another measure of general intelligence through multiple aptitude testing. This method is employed to obtain more detailed information about the cognitive, affective and psychomotor skills of individuals. One good example of such scale is the Differential Aptitude Scale (DAT) published by the Psychological Corporation, the Science Research Associates' Mental Abilities, and the Armed Services Vocational Aptitude Batteries (ASVAB) of the U.S. Military Enlistment Processing Command. Several other scales of individual and group intelligence measures are available in Anastasi (1982) and Hopkins (1998).

ACTIVITY III

1. Differentiate between Stanford-Binet and Wechsler Scales.

USING TESTS TO IDENTIFY GIFTED STUDENTS

Gifted children are those with considerable skills and abilities considered as more than average for their age group. These children show very high academic promise and will not find concepts considered commensurate with the readiness level of their peers interesting or challenging. Such children possess abundant talents and they usually constitute the top two to three percent of their group. They profit more in specially-

designed educational programmes that will allow them to develop or learn at their own rate. Subjecting children identified as gifted to the usual sequenced pattern of instruction as we have in conventional schools will amount to retarding their progress. This is the reason why a liberalized system that will allow each gifted child to progress at his or her own rate is desirable.

In the United States of America, the John Hopkins University pioneered the effort in identifying and educating the gifted children in 1977. In Nigeria, the National Examinations Council (NECO) conducts the Gifted Children Examinations every year to identify gifted children who would be admitted into the gifted children's school known as the Suleja Academy in Niger State. In most cases, one-shot examinations are those designed for identifying those to be admitted into the gifted education programme. Such examinations have the following limitations:

- (i) One-shot examinations may not be a valid reflection of the true abilities of exceptional children
- (ii) One-shot examinations for the identification of gifted children neglect the writing aspect of the usual 3r's emphasized in elementary education. Not only this, the affective (and even psychomotor) aspect of cognitive skills are left un-sampled by most gifted tests as they are usually presented in objective formats.
- (iii) Arising from the format of presentation of most gifted tests is the problem of validity of candidates' scores. This is due to the fact that such scores could emanate from guessing and chance factors.

However, most examinations designed to identify gifted children have been found to be valid and reliable although the reliability of most intelligent tests are not only due to hereditary factors but also due to environmental conditions. Environmental factors are usually responsible for changes in a child's score on different occasions in intelligence tests.

ACTIVITY IV

1. List and briefly discuss the problems associated with the procedures for identifying gifted children in Nigeria.

SUMMARY

- In this unit, you have been taught the meaning of intelligence and how intelligence is measured. Specifically, we have defined intelligence as that which almost translates to academic/scholastic aptitude. Intelligence is the construct which intelligence tests measure and it is a reflection of the reasoning, learning and adaptive abilities of an individual.

- The most popular scales for measuring intelligence are the Stanford-Binet Scale and Wechsler Scales. Intelligence is generally expressed in term of I.Q. (Intelligence Quotient). We have also shown that scores from intelligence tests are not stable because of changes in the testee's environmental conditions, which enhance or impair the normal functioning of his or her intellectual abilities.

REFERENCES

Anastasi, A. (1982). **Psychological Testing**. New York: Macmillan Publishing Co. Inc.

Hopkins, K.D. (1998). **Educational and Psychological Measurement and Evaluation**. Boston: Allyn & Bacon.

UNIT 4: MEASURING AFFECTIVE OR PERSONALITY VARIABLES

INTRODUCTION

In the preceding units of this module, we have discussed the development and administration of various types of tests such as achievement, aptitude and intelligence tests. In this unit you will be introduced to the measurement of affective or personality variables. Affective variables include those things that a person usually does or will do under natural circumstances. These include someone's temperament, acceptability in a group, attitude towards peer, attitude towards colleagues, etc.

OBJECTIVES

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

- i) explain affective measures;
- ii) list and define common personality variables;
- iii) list and describe the various methods by which affective variables are measured.

HOW TO STUDY THIS UNIT

1. Read through the unit and ensure that you understand the meaning of the key concepts in test construction.
2. Study the new words under the section on Word Study.
3. Use your dictionary and textbook to check the meanings of words which you find difficult in the text.
4. Attempt all the activities and assignments included in the unit without referring to the text.

WORD STUDY

Measurement: The assigning of number, grade or sign, to express in quantitative term the extent of presence of a particular trait.

Variable: Combination of attributes used in describing a person (e.g. honest, dishonest, aggressive, intelligence) or group of people. Thus, male and female are attributes of sex (a variable); occupation is a variable that has attributes like mining, engineering, law, etc as variables.

NATURE OF AFFECTIVE VARIABLES

This is discussed in detailed in Unit 6. You will also recollect that Assessment of the cognitive domain of behaviour has to do with the measurement of developed ability of an individual resulting from classroom instruction. However, in the affective domain, emphasis is on a person's character, values and feelings. Affective variables are those which describe a person's emotions, feelings, interest and attitudes in relation to the feelings and perception of other people. Variables of interest include punctuality, respect for elders, friendliness, neatness, attitude to colleagues, attitude to school work, attendance, self-concept, etc. All these are affective variables, which are usually sparingly represented in the teacher's assessment activities in the classroom.

The assessment of affective variables is a bit more difficult than those in the cognitive domain. This is simply because affective variables are abstract concepts which may require extra painstaking effort on the part of the teacher, psychologist, or counsellor. Notwithstanding the low level of employment of assessment tools in the affective domain in the educational system, the fact remains that nearly all cognitive objectives have an affective component. For example, before a student could learn, he has to be mature and ready for such learning. Motivation and readiness are affective variables that are capable of dictating the extent of cognitive achievement. They are therefore very important.

The taxonomy of affective educational objectives was developed by Krathwohl (1964). The taxonomy proposed 'degree of internalization', the common factor to all the hierarchies of the domain. The hierarchy of the affective domain ranges from the awareness of an affective variable to the characterization of such variable.

ACTIVITY I

1. Describe in your own words the affective domain of behaviour.

MEANING OF PERSONALITY

The term personality has numerous definitions in research literature. Some of the ways in which psychologists have defined the term include the following:

- The totality of qualities and traits, as of character or behavior, that are peculiar to a specific person.
- the habitual pattern which makes a person's ways of doing things immediately recognised by other people
- the characteristic way of thinking and acting which identifies each person as a unique individual.
- the culmination of all the relatively enduring dimensions of individual differences on which people can be measured.

- the dynamic organisation within the individual of those psychological systems that determine his characteristic behaviour and thoughts

COMMON PERSONALITY CHARACTERISTICS

Personality characteristics are dimensions of personality in which individual differences in people can be observed and measured. The following are a description of common personality characteristics in psychological literature:

Self- Concept

This refers to the overall perception of an individual about himself, his abilities, disabilities, strengths and weaknesses. An individual can be said to have a positive or negative, low or high, or favourable or unfavourable self-concept depending on his or her score on a self-concept scale.

Psychologists assert that the self-concept accounts for many personality differences since its content and organisation cause man to process information in a relatively stable manner, which results in relatively stable individual differences in behaviour (Mayer & Sutton, 1996).

Most personality theorists share the view that every human experience can be evaluated in terms of the self-concept and that most human behaviours can be understood as an attempt to maintain consistency between the self-concept and human actions. However, it has been realised that most of the time, the consistency between the self-concept and actual behaviour is not always achieved (Gross, 1992). Likewise, a person's self-concept may differ quite radically from how others perceive him. For instance, a person may be highly successful and respected by others and yet regard himself as a failure.

The self-concept is usually measured by the use of rating scales. Most of the currently published instruments measuring self-concept are of this type. Rating scales typically are composed of a set of statements to which the respondent expresses a degree of agreement or disagreement. The five-point Likert scale is common. Typical items on a self-concept scale might include "I am good at Physics" or "On the whole, I am satisfied with myself." Responses are then summed to form a score for a specific scale (e.g., Physics self-concept) or a measure of global self-concept. Some popular self-concept scales include the Tennessee Self-Concept Scale, Piers-Harris Children Self-Concept Scale, Puerto-Riccan Self-Concept Scale.

