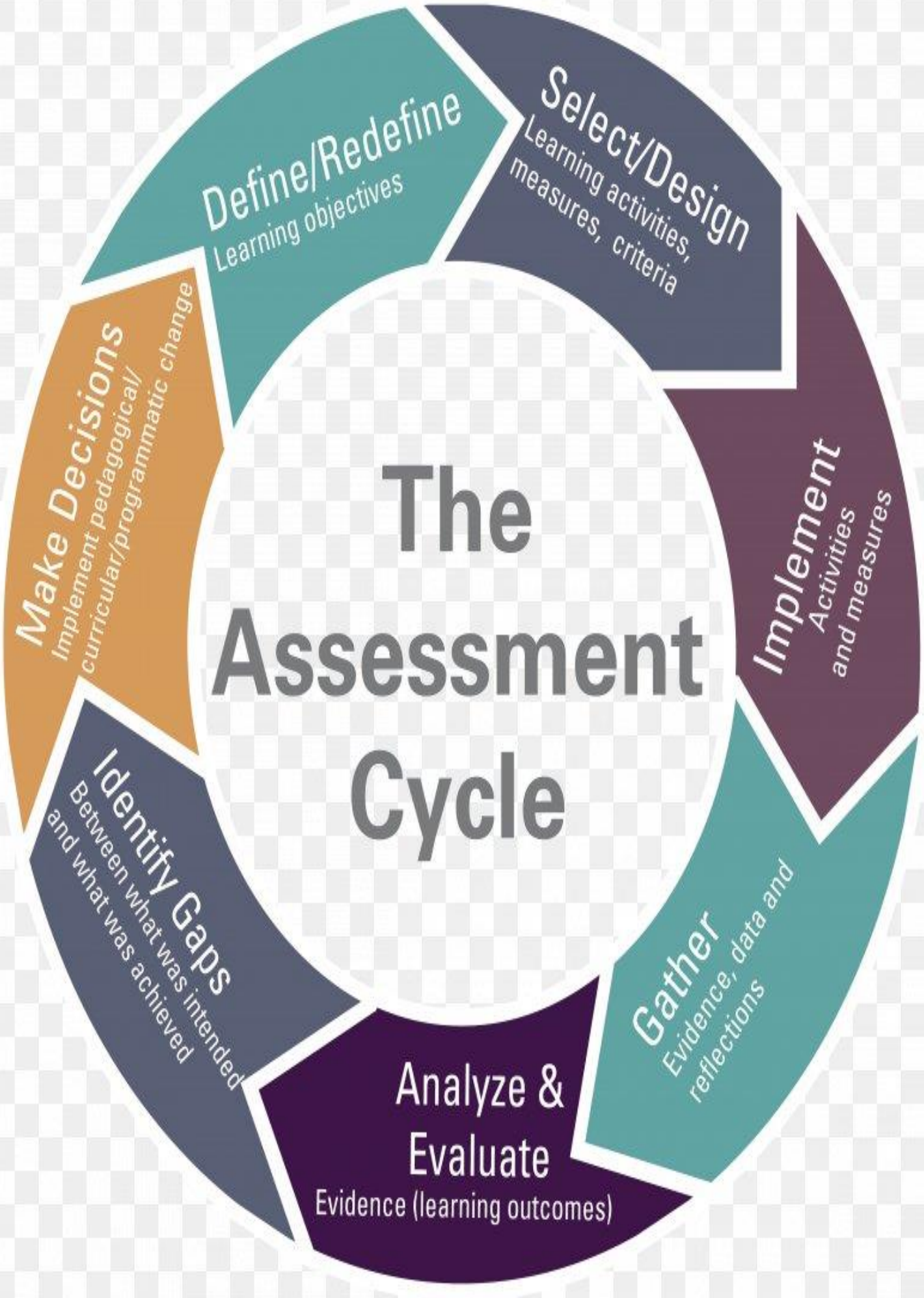

The Assessment Cycle
Define/Redefine
Learning objectives
Select/Design
Learning activities, measures, criteria
Implement
Activities and measures
Gather
Evidence, data and reflections
Analyze & Evaluate
Evidence (learning outcomes)
Identify Gaps
Between what was intended and what was achieved
Make Decisions
Implement pedagogical/ curricular/programmatic change

UNIVERSITY OF CAPE COAST

INSTITUTE OF EDUCATION

ASSESSMENT IN BASIC SCHOOLS
(EBS 234)

FOR COLLEGES OF EDUCATION (GHANA)

COMPILED BY;

OWUSU NATHANIEL ACHIAW

OF

GBEWAA COLLEGE OF EDUCATION
PUSIGA – BAWKU

PREFACE

Assessment is an integral part of teaching and learning. Teachers use informal and formal assessments on an ongoing basis to make decisions about their students, evaluate the success of their instruction, and to monitor classroom climate. The typical teacher spends about a third of his/her professional time engaged in assessment – related activities. Because classroom-based assessment is so critical to the instructional process, learning about assessment is essential to learning about teaching.

This course is mounted to equip year two (2) student teachers with basic skills to develop and evaluate formative and summative assessment of simple and complex student knowledge, beliefs and/or attitudes associated with classroom activities associated with a subject-matter domain and a particular target audience. Students will learn how to design assessments that are carefully aligned with educational objectives. The course will also include hands on activities to guide the creation, revision and use of quality assessment rubrics and coding schemes that work with the assessment of one's design.

TABLE OF CONTENTS

COURSE OUTLINE

PROGRAMME : 4 YEAR B.ED IN BASIC EDUCATION

COURSE TITLE : ASSESSMENT IN BASIC SCHOOLS

COURSE CODE : EBS 234

LEVEL : 200 **SEMESTER : TWO (2)**

CREDIT VALUE : 2

MODE OF DELIVERY OF THE COURSE

1. Face to Face
2. Practical Activity
3. Independent Study
4. e-learning opportunities

COURSE DESCRIPTION AND ASPECTS OF NTS AND NTECF TO BE ADDRESSED

The focus of the course is on the acquisition of the knowledge and skills needed to construct, administer, grade and interpret a variety of assessment instruments for classroom decisions making. Emphasis is on the objective and constructed response types of assessment instruments. The concepts of assessment for learning, test validity and reliability and the taxonomies of educational objectives are treated. (NTECF, NTS 3c, 3e, 3f, 3g, 3i, 3k).

COURSE CONTENT

UNIT/TOPIC	SUB-TOPIC	TEACHING AND LEARNING ACTIVITIES TO ACHIEVE LEARNING OUTCOME	DURATION
1. The Nature of Assessment	• Definition of concepts and Terms: Assessment, Test, Measurement, Evaluation • Continuous Assessment - Definition - Characteristics - Strengths and Weaknesses - Role of the classroom teacher	• Think –Pair-Share • Discussion to arrive at the meaning of concepts and terms in assessment. • Using individual and group presentations (being mindful	

	• Current trends in assessment in Ghana: - Profile Dimensions, - School-Based Assessment (SBA) - principles and purposes of Assessment - Assessment for Learning (AFL) - Assessment of Learning (AoL). - Assessment as Learning (AaL)	of gender roles) to present the concepts of continuous assessment and current trends in assessment.	
2. Goals and Learning Targets of Instruction	• Explanation of Terms: Goals, Outcomes, Objectives. • Importance of Learning Objectives for Assessment. • Taxonomies of Educational Objectives - Bloom's and Quell- malz's Cognitive domain. - Revised Bloom's taxonomy: Affective Domain; psychomotor Domain.	• Tutor-led discussion on meaning of the terms; goals, outcomes and objectives. Individual and group presentation on Taxonomies of educational objectives. • Questions and answers technique can also be employed where appropriate during the presentation.	
3. Characteristics of Tests	• Meaning, nature and principles of validity. • Categories of validity evidence • Factors affecting validity • Meaning of reliability concepts • Definition of Terms:	• Think-Pair-Share and running dictation to discuss and explain the meaning and nature of validity and reliability. • Using individual and group presentation	

	- Obtained scores - Error scores - True scores - Standard error of measurement • Methods of estimating reliability • Factors affecting reliability	(being mindful of gender roles) • Using questions and answers technique (being mindful of equity and inclusivity)	
4. Planning Classroom Tests and Assessment	• Characteristics of standardized Achievement Tests. • Teacher-made classroom Achievement Tests. • Stages in teacher made classroom achievement tests	• Teacher-led discussion on meaning, types and characteristics of achievement tests. • Power point presentation on comparing standardized and teacher-made achievement tests. • Questions and answers technique to be employed (being mindful of equity and inclusivity.	
5. Types of Tests Items	• Comparing Essay and Objective Tests • Objective-type tests - Description - Strengths and - Weaknesses • selection-type tests: - multiple choice; - true and False type; - matching type • supply-type tests: - short answer types;	• Teacher –led discussion on definition, types and characteristics of essay and object type tests; • Power point presentation on comparing essay and objective type tests. • Using individual and group presentations (being mindful of gender roles)	

	- fill in the blanks; - completion • essay-type tests and computational items - strengths, - weaknesses - constructing essay items - scoring essay tests - computational test items	on constructing the various item formats	
6. Assembling. Administering and Appraising Achievement Tests	• Guidelines for Assembling Achievement Tests. • Guidelines in Administering Achievement Tests • Appraising Achievement Tests	• Using individual and group presentations on assembling and appraising achievement tests. • Videos and whole class discussion can be used for presenting test administration practices.	
7. Interpretation of Test Scores	• Meaning of norm-referenced scores • Types of norm-referenced scores • Uses of norm-referenced scores • Meaning of criterion-referenced scores • Types of criterion referenced scores • Uses of criterion referenced scores	• Teacher-led discussion on the meaning, types and uses of norm-and criterion-referenced scores.	

UNIT 1: THE NATURE OF ASSESSMENT

Definition of Concepts and Terms

The general public uses the terms assessment, test, measurement and evaluation interchangeably, but is important for the student in assessment to distinguish among them. The meanings of the terms as applied to situations in schools are explained in the following paragraphs.

Assessment

The process of obtaining information that is used for making decisions about students, curricula and programs, and educational policy. It includes the full range of procedures used to gain information about student learning. The procedures may be formal (pencil and paper tests) or informal (observation). Certain concepts and terms are associated with assessment. These are tests, measurement and evaluation. Assessment can be said to be purposeful, systematic and on-going collection of information's as evidence for use in making judgment of students' learning and curriculum.