Extraversion

Introversion-extraversion describes an individual's general orientation to sociability and sensation seeking versus reserved and careful behaviour patterns. This personality characteristic produces two types of individuals known as introverts and extroverts. The 'typical' introvert is a quiet, retiring sort of person, introspective, fond of books rather than people. He is reserved and distant except to intimate friends. He tends to plan

ahead, 'looks before he leaps' and distrusts the impulse of the moment. The introvert is someone who does not like excitement, takes matters of everyday life with proper seriousness and does not lose his temper easily. He is reliable, somewhat pessimistic, and places great importance on ethical standards.

On the other hand, the 'typical' extrovert is one who is sociable, likes parties, has many friends, needs to have people to talk to, and does not like reading or studying by himself. He craves for excitement, takes chances, often sticks out his neck, acts on the spur of the moment and is generally an impulsive individual. The Eysenck Personality Inventory is one of the most popular instruments used in measuring the introversion-extroversion tendencies.

Locus of Control

This is a personality characteristic which indicates the degree to which an individual perceives events as being contingent upon personal initiative or upon factors outside the individual's control. The construct refers to an individual's perception of the locus of events as determined internally by his/her own behavior vs. fate, luck, or external circumstances. Individuals can be classified into either an internal or external locus of control orientation. If a student with an internal locus of control does badly on a test, he or she is likely to blame either her own lack of ability or preparation for the test. Conversely, another student with an external locus of control will tend to justify the poor performance by saying that the test was too hard or that the teacher graded unfairly.

The concept of locus of control was developed by psychologist Julian Rotter, who devised the Internal-External Locus of Control Scale (I-E) to assess this dimension of personality. Various studies have found that the Rotter's Scale is a valid predictor of behavior typically associated with locus of control.

Achievement Motivation

Achievement motivation or the need for achievement, as it is popularly known, has been identified as an important dimension of personality that is basic to the determination of an individual's overall personality. This need is defined as the desire to accomplish something difficult; to master, manipulate, or organise physical objects, human beings, or ideas; to do this as rapidly and as independently as possible; to overcome obstacles and attain a high standard; to excel one's self; to rival and surpass others (Murray, 1938).

The concept of need achievement arose from consistent observations that individuals differ in their responses to success situations and situations that could possibly lead to failure. The former disposition is termed achievement need (nAch) and the latter the fear of failure. These two personality factors are assumed to be relatively stable and persistent and their relative strength varies from person to person.

In the achievement motivation model, McClelland (1980) described the characteristics of people with high and low need for achievement. One of the prominent features of high-need achievers is that they are only moderate risk-takers. They want to succeed, but they

are also highly motivated to avoid failure. According to McClelland, such people are very energetic in their approach to work. They usually get 'pumped up' for tasks with the potential for some personal achievement. They are self-confident, willingly take on responsibility and are driven less by the desire for fame than by the need to live up to a high self-imposed standard of performance. Such people prefer jobs that give them personal responsibility for outcomes. Routine and boring jobs hold no interest for them.

Conversely, individuals with low need for achievement are attracted by either very easy or very difficult tasks. They exhibit anxiety or fear over failure in the face of challenging tasks. They are comfortable working in situations that will not ascribe to them personal responsibility for the success or failure of job outcomes (Burger, 1990).

Impulsiveness

Impulsiveness or impulsivity is the irresistible tendency to accomplish an act without first thinking about its consequences or relevance. It is an immediate response to thoughts or deeds without any consideration of their appropriateness or consequences. People who are impulsive tend to speak, act or move without thinking. Typical features of impulsiveness may include rapid decisions without reflecting on the consequences; acting without thinking; speaking without any thoughts about the content and the appropriateness of the situation and inability to wait for one's turn in games or group situation.

Many psychologists have argued that impulsiveness could sometimes be a positive personality trait. The ability to make rapid decisions or to act quickly can be important for leaders or become necessary in a case of emergency. The rating scale is commonly used to measure a person's level of impulsiveness.

ACTIVITY II

1. Define the term personality.
2. Mention two attributes of people with each of the following personality characteristics: positive self-concept, extraversion, impulsiveness and internal locus of control.

MEASUREMENT OF AFFECTIVE VARIABLES

There are various ways of measuring affective variables. Some of these are discussed in this section.

Questionnaires

Questionnaires are instruments used in collecting data from a survey. They usually contain questions and directive for response. In most cases, the response style usually

employed is either the Likert style or rating scale. In some other instances, the true/false or agree/disagree dichotomies are used in response.

Questionnaires are either administered directly by the investigator or assistant(s) appointed for the purpose or posted to the respondents. In the latter, series of reminders may have to be sent before the completed questionnaires could be returned to the investigator.

Questionnaire method of collecting affective information enables respondents to speak-out their mind or express their feelings or attitudes on issues as long as their confidentiality is guaranteed. Unlike face-to-face interview whereby the respondent could be shy or feel embarrassed when asked some questions, respondents feel freer in answering items on questionnaires.

The use of questionnaires is not without problems. One major issue is the respondents' willingness to spare time to complete the items on the questionnaire. This is why it is suggested that items on a questionnaire should be short and concise. Lengthy questionnaires are usually avoided, delayed or left unanswered. A lot of motivation and persuasion will be required to make respondents answer all the items of such lengthy questionnaire.

Another issue of note is the language of the questionnaire. When designing questionnaires, the characteristics of the potential respondent should be borne in mind. The language of the questionnaire should be one that respondents will understand.

The issue of the sincerity of response is also noteworthy. In situations where questionnaire items contain sensitive issues or one that deals with job security for example, responses to be given may not be the actual feelings of respondents as some deceptions might crop in.

References

This method of obtaining affective information is popular among job applicants and those seeking admission to institutions of higher learning. Usually, references or recommendations are letters written about an applicant by someone who might have worked with the applicant or taught him or her in school. This type of letter is not expected to be written by a blood relation; rather by someone who does so in an official capacity. Reference letters are written and sent directly to the agency or institution requesting for them. The information that reference letters should contain include those of the applicants' skills, brilliance, sincerity, honesty, perseverance, humility, innovativeness, among others.

The major problem of this method is that candidate could contract a relative or someone that will not give a sincere report about his/her applicant. The information provided by such a relative is capable of misleading the requesting authority in its decision-making process.

Observation Technique

This is a popular method used in collecting data on affective variables in child psychology. It is also employed in the training of pre-service teachers. Apart from these, a number of researches employ the observation technique as a means of gathering information on an aspect of a person's behaviour.

An observer could collect his observational data by identifying (i.e. interacting) with the subjects of the observation. He could as well do it as a 'bystander' to the group.

A major limitation of the observation technique is the conversion of observed behaviours to scores. Efforts should be made to avoid misrepresentation of subjects' behaviour by under-presenting or exaggerating it in figures. This is why it is suggested that standardized observation sheets where subjects' behaviours could easily be converted to scores with minimal efforts be used.

In classroom observation, the observer could make use of Flander's Interaction Analysis (Flanders, 1960) in which classroom interaction has been broadly divided into two broad categories – teacher talk and student talk. The teacher talk is further divided into direct and indirect influences. Four categories of teacher's direct influences are listed on the instrument while there are three categories of teacher's indirect influences.

As a result of advances in technology, more accurate and reliable observation data are now obtainable by the use of video/TV cameras, in which the subjects of observation may not even know that they are being observed. The tape could then be analyzed (re-observed) later for greater details.

Interview

This is another means of collecting affective data where the investigator and the interviewee talk to each other. The method is usually employed in selection processes into jobs or educational institutions or for the collection of data needed for research purposes. In an interview, the investigator asks questions from the respondents and he makes some recordings on the responses obtained.