Test

A task or series of tasks, which are used to measure specific traits or attributes in people. In educational settings, tests include paper and pencil instruments, which contain questions that students and pupils respond to. The responses provided to the questions help the test giver to obtain an estimate of the specific trait being measured. It answers the question, 'How well does the individual perform?'

Test Interpretations

Two interpretations can be given to scores from tests. These are norm-referenced and criterion-referenced interpretations.

Norm-referenced interpretation: These describe test scores or performance in terms of a person's position in a reference group that has been administered the assessment. The reference group is called the norm group. For example, a student's performance may place him/her as the 15^{th} out of 45. Or a student doing better than 90% of the class.

Criterion-reference interpretation: These describe test scores or performance in terms of the kinds of tasks a person with a given score can do. The performance can be compared to a

pre-established standard or criterion. For example, a student may be able to solve 8 problems out of 10 concerning fractions. A level of performance can be established at 6.

- Tests are instruments used to 'measure' students' characteristics or traits
- In scoring a test, numbers are assigned to students based on the degree of correctness of the answers provided.
- A score is the end product of testing.
- You can test your students when you want to assess them.

Measurement

The process of assigning numbers to the attributes or traits possessed by persons, events or a set of objects according to specific rules, educational measurement is the assignment of numerals to such traits as achievement, aptitude and performance in such a way that the numbers describe a degree to which the person possesses the attribute or trait. It is limited to the quantitative descriptions of students. It answers the question, 'How much?'

Steps in Measurement

Measurement involves three main steps. These are:

1. Identifying and providing a clear definition of the attribute/trait to be measured.
2. Determining the set of procedures/operations by which the attribute is to be manifested.
3. Establishing a set of procedures/rules for quantifying the attribute/trait being measured.

Scales of measurement

Depending upon the traits, /attributes/characteristics and the way they are measured, four levels of measurement dubbed measurement scales are identified. These are:

1. nominal;
2. ordinal;
3. interval and
4. ratio scales.

1. Nominal scale (features)

- A nominal scale classifies/labels/categorizes people or objects into two or more groups e.g. males/females

- Whatever the classification, a person can only be in one category, and members of a given category have a common set of characteristics
- For identification purposes categories are numbered
- Numbers can be used to label the categories of a nominal variable but the numbers serve as markers, not as indicators of amount or quantity e.g. if you wanted to, you could mark the categories of the variable called e.g. Gender with 1 = male and 2= male
- Some examples of nominal level variables are the countries you were born in, college major, personality type, religious affiliation and halls of residence.

2. Ordinal scale (features)

- Ordinal scales enable one to make ordinal judgement i.e. judgement about rank order.
- An ordinal scale not only classifies subjects but also ranks them in terms of the degree to which they possess a particular characteristic/ attribute of interest
- An ordinal scale puts subjects in order from highest to lowest, from most to least
- Though ordinal scale does indicate that some subjects are higher or better than others, they do not indicate **how much** higher or better, i.e. intervals between the ranks are not equal.
- Any variable where the levels can be ranked (but you don't know if the distance between the levels is the same) is an ordinal variable

3. Interval scale (features)

- An interval scale has all the characteristics of both nominal and ordinal scales and in addition has equal intervals
- It does not possess an absolute zero point. The zero point is arbitrary and does not mean the absence of the characteristic/trait. Here is the idea of a lack of a true zero point. Zero degrees celcius does not mean no temperature at all; in a Fahrenheit scale, it is equal to the freezing point or 32 degrees. Zero degrees in this scale do not mean zero or no temperature; just as zero percent in a test does not mean zero knowledge on the behaviour or characteristic measured.
- Values can be added and subtracted to and from each other, but not multiplied or divided
- Interval scale or interval level of measurement has the characteristics of rank order and equal intervals, i.e. the distance between adjacent values is the same

- Some examples are Celcius temperature, Fahrenheit temperature and IQ scores

4. Ratio scale

- A ratio scale has all the advantages of the types of scales and in addition it has a meaningful true zero point
- Values can be added, subtracted, multiplied, and divided. 60 minutes can be said to be 3 times as long as 20 minutes.
- This is a scale with a true zero point
- It has all of the lower level characteristics (i.e. the key characteristics of each of the lower level scales) of equal intervals (interval scale), rank order (ordinal scale), and ability to mark a value with a name (nominal scale).
- Some examples of ratio scale measurements are number correct; weight; height; response time; Kelvin temperature and annual income.
- If your annual income is exactly zero Ghana cedi, then you earned no annual income at all. You can buy absolutely nothing with zero cedi. Zero means zero

Evaluation

Stufflebeam (1973) defined evaluation as "the process of delineating, obtaining and providing useful information for judging decision alternatives". It is the systematic investigation of the worth or merit of an object (a person, programme or a book). For example, a teacher may judge a student's writing as exceptionally good for his grade placement. Evaluation involves gathering information, which can be qualitative or quantitative or about the effectiveness or worth of what is being assessed. The main concern of evaluation in the classroom is to arrive at a judgment on the worth or effectiveness of teaching and learning.

Forms of Evaluation

Evaluation may either be *formative* or *summative.*

Formative evaluation is the process of judging the worth of teaching and learning constantly during the period of instruction. Formative evaluation of student's achievement means we are judging the quality of student's achievement while the student is still in the process of learning. We make formative evaluations of students so we can guide their next learning steps. It requires the gathering of detailed information on frequent occasions through such means as teacher

observations, classroom questions, home assignments and short tests or quizzes. The main purpose is to provide feedback to both the teacher and the learner about progress being made and not to grade the student.

Summative evaluation is the process of judging the worth of teaching and learning at the end of the period of instruction. It is judgmental in nature. It attempts to determine to what extent the broad objectives of teaching and learning have been attained. In other words, it is judgement about the quality of student's achievement after instructional or learning process is completed. Such evaluators tend to summarise strengths and weakness; they describe the extent to which a properly implemented programmes or procedures has attained its stated goals and objective. For example, end of term or end of programme examinations – BECE, WASSCE and UCC end of semester examinations.

CONTINUOUS ASSESSMENT

Definition

Ogunniyi (1984, p. 113) defined continuous assessment as 'a formative evaluation process concerned with finding out, in a systematic manner, the overall gains that a student has made in terms of knowledge, attitudes and skills after a given set of learning experiences'. The definition implies that a student's final grade after a programme of instruction is an aggregation of all the performances exhibited in the cognitive, affective and psychomotor domains during the duration of the course.

Characteristics

Six desirable characteristics are expected in a continuous assessment programme. It is expected to be cumulative, comprehensive, diagnostic, formative, guidance-oriented, and systematic (Ipaye, 1982. Ogunniyi, 1984; Ministry of Education Ghana, 1988).

1. **Continuous assessment is cumulative:**

The final grade awarded a student at the end of the term or year is an accumulation of all the attainments throughout the term or year. Any decision on the student is based on all the scores obtained in all measurements during the period under review.

A student for example might be declared to have attained the acceptable level of mastery in Mathematics to warrant a promotion. This means that the sum total of all his scores in class

assignments, homework, weekly tests, mid-term tests, end of term tests, class discussions and projects, reached the desired level of competence. The decision does not centre on only one score in an end of year examination. Any decision made on the student therefore considers all the previous decisions made about him.

2. Continuous assessment is comprehensive

Opportunities are provided for the assessment of the total personality of the student. This involves the assessment of tasks, activities and outcomes and demonstrated in the cognitive (knowledge), affective (attitude) and psychomotor (skills) domains. It must however be emphasized that these three domains must not always be included in the programme before the process is described as continuous assessment.

In addition, many types of evaluation procedures are used. These include teacher-made tests, classroom observations, class assignments and projects, oral questions, standardized tests, interviews and autobiographies. The score obtained from using all these procedures are combined to arrive at a final grade or classification of the student.

3. Continuous assessment is diagnostic

Continuous assessment involves a constant and continual monitoring of a student's performance and achievement. The process enables each student's strengths and weaknesses to be identified. It also enables the teacher to identify which students' have difficulties and problems and in what areas. With this knowledge, specific remedial actions are recommended and taken.

4. Continuous assessment is formative

Continuous assessment allows for immediate and constant feedback to be provided to the student on his performance. The student often with the help of the teacher and school counsellor, can analyse the feedback results. On the basis of the information derived for such an analysis, various strategies are adopted.

5. Continuous assessment is guidance-oriented

Guidance aims at helping the individual to accept his/her 'worth'. He/she identifies and accepts strengths and weaknesses. He/she works hard to consolidate the strong areas and improve upon the weak areas.

Continuous assessment aims at playing this guidance role. As the student is actively involved in the teaching-learning activities and tasks, his areas of weakness and strengths are easily

identified early and from time to time. The teacher then helps the student to strengthen further his strong areas and attempts to improve upon his weaknesses to attain the level of mastery needed. The student is thus directed and motivated in his learning.

Learning and taking tests is not an end in 'itself'. It is a step to achieve the total growth and development of the student. Continuous assessment provides the facts and figures and all the necessary school information needed to guide the student to achieve growth and development. It also provides the necessary information for the student to decide his future and his world of work.

6. **Continuous assessment is systematic**

Continuous assessment operates on a well-scheduled programme. Assessment is not spontaneous. A plan is desirable at the beginning of each year, term and week. There could be long term, medium term and short-term plans. The plan should spell out what measurements are to be made. This includes the types of traits and behaviours to be assessed and the procedure for assessing them. The procedures should indicate the types of class assignment and exercises, homework and projects, class tests and non-test techniques to be employed to collect the relevant data.

A decision needs to be taken on the dates and periods on which the various measurements will be made. It is also important to include in the plan the types of instruments to be used. These could be teacher-made tests, standardized tests; interview schedules, questionnaires, project sheets, observation schedules and checklists. Specific times should also be stated for the filling in the scores that students obtain on the appropriate forms.

Strengths

1. **Continuous assessment provides an excellent picture of a student's performance over a period of time.** In summative evaluation, a student's attainment in a course of studies for example is measured by a single shot examination. However, several influences like, malpractices, illnesses and inability to follow instructions influence a student's final score. The reliability of such scores is therefore doubtful. In continuous assessment, judgement on a student's performance is based on several other previous performances. This enables the effect of extraneous variables to be minimized. A more representative sample of his performance is arrived at.
2. **It enables the classroom teacher as well as the school administration to be actively and more meaningfully involved in the assessment of the students throughout the**

period of teaching and learning. The teacher is expected to be alert, diligent and consistent in assessing the various behaviours expected. He gives examples, assignments and tests; scores them and discusses results with the pupil. He also observes students behaviours in various domains and provides help where needed. The school's administration also provides the 'backup' services like the provision of stationary needed for class assignments, projects, tests and questionnaires.

3. **It enables the measurements of the three important domains in the taxonomy of education objectives viz cognitive, affective and psychomotor domains.** This is important because while the cognitive objectives are measured under test and examination conditions, affective and psychomotor abilities such as courtesy, sociability, creativeness, leadership and responsibility could only be measured over a reasonable one-shot examination system, the emphasis has always been on the cognitive abilities. This ultimately provides a biased outcome of the student educational attainment. The possibility of spreading the measurement net over the other areas makes the assessment of the students' exhibited behaviour more total.
4. **It helps to minimize the students' fears and anxieties about failure in the examinations.** The fear of performing poorly leads students to engage in examination malpractices such as copying and the exchange of answer scripts. Since the student is aware that several scores will be used to assess his final performance, tensions are often reduced. More desirable learning habits are developed. Rote learning is discouraged. Creativity and initiative are encouraged. A poor performance in one course is counter-balanced by an improved performance in another. Failure in one aspect of the course of study does not spell the doom of the student. The student has a great advantage here in that he has several opportunities to demonstrate the behaviours and objectives being measured.
5. **Continuous assessment encourages students to work assiduously throughout the period of teaching and learning.** The student becomes more alert in the class. He is punctual and attends classes regularly. This attitude comes about as a result of the fact that every stage of the instructional process is assessed and these counts towards the ultimate grade or score he would obtain. He knows that complacency, absenteeism, laziness and

malingering would prove disastrous to his goals in academic achievement and he therefore works hard.

6. **Education under the traditional summative evaluation system could be termed, "syllabus -ish" or syllabus-pulled".** This means that the emphasis is purely on what the syllabus prescribes as related to internal and external examinations. Any activity or task, which is not directly related to the syllabus for the examinations, is met with profound resistance from students. Continuous assessment discourages this attitude to a very large extent. Topics which are found relevant and interesting are included in the teaching and learning process. The aim here is that teaching should bring about a total growth in the individual and not only passing examinations.
7. **Constant feedback is given and this provides the groundwork for teachers to engage in diagnostic teaching.** Feedback enables the teacher to identify the weaknesses of individual students early and across tasks. He is then in a position to provide remedial and individualized teaching. This corrective action reduces frustration, disappointment and disillusionment on the part of students. The student is thus helped to progress. Continuous feedback guides the students to the most effective means of improving his performance.
8. **Record keeping is an important aspect of the teaching and learning process.** Records acknowledge 'the totality of what pupils have done in order to improve their motivation and help schools identify their needs more closely. It also provides a testimonial respected and valued by employers and colleges. Records also help to place students in appropriate stages when they transfer to another institution. Continuous assessment is a great instrument in the achievement of these goals. It provides the opportunity for the collection, preparation and keeping of up-to-date records on students. This data includes family and health data, academic record, interests and hobbies, work experience and special talents.
9. **Parents are provided with better and clearer pictures of their wards' performance and achievement in school over a period of time and learning experience.** The 'One Shot' traditional examination in most cases colours the actual performance of the student because of the variety of influences that affect the performance of a student. Due to repeated performance on various activities and tests in continuous assessment, the influence of these factors is greatly minimized. Parents thus receive more accurate

information on their wards and are put in a better position to plan more relevant programmes towards the future careers of their wards.

Weaknesses

Even though continuous assessment achieves much in terms of student and teacher evaluation of the instructional process and product, there are problems and weaknesses.

1. **Continuous assessment brings about an increase in the workload of teachers.** Since the process is systematic and comprehensive, the teacher is expected to be active in-designing and producing a variety of assessment instruments. In addition, he is expected to be scoring the class tests, assignments, and projects and at the same time taking observations. He is also expected to provide up-to-date records on each pupil and simultaneously be involved in remedial and individual teaching. Where classes are large in size (and in Ghana most classes are large,) the load becomes unbearable. The teachers then resort to unfair means in providing the requisite data for each pupil..
2. **To implement a continuous assessment programme, it is assumed that teachers have the requisite skills in test construction.** However, in Ghana, most Ghanaian teachers lack the skills required for constructing tests, because most initial teacher training programmes do not make provision for a course in testing. In cases where teachers underwent a course of instruction in testing and assessment, few teachers use their knowledge in test construction. The effect is that, since each teacher designs his own instrument, the testing instruments yield unreliable information. Standards are also bound to vary from teacher to teacher.
3. **In Ghana, one problem is the inadequacy of materials and equipment. Continuous assessment is costly in terms of materials.** Finance is needed for the procurement of material and equipment such as cumulative record cards, stationary for testing instrument, chairs and tables, and well-built classrooms. The sizes of the classes are such that a huge financial outlay is needed. The experience in the Ghanaian classrooms is that these equipment and materials are woefully inadequate. This situation puts great inhibition on the success of any continuous assessment programme.
4. **Continuous assessment, especially in the first and second cycle levels, means less dependence on an external examining body.** This implies that the uniformity that goes

with external written examinations in the form of standard test items and scoring, are reduced to some extent. The fate of the individual student lies more in the hands of the classroom teacher. This situation generates fears, doubts and apprehensions in the minds of the public about the degree of fairness in assessing the achievement of students. It also makes it difficult to compare the performances from different schools since there is less uniformity in the use of instruments and techniques in assessing the performance of students.

5. **In the first and second cycle institutions, certificates obtained are based on performances and achievement in external examination in Ghana.** This situation enables the certificates to have credibility, since efforts are made to maintain standards across years and test items. However, with the continuous assessment, if schools award certificates based on the attainments of their own students, standards will vary from school to school as well as certificates. The credibility of certificates becomes doubtful in most cases. To handle this problem, schools contributes 30% of the total scores of each student in a subject while an external examining body (WAEC) contributes 70%.
6. **Another problem is that of supervision. Continuous assessment requires co-operation and co-ordination at different levels.** Close supervision is needed at all levels. Unfortunately, supervisors in most cases who are heads of institutions are already laden with loads of work. They are therefore not effective in their supervisory roles.
7. **There is also an additional problem of record maintenance.** Continuous assessment requires the collection and storage of records. In most institutions, adequate storage facilities are not available. Current storage and retrieval facilities like steel cabinets, personal computers and word processors are lacking in institutions. Handling continuous assessment data is therefore extremely difficult.

Role of the Ghanaian teacher in continuous assessment

The Ministry of Education, as a matter of policy expects each teacher to:

1. Give class assignment/exercises fortnightly and record the scores of four of them with a maximum score of 10 each;
2. Conduct three class tests in a term with a subtotal of 40.
3. Give pupils at least four projects/homework in a term with a subtotal of 20.

The three assessments give a total score of 100, which is scaled down to 30% as the internal mark for each pupil. The end of term examination is given 70%.

At the end of the junior and senior secondary schools, all the scores a pupil obtains are scaled to 30% and forwarded to the WAEC where 70% is obtained for external assessment.

For the policy to be successful, teachers are expected to perform the following roles.

1. **The teacher must accept the philosophy of continuous assessment.** He must be convinced beyond all reasonable doubt that continuous assessment is a better form of assessing student's academic attainment than the traditional summative system.
2. **The teacher needs to be knowledgeable about continuous assessment.** He must know the characteristics of continuous assessment, the strengths and weaknesses of the system as a procedure for assessing students' knowledge, attitude and manipulative skills. He must clearly understand and accept his roles and responsibilities as outlined by the programme and be willing to contribute to its successful implementation.
3. **At the beginning of each academic year and term (or semester), the teacher must make a timetable for the assessments to be made.** He must set specific dates on which the class tests, assignments/exercises, project/homework tasks will be performed. He also needs to decide, what instruments to use in his assessment.
4. **The teacher must break the learning programme of the period of instruction into smaller, specific and well-defined units.** A level of mastery must be set for each unit.
5. **The teacher must assess the learning outcomes and performances at the end of each unit of instruction.** He must follow the timetable laid out for assessment but should allow some degree of flexibility.
6. **The teacher must spread the assessment over all areas of student's behaviour.** These are the cognitive, affective and psychomotor domains. Appropriate assessment instruments such as pencil and paper tests, observation autobiographies, questionnaires, checklist, rating scales and inventories should be used.
7. **The teacher must formulate measurable, specific and attainable instructional objectives for each unit of instruction.** This helps him to make his teaching more effective and meaningful. It also makes his assessments easier since these are based on the objectives set out.

8. **The teacher must provide constant feedback.** Class assignments and exercises, projects, tests and homework must be promptly scored and returned to the students. This helps to direct and motivate student learning.
9. **The teacher must record all the assessment of the student in all the areas of learning and instruction in the appropriate records.** This must be done promptly at the end of each measurement. The records must be well kept and maintained.
10. **The teacher must be involved in remedial and individualized teaching.** At the end of the teaching assessment the teacher should find out whether the whole class attained the required level of mastery. Remedial programmes should be organized if the requisite level of mastery was not reached. In addition, the teacher must also devote time to individuals who do not perform well in the class assignments, exercises and tests and have difficulties.
11. **The teacher must also engage in guidance and counseling.** He must identify the weaknesses and strengths of students in the various areas of learning. He should then use the information to guide and counsel the student for his full personal development and growth as well as preparing the student for his future career.
12. **The teacher must engage in constant evaluation of himself and of the continuous assessment programme.** The scores obtained from the various assessments should be used to measure his own performance and the effectiveness of his methods and techniques. He must also evaluate the success of the programme regularly to identify the lapses and improve upon them. This could be done weekly, monthly and at the term/semester.

CURRENT TRENDS IN ASSESSMENT IN GHANA

The form of assessment in the new educational structure and reform, being implemented from September 2008, is based on two concepts and procedures. These are:

1. Assessment is based on profile dimensions of each subject.
2. School-based assessment (SBA) replaces continuous assessment.

Assessment is based on profile dimensions of each subject.

Definition of Profile Dimensions

A ‘dimension’ is a psychological unit describing a particular learning behaviour. More than one dimension constitutes a profile of dimensions. A specific objective may be stated with

an action verb as follows. The pupil will be able to **describe**….. etc. Being able to "describe" something after the instruction has been completed means that the pupils has acquired "**knowledge**". Being able to explain, summarize, give examples, etc. means that the pupil has **understood** the lesson taught.