The interview could be structured or unstructured. In structured interview, the interviewer has written out all the questions he intends to ask the respondent(s). The questions are then asked as listed on the interview schedule. In unstructured interview, the schedule of questions is not written down by the person conducting the interview. The type or nature or sequence of questions asked may depend on the responses given to an earlier question by the respondents.

It is important to note that interview questions could be culturally sensitive. Languages that are considered offensive, vulgar or impolite should be avoided as much as possible when conducting an interview. Another problem with interviews is the sincerity of the respondents. On a number of cases, respondents may not say the truth concerning the

question posed to them. A skillful interviewer, however, should be able to detect insincere responses.

Sociometric Techniques

Sociometric techniques of collecting affective data emanated from sociometry – the study of the interaction process of members of a group. The group could be a class, workers in a factory or members of a congregation. The sociometric technique is a system whereby members of an intact group supply information about one or more members of the same group. For example, in the secondary school, students' assessment of the leadership ability of some students nominated as prefects could aid the decision-making process of the school administration in appointing school prefects. Sociometric technique could also be employed by organizations in choosing their officers. This method of collecting affective data is commonly employed, though unnoticed.

Anecdotal Records

An anecdote is a short account of a remarkable behaviour or an interesting incident of a biographical nature which when can be used to illustrate the possession or absence of a particular quality in an individual. In the classroom, an incident might happen in which a student shows a remarkable/striking behaviour, the teacher is expected to record such behaviour as exhibited by the student without any value judgement. Such records are considered vital in enhancing the better understanding of students.

In most public schools today, this record is hardly kept by classroom teachers. It is suggested that when recording anecdotes, the information should not be exaggerated to eliminate bias. The record is expected to be part of the teacher's recorded needed for any future reference.

ACTIVITY III

1. Identify three factors that can influence the data collected through an unstructured interview
2. Highlight the advantages of the questionnaire method over the observational technique.

SUMMARY

- In this unit we defined affective measures as those variables which describe a person's qualities, emotions, values and attitudes in relation to the feelings of others. Examples of affective measures include friendliness, neatness, attitude to colleagues, attitude to school work, attendance, self-concept, etc.
- Affective variables can be measured in various ways. The questionnaire or rating scale is the most popular method of measuring the affective qualities of people.

Other methods that are also used include the observation technique, the interview, sociometric techniques and the use of anecdotal records. The strength and weaknesses of each of these methods were discussed.

ASSIGNMENT

1. Describe four factors that are capable of affecting data collected for effective measures.

REFERENCES

Anastasi, A. (1982). **Psychological Testing**. New York: Macmillan Publishing Co. Inc.

Brahm, N. (1981). "**The assessment of self concepts of educational achievement by a criterion referenced approach.**" (ERIC Document Reproduction Service No. ED 235 184)

Burger, J.M. (1990). **Personality**. California: Wadsworth.

Hopkins, K.D. (1998). **Educational and Psychological Measurement and Evaluation**. Boston: Allyn & Bacon.

Mehrens, W.A. (1975). **Measurement and Evaluation in Education and Psychology**. New York: Holt, Rinehart and Winston.

Murray, J.H.A. (1938). **Explorations in personality: A clinical and experimental study of fifty men of college age**. New York: Oxford University Press.

UNIT 5: TEST VALIDATION I

INTRODUCTION

Experience has shown that most teachers construct and administer tests on their pupils/students without caring to know anything about the qualities of items that make up the tests. Some of the test items, no doubt, are of good quality either because they have been carefully prepared by the teachers or because they were 'lifted' from past question papers of standardized examinations. Test validation therefore refers to all the procedures and processes that are followed in order to ensure that tests administered on students are of very good quality. In this unit, you will be taught some of the procedures involved in test validation.

OBJECTIVES

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

- i) give an acceptable definition of the terms 'validity' and 'reliability';
- ii) state the difference between validity and reliability;
- iii) discuss types of validity.

HOW TO STUDY THIS UNIT

1. Skim through this unit to have an idea of its content and the key concepts discussed in it.
2. Read through the unit carefully and ensure that you understand the key concepts and terminologies used.
3. Attempt all the activities and assignments included in the unit without referring to the text.

WORD STUDY

Consistency of scores:	This refers to the stability of scores over time by the same scorer or different scorer at the same time.
Coefficient:	Score or value obtained from the process of estimating the reliability or validity of a test or measuring instrument.
Validity:	Accuracy with which a test measures what it is designed to measure.
Reliability:	Consistency with which a test measures what is designed to measure.

RELIABILITY OF TESTS

The reliability of a test refers to the extent to which the test is consistent in measuring what it is designed to measure. It is a measure of the degree to which a test measures the same thing time after time and item after item. Reliability shows the extent to which a test score is consistent when scoring is done by the same person on different occasions or by different persons at a time. The key-word in reliability is *consistency*.

In reliability theory, each observed score of a student has a “true score” and an “error score” component and it is expressed as:

$$X = T + E \text{ where}$$

X is the observed score; T is the true score and E refers to the error score. This is to say that however small (especially for essay questions), every score awarded a candidate contains the true score and some error score. Theoretically, reliability is defined as the ratio of the true score and the observed score variances. As error variance decreases, reliability increases.

The standard error of measurement provides a measure of intra-individual variability in test scores. The standard error of measurement is given by:

$$S_e = S_x \sqrt{1 - r_{xy}}$$

where S_e = standard error of measurement,

S_x = score variance

r_{xy} = reliability coefficient.

Estimating Reliability

There are 3 procedures commonly used to estimate the reliability of a test. These are:

- (a) Test-Retest Reliability
- (b) Parallel Form Reliability
- (c) Internal Consistency.

The test-retest reliability is obtained by finding the correlation between two administrations of the same test over a period of time, say two or three weeks. The results emanating from the two administrations are then correlated using Pearson Correlational Analysis. The result of this analysis constitutes the reliability coefficient for the test.

The parallel form reliability is a measure of equivalence of two equivalent forms of a test administered on the same group of testees on the same day. The equivalence is established with respect to equality of content, means and variances. In this way, we are trying to generalize a testee's score to what he or she would receive if he or she took a test composed of similar but different questions.

The internal consistency measure of reliability is always employed when a form of a test is administered only once. There are three different ways of measuring the internal consistency of a test.

(i) **Split Half:**

Here, a single administration of a test is done and the result separated into two usually by using the even-odd number arrangements. The scores of odd numbered items are then correlated with those of the even numbered items using the Pearson Product Moment Correlation formula. After this, a correction formula will be employed because the correlation coefficient obtained is an estimate of the reliability of a test only half as long as the original. Thus, the estimate of the reliability of the whole test would be achieved by using Spearman Brown Prophecy formula which is expressed as:

$$r_{xx} = \frac{2r^{1/2}{}^{1/2}}{1 + r^{1/2}{}^{1/2}} \quad \text{where}$$

r_{xx} = estimated reliability of the whole test

$r^{1/2}{}^{1/2}$ = reliability of the half-test.

(ii) **Kuder Richardson Formula.** This formula is in two versions: K-R20 and K-R21.

$$\text{K-R20} = r_{xx} = \frac{n}{n-1} \left\{ \frac{1 - pq}{S_x^2} \right\}$$

$$\text{K-R21} = r_{xx} = \frac{n}{n-1} \left\{ \frac{1 - \bar{x}(n - \bar{x})}{nS_x^2} \right\}$$

where: n = number of test items

p = proportion of testees who answered an item correctly

q = proportion of testees who answered an item incorrectly

pq = variance of a single item

S_x = variance of the entire test

\bar{x} = mean of the entire test

(iii) **Coefficient Alpha:** This formula was developed by Cronbach (1951) as it is given as:

$$P_{\infty} = \frac{n}{n-1} \left\{ \frac{1 - \sum S_i^2}{S_x^2} \right\}$$

where S_i^2 is the variance of a single test item

S_x^2 is the variance of the entire test.

Examples of calculations using the “formulae” will be discussed in statistics lectures.