Similarly, being able to develop, plan, solve problems, construct, etc. means that the pupil can "apply" the knowledge acquired in some new context. Each of the specific objectives in each syllabus contains an "action verb" that describes the behaviours the pupils will be able to demonstrate after the instruction. "**Knowledge**", **Understanding**" and "**Application**", etc. are dimensions that should be prime focus of teaching and learning in schools. It has been realised unfortunately that schools still teach the low ability thinking skills of knowledge and understanding and ignore the higher ability thinking skills. Instruction in most cases has tended to stress knowledge acquisition to the detriment of the higher ability behaviours such as application, analysis, etc.

The persistence of this situation in the school system means that pupils will only do well on recall items and questions that require higher ability thinking skills such as application of mathematical principles and problem solving. For there to be any change in the quality of people who go through the school system, pupils should be encouraged to apply their knowledge, develop analytical thinking skills, develop plans, generate new and creative ideas and solutions, and use their knowledge in a variety of ways to solve problems while still in school.

Each action verb indicates the underlying profile dimension of each particular specific objective. Teachers are to read each objective carefully to know the profile dimension toward which you have to teach. The major profile dimensions are:

1. Knowledge and Understanding (KU)
2. Application of Knowledge. Use of Knowledge (AK/UK)
3. Attitudes and Values (AV)
4. Practical Skills (PS)

In development assessment procedures, select specific objectives in such a way that you will be able to assess a representative sample of the syllabus objectives. Each specific objective in the syllabus is considered a criterion to be achieved by the pupil.

When you develop a test that consists of items or questions that are based on a representative sample of the specific objectives taught, the test is referred to as a "Criterior-

Referenced Test". In many cases, a teacher cannot test all the objectives taught in a term, in a year etc. The assessment procedure you use i.e. class tests, homework, projects etc. must be developed in such a way that it will consist of a sample of the important objectives taught over a period.

School-based assessment (SBA) replaces continuous assessment

School Based Assessment

A new School Based Assessment system (SBA), formally referred to as Continuous Assessment is in use in Ghana as part of the new Educational Reforms starting September 2008. SBA is a very effective system for teaching and learning if carried out properly. The new SBA system is designed to provide schools with an internal assessment system that will help schools to achieve the following purposes:

1. Standardize the practice of internal school-based assessment in all schools in the country.
2. Provide reduced assessment tasks for each of the primary school subjects.
3. Provide teachers with guidelines for constructing assessment items/questions and other assessment tasks.
4. Introduce standards of achievement in each subject and in each class of the school system.
5. Provide guidance in marking and grading of test items/questions and other assessment tasks.
6. Introduce a system of moderation that will ensure accuracy and reliability of teachers' marks.
7. Provide teachers with advice on how to conduct remedial instruction on difficult areas of the syllabus to improve pupil performance.

The marks for SBA should together constitute the School Based Assessment component marked out of 60 percent. The emphasis is to improve students' learning by encouraging them to perform at a higher level. The SBA will hence consist of:

- End of month tests
- Homework assignments (specially designed for SBA
- Project

The SBA system will consist of 12 assessments a year instead of the 33 assessments in the previous continuous assessment system. This will mean a reduction by 64% of the workload compared to the previous continuous assessment system. The 12 assessments are labeled as Task

1, Task 2, Task 3 and Task 4. Task 1 – 4 will be administered in Term 1; Task 5-8 will be administered in Term 2, and Tasks 9 – 12 administered in Terms 3.

Task 1 will be administered as **an individual test** coming at the end of the first month of the term. The equivalent of Task 1 will be Task 5 and Task 9 to be administered in Term 2 and Term 3 respectively. Task 2 will be administered as a Group Exercise and will consist of two or three instructional objectives that the teacher considers difficult to teach and learn. The selected objectives could also be those objectives considered very important and which therefore need pupils to put in more practice. Task 2 will be administered at the end of the second month in the term. **Task 3** will also be administered as individual test under the supervision of the class teacher at the end of the 11^{th} or 12^{th} week of the term.

Task 4 (and also Task 8 and Task 12) will be a project to be undertaken throughout the term and submitted at the end of the term. Schools will be supplied with 9 projects topics divided into three topics for each term. A pupil is expected to select one project topic for each term. Projects for the second term will be undertaken by teams of pupils as Group Projects. Projects are intended to encourage pupils to apply knowledge and skills acquired in the term to write an analytic or investigative paper, write a poem (as may be required in English and Ghanaian Languages), use science and mathematics to solve a problem or produce a physical three-dimensional product as may be required in Creative Arts and in Natural Science.

Apart from the SBA, teachers are expected to use class exercises and home work as processes for continually evaluating pupils' class performance, and as a means for encouraging improvements in learning performance.

End-of-Term Examination

The end-of-term examination is a summative assessment system and should consist of a sample of the knowledge and skills pupils have acquired in the term. The end-of-term test for Term 3 should be composed of items/questions based on the specific objectives studies over the three terms, using a different weighting system such as to reflect the important of the work done in each term in appropriate proportions. For example, a teacher may build an end of term 3 test in such a way that it would consist of the 20% objectives studied in Term 1, 20% of the objectives studied in term 2, and 60% of the objective studied in Term 3.

Combining SBA marks and End-of-Term Examination Marks

The new SBA system is important for raising pupils' school performance. For this reason, the 60 marks for the SBA will be scaled to 50 in schools. The total marks for the end of term test will also be scaled to 50 before adding the SBA marks and end-of-term examination marks to determine pupils' end of term results. The SBA and end of term tests marks will hence be combined in equal proportions of 50:50. The equal proportions will affect only assessment in the school system. It will not affect the SBA mark proportion of 30% used by WAEC for determining examination results at the BECE.

Grading Procedure

To improve assessment and grading and also introduce uniformity in schools, it is recommended that schools adopt the following grade boundaries for assigning grades.

Grade A:	80-100%	-	excellent
Grade B:	70-79%	-	very Good
Grade C:	60-69%	-	good
Grade D:	45-59%	-	credit (satisfactory)
Grade E:	35-44%	-	pass
Grade F:	$\leq$ 34%	-	fail

The grading system presented above shows the letter grade system and equivalent grade boundaries. In assigning grades to pupils' test results, or any form of evaluation, you may apply the above grade boundaries and the descriptors. The descriptors excellent, Very Good etc) indicate the meaning of each grade. For instance, the grade boundary for "excellent" consists of scores between 80-100. Writing "80%" for instance, without writing the meaning of the grade, or the descriptor for the grade i.e. "Excellent" does not provide the pupil with enough information to evaluate his/her performance in the assessment. You therefore have to write the meaning of the grade alongside the score you write. Apart from the score and the grade descriptor, it will be important also to write a short diagnosis of the points the pupil should consider in order to do better in future tests etc. Comments such as the following may also be added to the grades:

Keep it up

Has improved

Could do better

Hardworking

Not serious in class

More room for improvement, etc

Note that the grade boundaries above are also referred to as grade cut-off scores. When you adopt a fixed cut-off score grading system as in this example, you are using the criterion-referenced grading system. By this system a pupil must make a specified score to earn the appropriate grade. This system of grading challenges pupils to study harder to earn the better grades. It is hence very useful for achievement testing and grading.

GENERAL PRINCIPLES OF ASSESSMENT

Principles are fundamental truths and doctrines accepted by most authorities as characteristics of assessment. Some of them are:

1. **Test developer must be clear about the learning target to be assessed.** This involves clearly specifying the intended learning goals and helps to select the appropriate assessment technique. Again, clear, accurate and timely information on assessment tasks and procedures should be available to students, staff and other external assessors or examiners. Thus, information should be explicit, accessible and transparent.
2. **The assessment technique selected must match the learning target.** The main criterion is whether the procedure is the most effective in measuring the learning target. Assessment tasks should primarily reflect the nature of the discipline or subject but should also ensure that students have the opportunity to develop a range of generic skills and capabilities.
3. **Assessment is inherently a process of professional judgement.** Proper use of assessment procedures requires that the user is aware of limitations of each technique. In interpreting the results of the assessment, these limitations must be considered. Therefore, all those involved in the assessment of students must be competent to undertake their roles and responsibilities.
4. **Assessment techniques must serve the needs of the learners.** They should provide meaningful feedback to the learners about how closely they have approximated the learning targets.
5. **Good assessments use multiple methods.** Multiple indicator of performance provides a better assessment of the extent to which a student has attained a given learning target. Assessment needs to be comprehensive. Formative and summative assessment should be

incorporated into the programmes to ensure that the purposes of assessment are adequately addressed. Many programmes may also wish to include diagnostic assessment.

6. **Assessment is a means to an end.** It is not an end in itself. Assessment influences student motivation and learning. The nature of assessment influences what is learned and the degree of meaningful engagement by students in the learning process. Students are therefore entitled to feedback on submitted formative assessment tasks and on summative tasks, where appropriate. The nature, extent and timing of feedback for each assessment task should be clear to students in advance.
7. **Assessment should be valid and reliable.** Evidence needs to be provided that the interpretations and use of students' assessment results are appropriate and reliable. There is the need for assessment to be reliable and this requires clear and consistent processes for setting, marking, grading and moderation of assignments/tests.
8. **Assessment of decision-making is influenced by a series of tensions.** Competing purposes, uses and pressures result in tension for teachers and administrators as they make assessment-related decisions. For example, good teaching is characterized by assessment that motivates and engages students in ways that are consistent with their philosophies of teaching and learning and with theories of development, learning and motivation. Most teachers want to use constructed-response assessment because they believe this kind of testing is best to ascertain student understanding. On the other hand, factors external to the classroom such as mandated large-scale testing, promote different assessment strategies such as using selected-responses tests and providing practice in objective test-taking (McMillan & M Nash, 2000).
9. **Good assessment is fair and ethical.** Usually, four views of fairness are presented by the Assessment Standards as
    i. absence of bias (e.g. offensiveness and unfair penalization)
    ii. Equitable treatment
    iii. Equality in outcomes and
    iv. Opportunity to learn.

In assessing students, the rights and responsibilities of test takers, testing individuals of diverse linguistic backgrounds and testing individuals with disabilities or special needs should be considered. Student's knowledge of learning targets and the nature of the assessment prior to

instruction (e.g. knowing what will be tested, how it will be graded, scoring criteria, anchors, exemplars and examples of performance) are also needed.

10. **Good assessment appropriately incorporates technology.** As technology advances and teachers become more proficient in the use of technology. There will be increased opportunities for teachers and administrators to use computer-based techniques (e.g. item banks, electronic grading, computer-adapted testing and computer-based simulation). Internet resources and more complex, detailed ways of reporting results. There is however, a danger that technology will contribute to the mindless use of new resources, such as using items on-line developed by some companies without adequate evidence of reliability, validity and fairness and crunching numbers with software programmes without sufficient thought about weighting, error and averaging.

PURPOSES OF ASSESSMENT

Educational assessment is conducted for a variety of reasons and the nature of the assessment often reflects the purpose for which it is being carried out.

Assessment provides information for decisions about students, curricula and programmes, and educational policy. These decisions are:

1. Instructional Management Decisions
2. Selection Decisions
3. Placement Decisions
4. Counseling And Guidance Decisions
5. Credentialing And Certification decisions
6. Instructional Management Decisions

Instructional Management Decisions

1. Assessment provides knowledge about the readiness of individuals (pupils, students) to learn a new set of curricular content.
2. Assessment enables the teachers to set realistic instructional goals and objectives for the class as well as individual pupils.
3. Assessment helps the teacher to discover the learning difficulties of the pupils and to provide remedial action. The diagnostic decision asks the question, ‘what learning

activities will best adapt to this student's individual requirements and thereby maximize the student's opportunities to attain the chosen learning target?

4. Assessment aids the teacher in the selection of the best instructional technique to adopt for the class and for each course.
5. Assessment helps in the evaluation of the degree to which objectives in the classroom are being achieved.
6. Assessment enables the teacher to determine the progress made by each individual student in learning.
7. Assessment serves as a source of motivation and directs and facilitates students' learning. It helps them to set goals.
8. Assessment provides feedback or knowledge of results to the students. This helps students to identify their own strengths and weaknesses as well as progress.
9. Assessment enables the teachers to assign grades to students which provide a record of their progress and achievement.

Selection decisions

Assessments provide information to select the right caliber of people for admission, promotion and awards of prizes. Those not acceptable are rejected.

Placement decisions

1. Assessment provides information to place students in courses and classes where they are likely to succeed in the future.
2. Assessments provide the basis for grouping individuals for instruction in view of their individual differences.

Counseling and Guidance decisions

1. Assessments aid in providing guidance and counseling in social and psychological adjustment problems that affect the pupil's performances in the classroom.
2. Assessments are used to assist students to explore and choose careers and in directing them to prepare for the careers they select.

Credentialing and Certification decisions

Assessments enable students to acquire certificates that are needed for employment in the world of work.

In all, purpose of assessment could be grouped under three main categories as assessment of learning, as learning and for learning.

1. **Assessment of Learning (AoL)**

Assessment of learning is the assessment that becomes public and results in statements or symbols about how well students are learning. It often contributes to pivotal decisions that will affect student's futures. It is important, then, that the underlying logic and measurement of assessment of learning be credible and defensible.

Assessment of learning (AoL) which is carried out at the end of instruction is done purposely for grading and reporting. It has well established guidelines including;

1. It is accompanied by a number or a letter grade (Summative)
2. It compares one student's achievement with standards.
3. The results can be communicated to the students and teachers.
4. It occurs at the end of the learning unit.

Teacher's roles in assessment of Learning

- A range of alternative mechanism for assessing the same outcomes should be provided
- Public and defensible reference point for making judgment should be available.
- Transparent approaches to interpretation must be made.
- Description of the assessment process should be clear.
- Strategies for recourse in the event of disagreement about the decision should be considered.

2. **Assessment as Learning (AaL)**

Through this process, students are able to learn about themselves as learners and become aware of how they learn-become metacognitive (knowledge of one's own thought processes) about their learning.

Students reflect on their work on a regular basis, usually through self and peer assessment and decide (often with the help of the teacher, particularly in the early stages) what their next

learning will be. Assessment, as learning helps students to take more responsibility for their own learning and monitoring future directions.

Teacher's roles in Assessment as Learning

- Guide students in developing internal feedback or self-monitoring mechanism to validate and question their own thinking and to become comfortable with ambiguity and uncertainty that is inevitable in learning anything new.
- Provide regular and challenging opportunities to practice, so that students can become confident, competent self-assessors.
- Monitor student's metacognitive processes as well as their learning and provide descriptive feedback.
- Create an environment where it is safe for students to take chances and where support is readily available.

3. Assessment for Learning (AfL)

Assessment for learning (AfL) is a recent term describing one of the important purposes of assessment. It is one of the powerful ways of improving learning and raising standards (Black and William, 1998). Assessment for learning (Afl) is the process seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there.

Black and William (1998) also define assessment for learning as "all those activities undertaken by teachers and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged".

Assessment for learning (Afl) focuses on the gap between where a learner is in his/her learning, and where he/she needs to be – the desired goal. This can be achieved through processes such as sharing with learners, effective questioning and feedback.

Assessment for Learning (Afl)

- Comprises two phases – initial or diagnostic assessment and formative assessment.
- Assessment can be based on a variety of information sources (e.g., portfolios, works in progress, teacher observation, and conversation).
- Verbal or written feedbacks to the student are primarily descriptive and emphasize strengths, identify challenges, and points to next steps.

- As teachers check on understanding they adjust their instruction to students on track.
- No grades or scores are given record-keeping is primarily anecdotal and descriptive.
- Occurs throughout the learning process, from the outset of the course of study to the time of summative assessment.

Assessment for learning (Afl) is distinguished from Assessment of learning (AoL) which is carried our purposely for grading and reporting. AoL involves decisions about the merit of student performance in relation to standards of performance. It is designed to measure student achievement and gauge what they have learned. AoL takes place at a point in time for the purpose of summarizing the current status of student achievement.

Assessment for Learning has well established guidelines including:

1. It is accompanied by a number or letter grade (summative).
2. It compares one student's achievement with standards.
3. The results can be communicated to the student and parents.
4. It occurs at the end of the learning unit.

The UK Assessment Reform Group (1999) identified the following seven key characteristics of assessment for learning.

1. It is embedded in a view of teaching and learning of which it is an essential part.
2. It involves sharing learning goals with learners.
3. It aims to help pupils to know and to recognize the standards for which they are aiming.
4. It involves pupils in self-assessment (and peer assessment).
5. It provides feedback that leads to pupils recognizing their next steps and how to take them.
6. It is underpinned by the confidence that every student can improve.
7. It involves both teacher and pupils reviewing and reflecting on assessment data.

The Assessment Reform Group in the UK in 2002 derived 10 principles for guidance in assessment for learning. These principles are explained below.

1. ***Assessment for learning should be part of effective planning of teaching and learning.***
   A teacher's planning should provide opportunities for both learner and teacher to obtain and use information about progress toward learning goals. It also has to be flexible to respond to initial and emerging ideas and skills. Planning should include strategies to ensure that learners understand the goals they are pursuing and criteria that will be applied

in assessing their work. How learners will receive feedback, how they will take part in assessing their learning and how they will be helped to make further progress should also be planned.

2. ***Assessment of learning should focus on how students learn.*** The process of learning has to be in the minds of both learner and teacher when assessment is planned and when the evidence is interpreted. Learners should become as aware of the 'how' of their learning as they are of the 'what'.
3. ***Assessment for learning should be recognised as central to classroom practice.*** Much of what teachers and learners do in classrooms can be described as assessment. That is, tasks and questions prompt learners to demonstrate their knowledge, understanding and skills. What learners say and do is then observed and interpreted, and judgments are made about how learning can be improved. These assessment processes are an essential part of everyday classroom practice and involve both teachers and learners in reflection, dialogue and decision making.
4. ***Assessment for learning should be regarded as a key professional skill for teachers.*** Teachers require the professional knowledge and skills to: plan for assessment; observe learning; analyse and interpret evidence of learning; give feedback to learners and support learners in self-assessment. Teachers should be supported in developing these skills through initial and continuing professional development.
5. ***Assessment for learning should be sensitive and constructive because any assessment has an emotional impact.*** Teachers should be aware of the impact that comments, marks and grades can have on learners' confidence and enthusiasm and should be as constructive as possible in the feedback that they give. Comments that focus on the work rather than the person are more constructive for both learning and motivation.
6. ***Assessment for learning should take account of the importance of learner motivation.*** Assessment that encourages learning fosters motivation by emphasizing progress and achievement rather that failure. Comparison with others who have been more successful is unlikely to motivate learners. It can also lead to their withdrawing from the learning process in areas where they have been made to feel they are 'no good'. Motivation can be preserved and enhanced by assessment methods which protect the learner's autonomy, provide some choice and constructive feedback, and create opportunity for self-direction.

7. ***Assessment for learning should promote commitment to learning goals and a shared understanding of the criteria by which they are assessed.*** For effective learning to take place learners need to understand what it is they are trying to achieve - and want to achieve it. Understanding and commitment follows when learners have some part in deciding goals and identifying criteria for assessing progress. Communicating assessment criteria involves discussing them with learners using terms that they can understand, providing examples of how the criteria can be met in practice and engaging learners in peer and self-assessment.
8. ***Learners should receive constructive guidance about how to improve*** Learners need information and guidance in order to plan the next steps in their learning. Teachers should:

- Pinpoint the learner's strengths and advice on how to develop them.
- Be clear and constructive about any weaknesses and how they might be addressed
- Provide opportunities for learners to improve upon their work.

9. ***Assessment for learning develops learner's capacity for self-assessment so that they can become reflective and self-managing.*** Independent learners have the ability to seek out and gain new skills, new knowledge and new understanding. They are able to engage in self-reflection and identify the next steps in their learning. Teachers should equip learners with the desire and the capacity to take charge of their learning through developing the skills of self-assessment.
10. ***Assessment of learning should recognize the full range of achievement of all learners.*** Assessment for learning should be used to enhance all learners' opportunities to learn in all areas of educational activity. It should enable all learners to achieve their best and to have their efforts recognised.

Assessment for learning (AfL) therefore results in assessments that:

- ✓ Encourage, not discourage;
- ✓ Build confidence, not anxiety;
- ✓ Bring hope, not hopelessness;
- ✓ Offer success, not frustration

UNIT 2: GOALS AND LEARNING TARGETS OF INSTRUCTION

Definition of terms

***Instructional Objective*:** A stated desirable outcome of education or an intended learning outcome in terms of the types of performance students are able to demonstrate at the end of instruction to show that they have learned what was expected of them. By the end of the lesson, students should be able to define the term, taxonomy.