ACTIVITY I

1. Define “reliability” in your own words
2. Discuss briefly two methods of estimating the reliability of a test.

ENSURING THE VALIDITY OF TESTS

Validity is perhaps the most important characteristic of a test. It refers to the extent to which a test measures what it purports to measure. Validity refers to the accuracy with which a test samples a particular trait for which the test is designed. There are four major types of validity:

(a) Content Validity

This is the extent to which the test items represent the content domain and universe of situations of the subject matter. Content validity could also be described as the extent to which a test covers all the topics covered or listed in a subject. The weight of representation of each topic will now be a function of the time or attention given to the teaching of each topic. The important objectives in terms of domain of behaviour must also be sampled or measured by the test.

(b) Criterion Related Validity

This type of validity is used by test experts to determine the validity of aptitude tests. Aptitude tests are those tests designed to find out the likely performance of somebody in an activity or subject or course of study. For example, the Banking Operations Aptitude Test (BOAT) is an aptitude test administered by NECO in Nigeria to find out how good a prospective banker is likely to be. The validity of aptitude tests is usually determined through the use of criterion related validation. Criterion related validity is of two types. These are predictive and concurrent validity.

The predictive validity of a test is determined by comparing the test result of a subject when examination/test is written after the aptitude test. Its validity is then determined by the extent to which the aptitude test (taken earlier) predicts the subject test (written later). For example, the predictive validity of a test of manual dexterity taken prior to the commencement of training as a secretary will be determined by such person’s performance in a typing course taken after registration for the programme.

The concurrent validity of a test is determined by comparing the result of an aptitude test which has just been taken with the result of a subject test written prior to the administration of the aptitude test. The concurrent validity of such a test written earlier will be determined by the extent of relatedness of the result with that of the aptitude test. The only distinguishing factor between predictive validity and concurrent validity is the time lag between the time an aptitude test is taken and the time a test in a related school subject or occupation is taken.

(c) **Face Validity:**

This refers to the quality of the whole test instrument. Face validity has to do with the typing or print legibility or clarity, appropriateness and extent of clarity of language used, quality of the test paper, perceived relevance of the materials used for test item, etc. Anything that has to do with the judgement which could be meted out to a test instrument is part of face validity. It has to do with whether or not the instrument is valid “on the face of it”

(d) **Construct Validity**

This is the degree to which test scores can be accounted for by certain explanatory constructs in a psychological theory. For example, if a construct like introversion is to be measured in a test, then a test with construct validity will have items that deal with those variables that tend to define introversion. When this happens, then the test can be said to have construct validity.

ACTIVITY II

1. In not more than twelve words, define validity.
2. Mention and explain the various types of validity.
3. Identify the essential difference between validity and reliability of test instruments.

SUMMARY

- In this unit, we have discussed the procedures involved in the validation of tests. These are procedures embarked upon to make test items to be of high quality so that they will have high validity and reliability. The validity of a test refers to the extent to which the test measures what it supposed to measure. Types of validity include content, criterion-related, face and construct validity. On the other hand, the reliability of a test refers to the consistency of test scores over time and across examiners. Reliability could be determined through the use of test-retest method, parallel method and internal consistency method.

ASSIGNMENT

1. In a tabular form identify four types of test validity and state the conditions appropriate for each of them.

REFERENCES

Anastasi, A. (1982). **Psychological Testing**. New York: Macmillan Publishing Co. Inc.

Hopkins, K.D. (1998). **Educational and Psychological Measurement and Evaluation**. Boston: Allyn & Bacon.

Mehrens, W.A. (1975). **Measurement and Evaluation in Education and Psychology**. New York: Holt, Rinehart and Winston.

Ughamadu, K.A., Onwuegbu, O.C. & Osunde, A.U. (1991). **Measurement and Evaluation in Education**. Benin City: World of Books Publishers.

UNIT 6: TEST VALIDATION II

INTRODUCTION:

In the previous unit, you have been introduced to two major procedures (validity and reliability) for ensuring that tests administered by classroom teachers as well as those administered by counsellors are of good quality. However, apart from these two procedures which normally take place before the final administration of a test, there are other steps which a test administrator takes after he or she has administered and scored his or her test. In this unit, you will be introduced to these other procedures, which are designed to improve not only the quality of the present test but also the quality or validity of similar tests to be administered in future.

OBJECTIVES

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

- i) explain the concept of difficulty level and discrimination index;
- ii) compute the difficulty level and discrimination index of test items;
- iii) explain the usefulness of distracter analysis in multiple-choice items.

HOW TO STUDY THIS UNIT

1. Read through the unit carefully making sure that you understand the key concepts and terminologies used.
2. Where mathematical formulas are given ensure you understand what they represent.
3. Attempt all the activities and assignments included in the unit.

WORD STUDY:

Index	A mathematical notation which expresses the ratio of one dimension of a variable being measured to the other.
Discrimination:	The ability or power to make a distinction among test items.
Distractor:	One of the incorrect answers presented as a choice in a multiple-choice test.

DETERMINING THE DIFFICULTY LEVEL AND DISCRIMINATION INDEX OF A TEST

Test difficulty level is the third procedure for ensuring that a test is good. It is also called the facility index of a test. The difficulty level of a test is the extent to which a test is said to be easy or difficult for the testees. This is determined by finding out the

proportion of testees that got each item on the test correctly. For example, the difficulty index of a test item taken by 20 students will be determined by finding the proportion of those who got a particular item correctly. If 12 testees out of 20 got item 1 correctly, it means the difficulty level of item 1 is $12/20 = 0.6$. The same procedures will be followed for all the other items or questions of the test and the difficulty index will be obtained by adding the difficulty index of all the items and finding the average.

One important thing that should be noted is that the higher the difficulty index of a test, the easier the test and vice-versa. The facility or difficulty index of a test is not expected to be too low or too high. It is expected to be moderate. One thing that will be of help is to generate test items from the content areas and domains of behaviour (as contained in the scheme of work or curriculum) list for the potential testees.

The Discrimination Index is the extent to which each item of a test differentiates between the superior ability students and inferior ability students. Such discrimination is expected to be in favour of the superior ability students. The discrimination index is defined as:

$$D = \frac{U_{25\%} - L_{25\%}}{N}$$

Where D is the discrimination index

$U_{25\%}$ is number of testees who got an item right in the upper 25% group

$L_{25\%}$ is the number of testees in the lower 25% group who got the item right

In computing the discrimination index, it is important to do the following:

1. Arrange the scores of all the testees in descending order of magnitude from highest to the lowest score
2. Group the scores into three parts by getting those in the upper 25% and lower 25% of the group
3. $U_{25\%}$ will be the number of testees within the upper 25% group who got an item correctly. The same for $L_{25\%}$
4. N refers to the number of testees in upper or lower 25% group.

It is important to note that in a good test, the discrimination index should be average. If for example, all the students in a class got an item correctly or wrongly, it means the item cannot be used to differentiate or “discriminate” between the brilliant and the less brilliant in the class. A good test item will be such that majority of students/testees that belong to the upper 25% will get it correctly while majority of those in the lower 25% are also expected to get it wrongly.

ACTIVITY I

1. What are the uses of difficulty and discrimination indices in test development
2. Explain the essential difference between difficulty index and discrimination index.

TEST VALIDATION USING DISTRACTER ANALYSIS

Distracter analysis is simply the examination of the strength of each of the distracters. The distracters in a test are the non-correct options. In multiple choice tests, it is expected that the options will be plausible i.e. as close to the correct answer (key) as possible so that candidates who lack the requisite knowledge will be 'tempted' to pick any of them as the answer to the item. If any of the options so provided is very far from the answer, then such an option or distracter should be replaced.

The procedure employed in doing this is a simple one. For each of the items, the test developer makes a list of the options and does a frequency count of how many of the testees picked each of them. For example, in a class of 30 students, if a two-item five-option multiple choice test was conducted in Mathematics, for each of the items, the five options will be listed and a count of the number of students that picked each of the distracters will be obtained. Any distracter with low frequency or with the least frequency when compared with others is considered poor and should be replaced by a better option that is closer to the answer. If the number of those that picked any of the distracters is high, it means the option is plausible and should be retained.

ACTIVITY II

1. Explain with examples the procedures for distracter analysis.

SUMMARY

- In this unit, you have discussed three other procedures for ensuring that our tests are of good quality. The first procedure is to determine the level of difficulty of test items. This is determined by finding out the proportion of testees that got each item on the test correctly. The second procedure is to determine the discrimination index, which is the extent to which each item of the test differentiates between the superior ability students and inferior ability students. Finally, if the test is a multiple-choice objective test, the test developer undertakes the distracter analysis. This is done by making a list of the options in the test and counting how many of the testees picked each option. You are encouraged as much as you can to employ the use of the procedures whenever you are developing or generating test items.

ASSIGNMENT

1. Construct a five-option multiple choice objective items in your teaching subject; administer it on a sample of students equivalent to those for which the test is meant. Thereafter, determine the difficulty level, discrimination index and distracter analysis for the items on the test.

REFERENCES

- Hopkins, K.D. (1998). **Educational and Psychological Measurement and Evaluation**. Boston: Allyn & Bacon.
- Mehrens, W.A. (1975). **Measurement and Evaluation in Education and Psychology**. New York: Holt, Rinehart and Winston.
- Ughamadu, K.A., Onwuegbu, O.C. & Osunde, A.U. (1991). **Measurement and Evaluation in Education**. Benin City: World of Books Publishers.

UNIT 7: TEST CONSTRUCTION

INTRODUCTION

It is an issue of common knowledge that the whole process of teaching and learning is incomplete without an attempt to evaluate what has been taught or learnt. The evaluation phase of the teaching-learning process usually provides insight into the effectiveness of the 'process' in the achievement of the objectives of any programme. Teachers, psychologists and counsellors make use of tests to evaluate the effectiveness of services provided by them. To be effective, tests used by teachers, psychologists and counsellors must satisfy some laid-down conditions which psychologists have agreed upon over the years. In this unit, you will be introduced to the standard procedures for test construction.

OBJECTIVES

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

- i) discuss the general principles of test construction
- ii) explain the procedures for conducting essay and objective test items.

HOW TO STUDY THIS UNIT

1. Read through this unit and make sure you understand the meaning of the key concepts in test construction.
2. Study the new words under the section on Word Study.
3. Use your dictionary to check the meanings of words which you find difficult in the text.
4. Attempt all the activities and assignments included in the unit without referring to the text.

WORD STUDY

Item: This is synonymous with question. It is the task presented to the learner or testee in order to determine the extent of presence of a trait or ability.

Table of Specification: A two-way dimensional table which helps to ensure that content and process objectives are adequately represented in a test.

PLANNING TO WRITE A TEST

The first step in test construction is planning. As mentioned earlier, planning is an important issue in the construction of achievement tests. A building without a plan may not have any appreciable beauty. Therefore, in planning a test, the content to be covered

by the test must be specified. Just like building a house, the area to be covered by the new building must be marked out for foundation-laying activities. Thus, the number of topics to be covered by the test is one of the focuses of test planning.

Test planning involves the specification of the objectives of writing the test. This means that before teachers construct their tests, they should make a list of topics and instructional objectives that are to be measured by the test. This should be done with great care so that no topic will be over-represented or under-represented in the test.

Also, for good planning, test items must be spread across the six domains of cognitive behaviour as specified by Bloom (1956). When items in a test adequately cover the topics listed in the curriculum or scheme of work and the behavioural domains, the test is said to have content validity. The difficulty level, length and format which achievement test will follow are a function of the purpose of the testing and the objectives to be measured.

ACTIVITY I

1. List and discuss the essential things that must be ensured in planning to construct a test.

DOMAINS OF INSTRUCTIONAL OBJECTIVES

Instructional objectives are statements which indicate “in active verbs” those activities which learners are expected to demonstrate at the end of a unit of instruction. They are also intended outcomes stated in such a way that their attainment (or otherwise) can be observed or measured. Bloom (1956) led other educationists in writing on this issue. To Bloom, objectives used in classroom teaching-learning processes could be categorized into three main domains: (1) cognitive (2) affective, and (3) psychomotor.

The cognitive domain includes those objectives that emphasize the attainment, retention and development of knowledge and intellect. The acquisition of subject-matter skills/concepts is a function of the cognitive ability of individuals. Cognitive domain is the centre-point of most educational activities including testing. Objectives that are based on cognitive domain emphasize students’ behaviour in the area of knowledge, ability and understanding. These involve the development of the thinking of gradual improvement in the child’s intellectual make-up. An example of a cognitive objective is to state that at the end of a lesson, students should be able to explain the meaning of inflation.

Affective objectives encompass objectives that emphasize behaviours characterized by feelings, values and emotions. Such feelings, values and emotions may be positive or negative. Affective objectives are those that seek to express the control of negative emotions and the students’ appreciation of positive ones.

The psychomotor domain includes those objectives that emphasize physical manipulative skills in and out of the classroom, laboratory, workshops, auditorium or dance theatre. Learning objectives in Fine Arts and Physical Education usually fall under this domain.

It should however be pointed out that the cognitive domain has been over-emphasized in most educational settings especially in elementary and secondary school levels at the detriment/disadvantage of the other two domains. Therefore, in teaching and testing at these two levels of schooling, efforts should be made to incorporate the affective and psychomotor domains of behaviour.

BLOOM'S COGNITIVE DOMAIN OF LEARNING

Bloom classified the cognitive domain of learning into six levels in ascending order of complexity. The levels are knowledge, comprehension, application, analysis, synthesis and evaluation. It is at each of these levels that questions prepared by teachers are meant to test students' ability.

Questions to be prepared under the cognitive domain should begin from simple to complex. This means that testing of students will be meaningless unless they have acquired certain amount of information in the knowledge domain. Knowledge is required before one can attain the understanding of a concept. This is to say that each of the six levels of cognitive skills proposed by Bloom is a prerequisite for the subsequent higher level. To be able to perform well at the high levels (analysis, synthesis and evaluation), the lower ones (knowledge, comprehension and application) must have been mastered. The higher levels require much thinking on the part of the students than needed at the lower levels. Thus, tasks or questions at this level measure higher mental processes. At each of the levels, some active verbs are expected to be used by objective writers. For instance, at the level of knowledge, verbs such as *define, state, list, name, write, recall, label, reproduce, mention, etc* are used in stating the behavioural objectives. At the level of comprehension or understanding, associated verbs include *identify, describe, explain, name, classify, etc*. Verbs used at the level of application include *demonstrate, construct, compute, perform, manipulate, show, solve, use, etc*. At the level of analysis, verbs such as *breakdown, separate, differentiate, classify, compare, contrast, etc* are used. At the level of synthesis, verbs like *summarise, compose, derive, design, device, create, discover, generalize, etc* are used in writing objectives. Also, at the level of evaluation, verbs that are used include *criticize, assess, appraise, justify, defend, evaluate, judge, conclude, etc*.

ACTIVITY II

1. Select a topic in your teaching subject to be taught to students and write one objective for each of the three domains of instructional objectives.

GUIDELINES FOR CONSTRUCTING ESSAY QUESTIONS

Essay questions are those requiring testees to present their responses in their own language. These could be in two forms: (i) Questions requiring extended response and (ii) those requiring restricted response.

In the extended response type, candidates are not limited in the form and length of their answers while in the restricted response type, candidates are limited in the choice of language or/and scope of their writing.

One advantage of essay item requiring extended response is that it is a good means of testing and developing students' writing abilities. The extended response form of essay writing gives total freedom to students' in their choice of the pattern of answering questions. It is also very useful in measuring higher level mental processes.

The disadvantage of an essay question requiring extended response is that its marking is very subjective. There is the possibility of scoring the response with bias. Some extraneous factors like candidate's handwriting, style or skill of presentation and examiner's mood can influence the award of marks.