***Behavioural objectives*:** A statement that specifies what observable performance the learner should be engaged in when the achievement of the objective is evaluated. Behavioural objectives require action verbs such as discuss, write, read, state.

***Learning objectives*:** These specify what the teacher likes the students to do, value, or feel at the completion of an instructional segment.

Importance of learning objectives (targets) for classroom assessment

1. Learning objectives make the general planning for an assessment procedure easier through the knowledge of specific outcomes.
2. The selection, designing and construction of assessment instruments depend on knowing which specific outcome should be assessed.
3. Evaluating an existing assessment instrument becomes easier when specific outcomes are known.
4. They help to judge the content relevance of an assessment procedure. Specific learning outcomes provide information for the judgment.

TAXONOMIES OF EDUCATIONAL OBJECTIVES

Taxonomies are hierarchical schemes for classifying learning objectives into various levels of complexity. There are three main domains of educational objectives. These are:

1) cognitive,
2) affective,
3) psychomotor

Cognitive domain objectives produce outcomes that focus on knowledge and abilities requiring memory, thinking, and reasoning processes.

Affective domain objectives produce outcomes that focus on feelings, interests, attitudes, dispositions and emotional states.

Psychomotor domain objectives produce outcomes that focus on motor skills and perceptual processes.

The Cognitive domain

This domain was developed by Benjamin Bloom in 1956, hence it is known as Bloom's taxonomy of educational objectives. The taxonomy classifies educational objectives into 6 main headings.

i. **Knowledge.** This involves the recall of specific facts, methods and processes. They include those objectives, which deal with the recall or recognition of knowledge and the development of intellectual abilities and skills. It is often defined as the remembering of previously learned material. Illustrative verbs include define, identify, label.

ii. **Comprehension.** It is the ability to grasp the meaning of material. It is shown by translating material from one form to another, or by interpreting material (explaining or summarizing). Illustrative verbs include convert, explain, summarize

iii. **Application.** This refers to the ability to use learned material in new and concrete situations. This includes the application of such things as rules, methods, concepts, principles etc. Illustrative verbs include change, compute, and prepare.

iv. **Analysis.** This is the ability to break down material into its component parts so that its organizational structure may be understood. This includes the identification of parts, analysis of the relationships between parts etc. Illustrative verbs include break down, differentiate, illustrate

v. **Synthesis.** This refers to the ability to put parts together to form a new whole. This may involve the production of a unique communication, or a plan of operations. Illustrative verbs include categorize, combine, organize

vi. **Evaluation.** This is the ability to judge the value of material (e.g. novel, poem, and research report) for a given purpose. The judgments are based on definite criteria. Illustrative verbs include appraise, contrast, support.

Revised Bloom's Taxonomy

1. **Remember** Retrieve relevant knowledge from long-term memory.
   - ✓ Recognizing (identifying)
   - ✓ Recalling (retrieving)
2. **Understand** Construct meaning from instructional messages, including oral, written, and graphic communication
   - ✓ Interpreting (clarifying, paraphrasing, representing, translating)
   - ✓ Exemplifying (illustrating, instantiating)
   - ✓ Classifying (categorizing, subsuming)
   - ✓ Summarizing (abstracting, generalizing)
   - ✓ Inferring (concluding, extrapolating, interpolating, predicting)
   - ✓ Comparing (contrasting, mapping, matching)
   - ✓ Explaining (constructing models)
3. **Apply** Carry out or use a procedure in a given situation.
   - ✓ Executing (carrying out)
   - ✓ Implementing (using)
4. **Analyze** Break material into its constituent parts and determine how the parts relate to one another and to an overall structure or purpose
   - ✓ Differentiating (discriminating, distinguishing, focusing, selecting)
   - ✓ Organizing (finding coherence, integrating, outlining, parsing, structuring)
   - ✓ Attributing (deconstructing)
5. **Evaluate** Make judgments based on criteria and standards
   - ✓ Checking (coordinating, detecting, monitoring, testing
   - ✓ Critiquing (judging)
6. **Create** Put elements together to form a coherent or functional whole; reorganize elements into a new pattern or structure.

- ✓ Generating (hypothesizing)
- ✓ Planning (designing)
- ✓ Producing (constructing)

Quellmalz (1985) also proposed a cognitive taxonomy, which has five categories. These are:

i. **Recall.** This requires that students recognize or remember key facts, definitions, concepts, rules, and principles. They require students to repeat verbatim or to paraphrase given information. e.g. Who wrote the story?

ii. **Analysis.** Students divide a whole into component elements, e.g. What are the different parts of the story?

iii. **Comparison.** This requires students to recognize or explain similarities and differences. e.g. How was this story like the last one?

iv. **Inference.** Students are given a generalization and are required to recognize evidence or details and are required to come up with the generalization. e.g. What might be a good title for the story?

v. **Evaluation.** Students are required to judge quality, credibility, worth, or practicality. e.g. Is this a good story?

The Affective domain

This was developed by David Krathwohl, Benjamin Bloom and Masia in 1964. They classified educational objectives in the affective domain into 5 categories.

i. **Receiving:** It is the lowest level of learning outcomes in the affective domain. It is the willingness of a student/pupil to attend to particular phenomena/stimuli (e.g. classroom activities, reading textbook or library books, doing class assignments etc.). Examples of general instructional objectives include; listen attentively and attends closely to the classroom activities. Illustrative verbs that are used include asks, chooses, follows, gives, holds, names.

ii. **Responding:** It is the active participation of a student/pupil in given activities. The student/pupil does not only attend to particular stimuli but also reacts to it in some way.

The student/pupil may read an assigned material or does an assignment or project. Examples of general instructional objectives include; completes assigned homework, obeys school rules and regulations. Illustrative verbs that are used include answers, assists, complies, conforms, discusses, greets, practices, writes.

iii. **Valuing:** It is concerned with the worth or value a student/pupil attaches to a particular object, phenomenon or behaviour. The value ranges from a simple acceptance of a value to a more complex level of commitment. Examples of general instructional objectives include; shows concern for the welfare of others, appreciates the role of science in everyday life. Illustrative verbs include completes, describes, differentiates, explains, follows, initiates, invites, joins, reads.

iv. **Organization:** It is the ability to bring together different values, resolving conflicts between them, and beginning to build and internally consistent value system. Students/pupils begin to develop philosophies of life. Examples of general instructional objectives include; accepts responsibility for own behaviour, understands and accepts own strengths and weaknesses. Illustrative verbs include adheres, alters, arranges, combines, compares, completes, defends.

v. **Characterization by a value or value complex:** This is the highest level in the affective domain. At this level, the individual student/pupil has a value system that has controlled his/her behaviour for a sufficiently long time for him/her to have developed a characteristic lifestyle. Examples of general instructional objectives include; practices cooperation in group activities, maintains good study habits. Illustrative verbs include acts, discriminate, displays, influences, listens, modifies, performs, practices, proposes, qualifies.

The Psychomotor domain

Simpson (1972) and Harrow (1972) developed categories in this domain. Simpson produced 7 categories while Harrow had 6 categories.

Simpson's categories

1. **Perception.** This is the lowest level. It is the ability to use the sense organs to obtain cues that guide motor activity. For example, relating the sound of drums to the dance type. Illustrative verbs include, choose, describe, detect, identify.
2. **Set.** It is the readiness to take a particular type of action. Demonstrating a proper position to save a penalty kick in a soccer game. Illustrative verbs include, begin, displays, explains, shows, starts.
3. **Guided response.** It involves the early stages in learning a complex skill. For example, starting a car while beginning to learn how to drive. Illustrative verbs include, assemble, build, construct, display.
4. **Mechanism.** This occurs when a learned activity has become habitual and movements are performed with confidence and proficiency. For example, typing, operating a video recorder. Illustrative verbs include sketch, fix, fasten, dissect, assemble.
5. **Complex Overt Response**. It is the ability to perform complex acts. For example, driving an articulator truck, performing skilfully on the piano. Illustrative verbs include assemble, build, construct, organize.
6. **Adaptation.** It is the ability to modify movement patterns from well-developed skills to fit special requirements or situations. For example, modify piano rhythms to suit local songs. Illustrative verbs include adapt, alter, change, reorganize.
7. **Origination.** This is the highest level. It involves the ability to create new movement patterns to meet a specific need or particular problem. Creativity and originality are emphasized. For example, design a new computer software, create a new musical dance. Illustrative verbs include arrange, create, design, originate.

Harrow's (1972) categories

1. **Reflex movements** are actions elicited without learning in response to some stimuli. Examples include: flexion, extension, stretch, postural adjustments.

2. **Basic fundamental movements** are inherent movement patterns which are formed by combining of reflex movements and are the basis for complex skilled movements. Examples are: walking, running, pushing, twisting, gripping, grasping, manipulating.

3. **Perceptual abilities** refers to interpretation of various stimuli that enable one to make adjustments to the environment. Visual, auditory, kinesthetic, or tactile discrimination. Suggests cognitive as well as psychomotor behavior. Examples include: coordinated movements such as jumping rope, punting, or catching.

4. **Physical activities** require endurance, strength, vigor, and agility which produces a sound, efficiently functioning body. Examples are: all activities which require a) strenuous effort for long periods of time; b) muscular exertion; c) a quick, wide range of motion at the hip joints; and d) quick, precise movements.

5. **Skilled movements** are the result of the acquisition of a degree of efficiency when performing a complex task. Examples are: all skilled activities obvious in sports, recreation, and dance.

6. **Non-discursive communication** is communication through bodily movements ranging from facial expressions through sophisticated choreographies. Examples include: body postures, gestures, and facial expressions efficiently executed in skilled dance movement and choreographies.

UNIT 3: CHARACTERISTICS OF TESTS

TEST VALIDITY AND RELIABILITY

TEST VALIDITY

Nitko (1996, p. 36) defined validity as the "soundness of the interpretations and use of students' assessment results". Validity emphasizes the interpretations and use of the results and not the test instrument. Evidence needs to be provided that the interpretations and use are appropriate.

Nature of validity:

In using the term validity in relation to testing and assessment, five points have to be borne in mind.

1. Validity refers to the appropriateness of the interpretations of the results of an assessment procedure for a group of individuals. It does not refer to the procedure of instrument itself.
2. Validity is a matter of degree. Results have different degrees of validity for different purposes and for different situations. Assessment results may have high, moderate or low validity.
3. Validity is always specific to some particular use or interpretation. No assessment is valid for all purposes.
4. Validity is a unitary concept that is based on various kinds of evidence.
5. Validity involves an overall evaluative judgment. Several types of validity evidence should be studied and combined.