The restricted essay questions are easier to use in measuring students' learning of basic facts and are more useful at the lower levels of schooling. They are easier to mark and are less subjective than the extended essay responses. However, restricted essay questions are more difficult to set as their development takes more time, consume more stationery and printing materials.

The following guidelines are helpful in constructing essay questions:

- (i) Ambiguous or difficult language especially words with multiple interpretation must be avoided. The language of a test must not confuse the testees but must communicate unambiguously to them what is required of them.
- (ii) Teachers or test administrators should not give too long or too short time as the duration for the test. If the test duration is too long, students will get bored and if it is too short, it will lead to disappointment as most testees may not finish the test.
- (iii) Marking guides/scheme must be prepared along with the test items.
- (iv) Teachers should avoid questions that require mere recall or reproduction of facts either from textbooks or from class notes. Instead, teachers should include items that will ensure the development of students' expressive and analytical abilities.

GUIDELINES FOR CONSTRUCTING OBJECTIVE TESTS

Objective tests otherwise known as structured response tests have a stem which ask the questions or present the problem as well as suggested answers otherwise called options or foil. The commonly used types of objective tests are the multiple choice, true-false, matching, short-answer, completion and close tests. The multiple-choice, short-answer and completion types are more frequently used in day-to-day classroom achievement testing.

In multiple choice objective tests, there is a correct or best answer called the "key" while the other options that are incorrect are called "distracters". Distracters are so called

because their function is to confuse or distract the testees away from the correct answer. To be able to perform this function, distracters should be plausible i.e. be as close to the correct option as possible.

One of the major advantages of an objective test is that the structured manner of the responses makes marking easy to the point that several independent marking by the same person or different scores result in similar or same score. Another benefit apart from the ease of scoring is that it is useful in covering a large part of a course or subject. Also, it does not give room for bias in its scoring. However, objective tests may take much longer time to prepare; it consumes more stationery, gives room for surface-level or superficial learning and is vulnerable to guessing and cheating by testees.

The following guidelines should be adhered to in the preparation and administration of objective tests:

- (a) Course objectives and course content should be taken into consideration. All the items to be included in the test should be relevant to the series of instructional objectives specified during teaching and the contents of the course. These two considerations affect the validity of the test. To ensure adequate content validity, teachers and test developers must ensure that they develop test-blueprint or table of specification and use it as a guide in determining the number and nature of their objective test items.
- (b) Questions should be prepared well in advance.
- (c) More questions than needed should be prepared for each test. This will enable the teacher to leave out some items found to be inadequate before the final production of the entire test.
- (d) When multiple choice items are used, teachers must ensure that the distracters are plausible. This can be done by administering the test on a different but equivalent sample and computing the discrimination index, difficulty index and distracter analysis before the actual administration of the test.
- (e) Teachers must use clear, unambiguous language when preparing objective tests.
- (f) Teachers should plan their tests to cover all the six domains of learning especially the higher levels.
- (g) Teachers should allow others especially senior colleagues to look at the questions and offer their suggestions/criticisms in order to improve its overall quality.

ACTIVITY III

1. Discuss five important things which a teacher must do while constructing a test.
2. State the major difference between objective and essay tests
3. Mention two advantages of objective tests over the essay type.

SUMMARY

- In this unit, we have discussed test planning as the first step in item construction. You will recall that we stated that test planning involves the specification of the content area to be covered, the behavioural objectives to be sampled and the level at which content area and the behavioural objectives will be tested. This is best done through the preparation of a table of specification otherwise called the test blueprint. After this initial planning comes the determination of the format to be employed in testing. This could be essay or objective format. Essay questions allow candidates to supply the answer in their own words and are either in the form requiring extended or restricted responses. Objective tests are those in which candidates are provided with options from where a correct answer is to be picked. Since the motive of conducting tests is not to make testees fail or to get them confused, questions should be prepared in simple, unambiguous language and instructions should be clearly specified.

ASSIGNMENT

1. Construct a ten-item multiple-choice objective test in your teaching subject making sure that the distracters are plausible and the instructions are simple and clearly stated.

REFERENCES

- Anastasi, A. (1982). **Psychological Testing**. New York: Macmillan Publishing Co. Inc.
- Hopkins, K.D. (1998). **Educational and Psychological Measurement and Evaluation**. Boston: Allyn & Bacon.
- Krathwohl, D.R. *et al* (1964). **Taxonomy of educational objectives: Handbook II, Affective domain**. New York: D. McKay.
- Ughamadu, K.A., Onwuegbu, O.C. & Osunde, A.U. (1991). **Measurement and Evaluation in Education**. Benin City: World of Books Publishers.

UNIT 8: TEST ADMINISTRATION AND SCORING

INTRODUCTION

You have at one time or the other taken a test or a public examination in the course of pursuing an educational career. On one of those occasions, you must have been exposed to some conditions in the test or examination room which tend to have negative effect on your performance in the test or examination. Can you remember any of those conditions? How did you feel when you were unduly harassed by the invigilators? What about the anxiety you nursed over the sincerity of the person that will grade your scripts? Besides from your academic abilities, these are factors which might have affected your performance in the test or examination. In this unit, we shall discuss certain factors in the administration and scoring of tests which have implications on the eventual performance of students in a test or examination.

OBJECTIVES

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

- i) identify the factors necessary for the smooth conduct or administration of tests;
- ii) discuss common problems associated with test administration;
- iii) describe the methods of scoring test responses; and
- iv) discuss factors that could introduce bias into test scoring.

WORD STUDY

Scoring: This refers to the procedure involved in converting the responses of testees into scores. Note that in psychological testing, responses are not scored as right or wrong; rather they are described in scores which represent the level of presence or absence of the variables of interest.

Bias: A statistical sampling or testing error caused by systematically favoring some outcomes over others

Malpractice: Illegal and improper conduct of testees which are contrary to established rules and regulations guiding the conduct of tests or examinations.

TEST ADMINISTRATION

The administration of psychological tests is a vital aspect of psychological testing. Psychological tests could be administered on individual or group basis. In whichever case, the tests should be administered on testees/clients in a tension-free and conducive atmosphere. This is to say that tests should be administered in serene and comfortable environmental situations free of intimidating factors. Hence, psychological as well as physical/environmental factors should be taken into consideration in test administration.

Physical/Environmental Factors:

Physical factors to be considered in the administration of test include conducive examination room and furniture, adequate lightening, good ventilation of the test room, etc.

Conducive physical conditions for the testees begin with the availability of comfortable chairs and desks on which they are to sit and write the test. Situations whereby testees have no comfortable furniture to themselves could mar the success of the test.

It is also necessary that examination/test room is well lit. If the test venue is poorly lit, testees may be unable to perform well and the validity of their responses and the results arising from the process may become questionable.

Of equal importance is the issue of ventilation. Adequate ventilation is needed for any testing situation – whether group or individual. Counsellors, psychologists and teachers should ensure that test venues are adequately ventilated. Poor ventilation could be dangerous for the safety of testees especially when gases or chemicals are to be used in testing, e.g. in a Chemistry practical test. Adequate provision should be made to avoid suffocation and to ensure that the temperature of the test venue is not unbearable for the testees.

It is also important that undesired sounds or noise from within or without the test venue should be checked. If noise is not checked, it could easily distract the attention and concentration of the testees. It is therefore important that test administrators should endeavour to control noise from the immediate surrounding of the test venue. This could be done by writing inscriptions telling passers-by of the conduct of the test. In the case of noise within the test venue, it is essential that invigilators or the person(s) administering the test maintain the requested silence by making sure that all the needed instructions are passed before the commencement of the test. Not only this, intermittent announcements of the duration or ‘time left’ for the test should not be allowed to constitute noise and disturbance to the testees. A wall-clock, which could be hung conspicuously in the test venue, should be used for this purpose. Also, testees should not be allowed to sit too close to one another to reduce if not totally eliminate communicating with one another.

Psychological Factors

Psychological factors to be considered in test administration are factors that have to do with the emotional stability of the testees. Most test administrators are not careful enough on this issue. First, test administrators should ensure that testees are informed in advance about the conduct of the test. Tests should not be conducted unannounced as this destabilizes the testees. Second, when administering tests, testees should be made to enjoy some level of emotional stability. The use of threats on the testees should be avoided. There should be efficient distribution of test materials and these should round all testees to avoid anxiety. When all these are done, testees will be able to demonstrate their maximum ability and sincere and accurate responses.

ACTIVITY I

1. Discuss three physical and three psychological conditions that must be considered for successful administration of tests.

PROBLEMS ASSOCIATED WITH TEST ADMINISTRATION

There are two major problems confronting the administration of tests, especially those tests that are administered within the classroom or school setting. These are (i) logistic problems, and (ii) malpractices. Test administrators should consider these problems and plan ahead to take care of them when planning to administer school-based tests or examinations.

The problem of logistics includes the availability of test materials such as question papers, answer booklets, etc. in adequate numbers. There are also the problem of efficient distribution of test materials (when testees are large in number) and all those factors that have to do with availability and suitability of test venues. The person wishing to administer a test should ensure that the factors mentioned above are taken into consideration if a successful test administration regime is to be achieved.

Apart from the problems of logistics and suitability of test venues, there is the problem of examination malpractice which test administrators today have to contend with. Malpractice involves a deliberate act of wrong-doing, contrary to official examination rules, intended to achieve an unfair advantage for a particular candidate or group of candidates. Cheating is a word that is regularly used interchangeably with malpractice. For clarity, cheating involves activities that contradict fair conduct at the test venue while malpractice encompasses all the efforts aimed at getting undue advantage either within or outside the test venue. While all forms of cheating constitute malpractice, certain activities in the examination room such as noise making, harassment, disrespect for invigilators, etc may not be regarded as cheating but purely as examination malpractices.

A number of reasons account for why testees engage in malpractices during the conduct of tests or examinations. These include the high value or premium placed on paper qualification, quest for teacher/school status protection, inadequate school facilities, inadequate preparation by students, lapses on the part of the person or body administering the test, location of test centres and general level of moral decadence in the society.

Most malpractices go undetected. However a number of counter measures designed to discourage malpractice would include gender separation in the hall, strict monitoring by officials of examination bodies, thorough supervision, daily distribution of question papers, arrest and prosecution of offenders and also recently, the passage of examination malpractice decree number 33 of 1999 by the Federal Government.

ACTIVITY II

1. Using your experience as a classroom teacher, highlight what you consider the commonest forms of examination malpractice among students and discuss what can be done to minimize the malpractices.

SCORING OF TEST RESPONSES

In test scoring, the type of scoring method to be employed depends on the type of test administered on testees/clients. Essentially, there are two ways by which scoring could be done. These are by manual scoring and machine scoring.

(a) Manual Scoring

Irrespective of the nature of test items (i.e. whether objective or subjective) the response could be scored manually i.e. by hand. Manual or hand scoring refers to the scoring of responses to test items either by the teacher/counsellor, the testees themselves or by someone appointed to do so.

Manual/hand scoring are of two types. The first type is global or impressionistic marking, which is usually employed when scoring the extended essay responses. In this type of scoring, the scorer reads through the responses and forms an impression (which is rather subjective) on the number of point to be awarded. Under this system, a number of factors could easily influence the scoring. These include legibility of handwriting, grammar and the scorer's mood.

The other type of manual scoring is the analytic scoring otherwise referred to as the point scoring method. Under this method, the scorer bases his or her judgement on a number of pre-specified criteria. These might include things like clarity of expression, number of points raised (in case of achievement tests), logical presentation of facts, mechanical accuracy, etc. Usually, the scorer makes use of a frame of reference (or marking guide) in doing the scoring.

(b) Machine Scoring

This method is suitable for scoring objective items. Objective scoring means that respondents' scores remain constant irrespective of who does the scoring or how many times it is done. Due to advances in technology, machines (called scanners) are now used to score objective answer scripts. This method is faster and more efficient than hand or manual scoring. It is a method employed by examination bodies with large number of scripts to be scored within a limited time. The testees are to make use of answer sheets specifically designed to the specification of the scoring machines. These sheets are called the Optical Marks Recorder (OMR) sheets. From the pages of the OMR sheets, the scoring machine could pick the correct answers and immediately record the score for each candidate. This method is fast, almost infallible but very expensive.

ACTIVITY II

1. Identify two methods of scoring psychological tests and discuss the problems/limitations of each method.

BIAS IN SCORING

Every score awarded a testee, especially in essay tests, is said to contain a true score and some error scores. The error could be as a result of bias. Any factors that tend to introduce error component into the total score of a testee is introducing bias. Such factors, which eventually affect the validity of the scores, are discussed below:

Testees' Handwriting

Some testees have very good (i.e. legible and attractive) handwriting while others have illegible handwriting. This is capable of making the scorer exaggerate either negatively or positively the response. To ensure that testees with illegible handwriting do not suffer unduly, scores should always take the pain to read carefully even the worst of handwritings and award scores on the basis of the substance in the testees' scripts and not on the basis of the beauty/legibility of handwriting.

Relationship to the Scorer

In classroom situations, some testees may be known to the scorer (in this case, the teacher) as 'good boys' or 'bad boys'. This impression could influence the score to be awarded such testees. For this reason, the use of examination numbers is recommended for all examination/test situations. This will help to reduce the bias which relationship or personal knowledge of testees by scorers could bring.

Number of Scripts to the Scored

Number of scripts to be marked and the time limit within which the scoring is to be done is capable of influencing score reliability. This problem is peculiar to those scoring standardized test scripts (e.g. the SSCE). Since the scorers of such examinations are given a time frame which hardly allows the scorers to do any other thing during the period, there is the tendency for them to 'rush' the scoring. This might affect the accuracy with which the scripts are graded.

Testwiseness by Testees

Some testees have the knowledge of how to present responses in logical order, others do not. 'Testwiseness' could be described as the logicity with which testees present their responses to questions. When facts are presented in logical order, the scorer may be 'cajoled' to the point of positively exaggerating the response. The other way round is the case with illogical presentation of responses. To avoid biases caused by 'testwiseness', it

is necessary for scorers to carefully read through all the responses to each item and not to rush into impressionistic judgement.

ACTIVITY III

1. List and discuss the factors capable of affecting the reliability of scores.

SUMMARY

- In this unit, we have discussed the methods of scoring test responses. These are (a) manual scoring and (b) machine scoring. Manual scoring could be global or analytical. Global method of manual scoring involves scorers basing their judgement on the impressions created in them by the quality of testee's responses. Analytical method involves the basing of scores on pre-specified criteria usually spelt out in a marking guide. Machine scoring is applicable to the scoring of objective test items. It involves the use of machines or scanners in scoring testees' responses.
- Also examined in this unit are the sources of bias in test scoring. These include testees' handwriting, relationship to the scorer, examiners' mood, number of scripts to be scored vis-à-vis the time frame allowed for scoring.

ASSIGNMENT

Identify two major methods of test scoring and in a tabular form, highlight the advantages and disadvantages of each method.

REFERENCES

- Hopkins, K.D. (1998). **Educational and Psychological Measurement and Evaluation**. Boston: Allyn & Bacon.
- Mehrens, W.A. (1975). **Measurement and Evaluation in Education and Psychology**. New York: Holt, Rinehart and Winston.
- Ughamadu, K.A., Onwuegbu, O.C. & Osunde, A.U. (1991). **Measurement and Evaluation in Education**. Benin City: World of Books Publishers.

UNIT 9: CONSOLIDATION

INTRODUCTION

This module has introduced you to basic concepts in psychological testing. The module can be divided into three parts. The first part consists of unit 1 which provided you a working definition of psychological tests and also described the uses of psychological tests by counsellors. The second part (units 2, 3, and 4) discussed specific types of psychological tests while the third part dealt with basic principles of test construction and administration. This unit will serve to refresh your memories on the important issues discussed in each unit of the module.