Principles for validation

There are four principles that help a test user/giver to decide the degree to which his/her assessment results are valid.

1. The interpretations (meanings) given to students' assessment results are valid only to the degree that evidence can be produced to support their appropriateness.

2. The uses made of assessment results are valid only to the degree that evidence can be produced to support their appropriateness and correctness.
3. The interpretations and uses of assessment results are valid only when the educational and social values implied by them are appropriate.
4. The interpretations and uses made of assessment results are valid only when the consequences of these interpretations and uses are consistent with appropriate values.

Categories of validity evidence

There are 3 major categories of validity evidence.

1. Content-related evidence

This type of evidence refers to the content representativeness and relevance of tasks or items on an instrument. Judgments of content representativeness focus on whether the assessment tasks are a representative sample from a larger domain of performance. Judgments of content relevance focus on whether the assessment tasks are included in the test user's domain definition when standardized tests are used.

Content-related evidence answers questions like:

i. How well do the assessment tasks represent the domain of important content?
ii. How well do the assessment tasks represent the curriculum as defined?
iii. How well do the assessment tasks reflect current thinking about what should be taught and assessed?
iv. Are the assessment tasks worthy of being learned?

To obtain answers for the questions, a description of the curriculum and content to be learned (or learned) is obtained. Each assessment task is checked to see if it matches important content and learning outcomes. Each assessment task is rated for its relevance, importance, accuracy and meaningfulness.

One way to ascertain content-related validity is to inspect the table of specification which is a two-way chart showing the content coverage and the instructional objectives to be measured.

2. Criterion-related evidence

This type of evidence pertains to the empirical technique of studying the relationship between the test scores or some other measures (predictors) and some independent external measures (criteria) such as intelligence test scores and university grade point average. Criterion-related evidence answers the question, How well the results of an assessment can be used to infer or predict an individual's standing on one or more outcomes other than the assessment procedure itself. The outcome is called the criterion.

There are two types of criterion-related evidence. These are concurrent validity and predictive validity.

Concurrent validity evidence refers to the extent to which individuals' current status on a criterion can be estimated from their current performance on an assessment instrument. For concurrent validity, data are collected at approximately the same time and the purpose is to substitute the assessment result for the score of a related variable. e.g. a test of swimming ability vrs swimming itself to be scored.

Predictive validity evidence refers to extent to which individuals' future performance on a criterion can be predicted from their prior performance on an assessment instrument. For predictive validity, data are collected at different times. Scores on the predictor variable are collected prior to the scores on the criterion variable. The purpose is to predict the future performance of a criterion variable. e.g. Using WASSCE results to predict the first year GPA in the University of Cape Coast.

Criterion-related validation is done by computing the coefficient of correlation between the assessment result and the criterion. The correlation coefficient is a statistical index that quantifies the degree of relationship between the scores from one assessment and the scores from another. This coefficient is often called the validity coefficient and takes values from

–1.0 to +1.0.

Expectancy tables can also be used for validation. An expectancy table is a two-way table that allows one to say how likely it is for a person with a specific assessment result to attain each criterion score level.

An example of an expectancy table

Predictor Test Score	Percent of pupils receiving each grade					
	F	D	C	B	A	Totals
90-99			20	60	20	100
80-89		8	33	42	17	100
70-79		20	33	40	7	100
60-69		22	44	28	6	100
50-59	6	28	44	22		100
40-49	7	40	33	20		100
30-39	17	42	33	8		100
20-29	25	50	25			100
10-19	100					100

Determine the degree of success by using a grade e.g. C or better. A question will be, What is the probability that a person with a score of 65 will succeed in this course (i.e. obtaining grade C or better)? The score of 65 lies in the 60-69 class and for this class, 78% (44+28 +6) are successful, so the person has a 78% chance of success.

3. Construct-related evidence:

This type of evidence refers to how well the assessment results can be interpreted as reflecting an individuals' status regarding an educational or psychological trait, attribute or mental process. Examples of constructs are mathematical reasoning, reading comprehension, creativity, honesty and sociability.

Methods of construct validation

i. **Define the domain or tasks to be measured**. Specifications must be very well defined so that the meaning of the construct is clear. Expert judgment is then used to judge the extent to which the assessment provides a relevant and representative measure of the task domain.

ii. **Analyze mental process required by the assessment tasks**. Examine the assessment tasks or administer the tasks to individual students and have them "think aloud" as they perform the tasks.

iii. **Compare the scores of known groups**. Groups that are expected to differ on the construct, e.g. by age, or training may both be given the same assessment tasks.

iv. **Correlate the assessment scores with other measures**. Similar constructs are expected to produce high correlation. E.g. two assessments on creativity are expected to produce high correlation.

Factors affecting validity

1. **Unclear directions.** Validity is reduced if students do not clearly understand how to respond to the items and how to record the responses or the amount of time available.
2. **Too difficult reading vocabulary and sentence structure tends to reduce validity**. The assessment may be measuring reading comprehension which is not to be measured.
3. **Ambiguous statements in assessment tasks and items.** This confuses students and makes way for different interpretations thus reducing validity.
4. **Inadequate time limits.** This does not provide students with enough time to respond and thus may perform below their level of achievement. This reduces validity.
5. **Inappropriate level of difficulty of the test items.** Items that are too easy or too difficult does not provide high validity.
6. **Poorly constructed test items.** These items may provide unintentional clues which may cause students to perform above their actual level of achievement. This lowers validity.
7. **Test items being inappropriate for the outcomes being measured lowers validity.**
8. **Test being too short.** If a test is too short, it does not provide a representative sample of the performance being interested in and this lowers validity.
9. **Improper arrangement of items.** Placing difficult items in the beginning of the test may put some students off and cause them to become unstable thereby performing below their level of performance thus reducing validity.
10. **Identifiable pattern of answers.** Placing the answers to tests like multiple-choice and true/false types enables students to guess the correct answers more easily and this lowers validity.

11. **Cheating.** When students cheat by copying answers or helping their friends with answers to test items, validity is reduced.
12. **Unreliable scoring.** Scoring of test items especially essay tests may lower reliability if they are not scored reliably.
13. **Student emotional disturbances.** These interfere with their performance thus reducing validity.
14. **Fear of the assessment situation.** Students can be frightened by the assessment situation and are unable to perform normally. This reduces their actual level of performance and consequently, lowers validity.

TEST RELIABILITY

Definition

Reliability is the degree of consistency of assessment results. It is the degree to which assessment results are the same when (1) the same tasks are completed on two different occasions (2) different but equivalent tasks are completed on the same or different occasions, and (3) two or more raters mark performance on the same tasks.

Points to note when applying the concept of reliability to testing and assessment.

i. Reliability refers to the results obtained with an assessment instrument and not to the instrument itself.
ii. An estimate of reliability refers to a particular type of consistency.
iii. Reliability is a necessary condition but not a sufficient condition for validity.
iv. Reliability is primarily statistical. It is determined by the reliability coefficient, which is defined as a correlation coefficient that indicates the degree of relationship between two sets of scores intended to be measures of the same characteristic. It ranges from 0.0 - 1.0

Definition of terms:

***Obtained (Observed) score*:** Actual scores obtained in a test or assessment.

***Error score*:** The amount of error in an obtained score.

***True score*:** The difference between the obtained and the error scores. It is the portion of the observed score that is not affected by random error. An estimate of the true score of a student is the mean score obtained after repeated assessments under the same conditions.

X = T + E

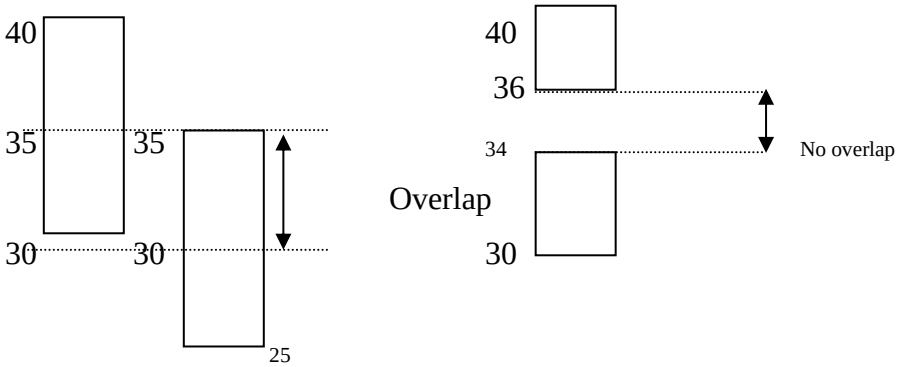
Reliability can be defined theoretically as the ratio of the true score variance to the observed score variance. i.e. rxx = $\frac{s_t^2}{s_x^2}$

***Standard error of measurement*:** It is a measure of the variation within individuals on a test. It is an estimate of the standard deviation of the errors of measurement. It is obtained by the formula: Se = Sx$\sqrt{1-r_{xx}}$ or SEM = SDx$\sqrt{\sqrt{1-reliabilitycoefficient}}$, where Sx or SDx is the standard deviation of the obtained scores. For example, given that, rxx = 0.8, Sx = 4.0,

SEM = 4$\sqrt{1-0.8}$ = 4$\sqrt{0.2}$ = 4 x 0.447 = 1.788

Interpreting standard errors of measurement

1. **It estimates the amount that a student is likely to deviate from her/his true score.** e.g. SEM=4 indicates that a student's obtained scores lies 4 points above or below the true score. An obtained score of 75 means the true score is either 71 or 79. The true score therefore lies between 71 and 79. 71-79 therefore provides a confidence band for interpreting an obtained score. A small standard error of measurement indicates high reliability providing greater confidence that the obtained score is near the true score.
2. **In interpreting the scores of two students, if the ends of the bands do overlap as in Example 1, then there is no real difference between the two scores.** However, if the two bands do not overlap as in Example 2, there is a real difference between the scores.



Example 1. SEM = 5	Example 2. SEM = 2
Suppose Grace had 34 and Fiifi 32 in a quiz. There is overlapping.	Suppose George had 38 and Aku 32 in a quiz. There is no overlapping.