OBJECTIVES

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

- i) explain the meaning and uses of psychological tests;
- ii) define achievement and aptitude tests and identify the major similarities and differences between them;
- iii) explain what is meant by intelligence and describe some popular intelligence test measures;
- iv) list and describe the various methods by which affective or personality variables are measured;
- v) discuss types of validity and reliability;
- vi) explain the concept of difficulty level and discrimination index;
- vii) explain the usefulness of distracter analysis in multiple-choice items; and
- viii) discuss the general principles of test construction.

HOW TO STUDY THIS UNIT

1. Read through the summaries of the various themes from Units 1 to 8 and ensure that you understand the main points.
2. Read through units 1 – 9 on which you have taken notes
3. Check the word study section for the meaning of unfamiliar words
4. Study the units as presented step by step.

MEANING AND USES OF PSYCHOLOGICAL TESTS

You will remember that in unit 1, we attempted a working definition of psychological tests and also mentioned the uses of some psychological tests to the counsellor and the classroom teacher. Specifically we defined psychological tests as series of instruments

containing statements or tasks seeking information that could help in the determination of the extent to which a particular trait, attribute or quality is present in an individual.

Specific examples of common psychological tests include:

- (i) Tests measuring intelligence
- (ii) Scholastic aptitude
- (iv) Achievement tests,
- (v) Vocational interest inventory
- (vi) Personality tests.

Psychological tests are used for the measurement of differences in attitudes of people and for classification and guidance. They are also used to determine individual differences and for the measurement and understanding of affective variables.

ACTIVITY I

1. Name four common psychological tests and discuss their usefulness to the counsellor and classroom teachers.

ACHIEVEMENT VERSUS APTITUDE TESTS

Achievement tests as those designed to determine the level of ability of learners in a given instructional task while aptitude tests are designed to measure ability that is likely to be demonstrated at a later time. While achievement test is present and past-oriented and content specific, aptitude test is future-oriented as it predicts ability that is likely to be demonstrated at a later time. It could also be said that achievement tests rely on prior and present learning experiences but aptitude tests forecast learning or benefit/performance in an unknown situation in the future. Although both achievement and aptitude tests measure ability, the difference comes in the aspect of the test content. While achievement tests are content-specific, aptitude tests are construct-specific i.e. the items focus largely on psychological constructs such as scholastic aptitude, manual dexterity, etc whereas the items in achievement tests are drawn from previously-specified content areas which have or ought to have been taught.

You should however bear it in mind that sometimes the difference between the two tests may not be clear-cut. For instance, it is possible to use achievement tests to 'predict' future abilities and skills or benefits from a programme. In the same way, aptitude tests may contain items which reflect the testee's prior learning and/or developed skills.

INTELLIGENCE VERSUS PERSONALITY TESTS

Intelligence tests are designed to measure the level of reasoning, learning and adaptive abilities of an individual. The measurement of intelligence is generally expressed in term

of Intelligence Quotient (I.Q). The I.Q. is measured by finding the ratio of a person's mental age (MA) to his or her chronological age (CA).

The most popular scales for measuring intelligence are the Stanford-Binet Scale and Wechsler Scales. However, scores obtainable from intelligence tests may be unstable because of changes in the testee's environmental conditions, which enhance or impair the normal functioning of his or her intellectual abilities.

Personality tests differ from intelligence tests in that they are essentially affective measures. Affective variables are those which describe a person's qualities, emotions, values and attitudes in relation to the feelings of others. Examples of affective measures include friendliness, neatness, attitude to colleagues, attitude to school work, attendance, self-concept, etc.

Personality variables can be measured in various ways. The questionnaire or rating scale is the most popular method of measuring them. Other methods that are also used include the observation technique, the interview, sociometric techniques and the use of anecdotal records. Each of these methods has its peculiar strength and weaknesses.

ACTIVITY II

1. In a tabular form, highlight the differences and similarities between (i) achievement and aptitude tests, and (ii) intelligence and personality tests.

TEST CONSTRUCTION AND ADMINISTRATION

Teachers and counsellors use psychological tests to arrive at various decisions about their students or clients. In order to ensure that information obtained from these tests do not mislead them in the process of decision making, it is expedient that the tests are properly constructed and administered under the right conditions. Several procedures are embarked upon to make test items to be of high quality so that they will have high validity and reliability.

The validity of a test refers to the extent to which the test measures what it supposes to measure. Types of validity include content, criterion-related, face and construct validity. On the other hand, the reliability of a test refers to the consistency of test scores over time and across examiners. Reliability could be determined through the use of test-retest method, parallel method and internal consistency method.

Other procedures for ensuring that our tests are of good quality include the determination of the level of difficulty and discrimination index of the test items as well as analyzing the distracters (for tests which contain multiple-choice items).

The level of difficulty of test items is determined by finding out the proportion of testees that got each item on the test correctly. An 'easy' item is that which a high proportion of testees got correctly while 'difficult' items are those which a high proportion of testees did not get correctly.

The discrimination index is the extent to which each item of a test differentiates between the superior ability students and inferior ability students. In a good test, the discrimination index is expected to be on the average i.e. that majority of students/testees that belong to the upper 25% will get it correctly while majority of those in the lower 25% are also expected to get it wrongly.

Distracter analyses are done by making a list of the options in the test and counting how many of the testees picked each option. Any distracter with low frequency or with the least frequency when compared with others is considered poor and should be replaced by a better option that is closer to the answer. On the other hand, if the number of testees that picked a distracter is high, it means the option is plausible and should be retained.

In developing tests the principles to be borne in mind include the following:

- (i) Test items must adequately cover the course content if the test is an achievement test. In the case of intelligence, aptitude and personality tests, the items must adequately reflect the constructs being measured.
- (ii) Avoid ambiguous or difficult language especially words with multiple interpretation.
- (iii) Teachers or test administrators should not give too long or too short time as the duration for the test.
- (iv) Marking guides/scheme or scoring procedure must be prepared along with the test items.
- (v) Questions should be prepared well in advance.
- (vi) More questions than needed should be prepared for each test.
- (vii) When multiple choice items are used, teachers must ensure that the distracters are plausible.
- (viii) In classroom testing teachers should plan their tests to cover all the six domains of learning especially the higher levels.
- (ix) In test scoring, all factors which might introduce bias into the scoring procedure should be controlled as much as possible.

ACTIVITY III

1. Discuss two major procedures for ensuring that psychological tests are of good quality.

SUMMARY

- This unit has summarized all the units in this module with the following points to be noted. The first part of the module gave a definition of psychological tests as

instruments which seek information that could help in the determination of the extent to which a particular trait, attribute or quality is present in an individual. It also identified common uses of psychological tests to include the measurement of differences in attitudes of people; classification and guidance; determination of individual differences in people and the measurement and understanding of affective variables.

- The second part discussed specific types of psychological tests such as achievement, aptitude, intelligence and personality tests while the third part dealt with basic principles to be borne in mind by test developer and administrators in ensuring that psychological tests are of good quality and are administered or scored under the right conditions.

ASSIGNMENT

1. Discuss two types of psychological tests and highlight specific conditions which may affect the validity and reliability of the tests.
2. Explain the steps to be taken to improve the quality of the tests identified in (1) above.

REFERENCES

- Anastasi, A. (1982). **Psychological Testing**. New York: Macmillan Publishing Co. Inc.
- Hopkins, K.D. (1998). **Educational and Psychological Measurement and Evaluation**. Boston: Allyn & Bacon.
- Mehrens, W.A. (1975). **Measurement and Evaluation in Education and Psychology**. New York: Holt, Rinehart and Winston.
- Ughamadu, K.A., Onwuegbu, O.C. & Osunde, A.U. (1991). **Measurement and Evaluation in Education**. Benin City: World of Books Publishers.