***Reliability coefficient*:** A correlation coefficient that indicates the degree of relationship between two sets of scores intended to be measures of the same characteristic (e.g. correlation between scores assigned by two different raters or scores obtained from administration of two forms of a test)

Methods of estimating reliability

1. **Test-retest method.** This is a measure of the stability of scores over a period of time. The same test is given to a group of students twice within an interval ranging from several minutes to years. The scores on the two administrations are correlated and the result is the estimate of the reliability of the test. The interval should be reasonable, not be too short nor too long.
2. **Equivalent forms method.** Two test forms, which are alternate or parallel with the same content and level of difficulty for each item, are administered to the same group of students. The forms may be given on the same or nearly the same occasion or a time interval will elapse before the second form is given. The scores on the two administrations are correlated and the result is the estimate of the reliability of the test.
3. **Split-half method.** This is a measure of internal consistency. A single test is given to the students. The test is then divided into two halves for scoring. The two scores for each student are correlated to obtain the estimate of reliability. The test can be split into two halves in several ways. These include using (i) odd-even numbered items, and (ii) first half-second half. The Spearman-Brown prophecy formula is often used to obtain the reliability coefficient. This is given by:

$$r_{yy}\left(\text{Whole test reliability}\right) = \frac{2 \text{ x correlation between half test scores}}{1 + \text{correlastion between half test scores}}$$

Suppose correlation between half test scores was 0.75.

$$r_{yy} = \frac{2 \text{ x } 0.75}{1 + 0.75} = \frac{1.50}{1.75} = 0.86$$

4. **<u>Kuder-Richardson method</u>.** This is also a measure of internal consistency. A single administration of the test is used. Kuder-Richardson Formulas 20 and 21 (KR20 & KR21) are used mostly for dichotomously scored items (i.e. right or wrong). KR20 can be generalized to more than one-correct response items (e.g. attitude scales ranging from 5 to 1 on a 5-point scale). Such estimates are called Coefficient Alpha.

$$KR21 = \frac{n}{n-1}\left[1 - \frac{\bar{x}\left[n - \bar{x}\right]}{ns_x^2}\right]$$

5. **<u>Inter-rater reliability</u>**. Two raters, each score a student's paper. The two scores for all the students are correlated. This estimate of reliability is called scorer reliability or inter-rater reliability. It is an index of the extent to which the raters were consistent in rating the same students.

Factors influencing reliability

1. **Test length.** Longer tests give more reliable scores. A test consisting of 40 items will give a more reliable score than a test consisting of 25 items. Wherever practicable, give use more items.
2. **Group variability.** The more heterogeneous the group, the higher the reliability. The narrower the range of a group's ability, the lower the reliability. Differentiate among students. Use items that differentiate the best students from the less able students.
3. **Difficulty of items.** Too difficult or too easy items produce little variation in the test scores. This in turn lowers reliability. The difficulty of the assessment tasks should be matched to the ability level of the students.
4. **Scoring objectivity.** Subjectively scored items result in lower variability. More objectively scored assessment results are more reliable. For subjectively-scored items, multiple markers are preferred.

5. **Speed.** Tests, where most students do not complete the items due to inadequate allocation of time, result in lower reliability. Sufficient time should be provided to students to respond to the items.
6. **Sole marking.** Using multiple markers improves the reliability of the assessment results. A single person grading may lead to low reliability especially of essay tests, term papers, and performances. Averaging the results of several markers increases reliability.
7. **Testing conditions**. Where test administrators do not adhere strictly to uniform test regulations and practices, students' scores may not represent their actual level of performances and this tends to reduce reliability. In cases of the test-retest method of estimating reliability, this issue is of a great concern.

SAMPLE TEST ITEMS

1. The process of assigning numbers to the attributes or traits possessed by persons according to specific rules is
   A. assessment.
   B. evaluation.
   C. measurement.
   D. test.

2. Which of the following situations is a norm-referenced interpretation of a test score?
   A. Comfort obtained Distinction in her teaching practice.
   B. Jane obtained 8 As in the WASSC Examination.
   C. Joseph scored 75% in his Statistics examination.
   D. Stephen won the 1st prize in EPS 311 course.

3. The main purpose of formative evaluation is to
   A. attain total growth and development of the student.
   B. determine extent of achievement of objectives of education.
   C. plan the types of traits and behaviours to be assessed.
   D. provide feedback about the progress being made in school.

4. One of the characteristics of continuous assessment is that it is
   A. apprehensive.
   B. diagnostic.
   C. selective.
   D. systemic.

5. One role of the Ghanaian teacher in the implementation of the continuous assessment programme is that the teacher must
   A. concentrate on the cognitive domain objectives.
   B. constantly evaluate the assessment programme.

C. make an end-of-year time table for assessments.
D. provide information to place students in courses.

6. One purpose of the school-based assessment (SBA) in Ghana is to provide teachers with
   A. a list of objectives in constructing assessment items.
   B. guidance in dealing with social and psychological problems.
   C. sources of motivation for directing students' learning.
   D. standards of achievement in each class of the school system.

7. Which of the following constitutes a placement decision in teaching and learning?
   A. Acquiring certificates for employment in the world of work.
   B. Assigning grades to students as a record of progress and achievement.
   C. Grouping individuals for instruction in view of individual differences.
   D. Selecting students for award of prizes in the measurement class.

8. One of the general principles of assessment is that
   A. users become aware of assessment techniques available.
   B. assessment techniques require knowledge about student learning.
   C. good assessment techniques must serve the needs of the teachers.
   D. good assessments are provided by multiple indicators of performance.

9. Learning objectives are important for classroom assessment because they help teachers to
   A. assess students' performances through the knowledge of specific outcomes.
   B. design appropriate assessment procedures based on unknown outcomes.
   C. evaluate existing assessment instruments when specific outcomes are known.
   D. obtain information for judging the reliability of assessment procedures.

10. In the cognitive domain classification of educational objectives, the ability to put parts together to form a new whole is referred to as
    A. analysis.
    B. application
    C. evaluation.
    D. Synthesis.

11. Content-related evidence for the validity of an assessment result is obtained by
    A. analyzing mental processes.
    B. computing correlation coefficients.
    C. examining course objectives.
    D. inspecting specification tables.

12. One of the principles for deciding on the degree to which an assessment result is valid is that evidence can be produced to support the
    A. appropriateness of the test instrument.
    B. appropriateness of educational values.
    C. consequences of the uses of the test items.
    D. content-relevance of the test items.

13. One method of obtaining construct-related evidence is to
    A. analyze behavioral processes required by assessment tasks.
    B. compare scores of groups that differ on the construct.
    C. correlate assessment scores of two different constructs
    D. use novice judgment on tasks that have been defined.

14. Which of the following factors does **NOT** decrease the degree to which a test is valid for a particular purpose?
    A. Ambiguous statements in tasks.
    B. Difficult reading vocabulary.
    C. Unclear directions to students.

D. Unidentifiable patterns of answers.

15. A small value of the standard error of measurement indicates that the

    A. observed score is low.

    B. error score is positive.

    C. reliability is high.

    D. reliability is low.

16. Amina obtained 97 in a Measurement & Evaluation quiz. It is known that the error score for the quiz was − 3.0. An estimate of her true score is

    A. 100

    B. 97

    C. 94

    D. 3

UNIT4: PLANNING CLASSROOM TESTS AND ASSESSMENTS

Definition

Achievement tests are test that measure the extent of present knowledge and skills. In achievement testing, test takers are given the opportunity to demonstrate their acquired knowledge and skills in specific learning situations.

TYPES OF ACHIEVEMENT TESTS

There are two types of achievement tests. These are:

1. Standardized achievement test
2. Teacher-made/classroom achievement test

Standardized Achievement Test

The major difference between these two types of tests is that standardized achievement tests are carefully constructed by test experts with specific directions for administering and scoring the tests. This makes it possible for standardized Achievement tests to be administered to individuals in different places often at the same time.

Characteristics of standardized achievement tests

1. Standardized specific instructions are provided for test administration and scoring. Directions are so precise and uniform that the procedures are standard for different users of the test.
2. The test items are developed by test experts and specialists who follow well formulated procedures for test development. The tests are thus of high quality. Reliability is often over 0.90.
3. They use test norms which are based on national samples of students in the classes/forms where the tests are intended for use.
4. Test content is determined by curriculum and subject-matter experts and involves extensive investigations of existing syllabi, textbooks and programs.
5. Equivalent and comparable forms of the test are usually provided and administrated.

6. A test manual is available as a guide for test administration and scoring. It provides information for evaluating the test for technical quality and interpretation and use of the results.
7. They are useful for measuring broader curriculum objectives and for school, district, regional and national comparisons.

Teacher-Made/Classroom Achievement Tests

These tests are constructed by classroom teachers for specific uses in each classroom and are closely related to particular objectives. They are usually tailored to fit the teacher's instructional objectives. The content of the test is determined by the classroom teacher. The quality of the test is often unknown but usually lower that standardized tests.

Stages in Teacher-Made/Classroom Achievement Testing

The main goal of classroom assessment is to obtain valid, reliable and useful information concerning student achievement. It is therefore important that good and quality tests and assessment tasks are constructed. Four principal stages are involved in classroom testing. These are:

1. Constructing the test
2. Administering the test
3. Scoring the test
4. Analyzing the test results

Stage1: Constructing the test

There are eight steps in the construction of a good classroom test.

Step 1: Define the purpose of the test

The basic question to answer is "why am I testing?" the tests must be related to the teacher's classroom instructional objectives. Several purposes are served by classroom tests and the teacher has to be clear on the purpose of the test. The teacher has to answer other questions such as;

- 'why is the test being given at this time in the course?
- Who will take the test?
- 'Have the test takers been informed?
- 'How will the scored be used'?

Step 2: Determine the item format to use

Test items could either be essay, objective or performance types. Objective-type tests include multiple-choice, short-answer, matching and true and false. The choice of format must be appropriate for testing particular topics and objectives. It is sometimes necessary to use more than one format in a single test. Mehrens and Lehmann (1991) mentioned 8 factors to consider in the choice of the appropriate format. These include:

i. The purpose of the test,
ii. The time available to prepare and score the test,
iii. The number of students to be tested,
iv. Skills to be tested,
v. Difficulty desired,
vi. Physical facilities like reproduction materials,
vii. Age of pupils,
viii. Skills in writing the different types of items.

Step3: Determine what is to be tested

The teacher asks himself or herself the question, 'what is it I wish to measure?', The teacher has to determine what chapters or units the test will cover as well as what knowledge, skills and attitudes to measure. Instructional objectives must be defined in terms of student behaviour and linked to what has been stressed in class. A test plan made up of a table of specifications or blueprint must be made. The specification table matches the course content with the instructional objectives.

To prepare the table, specific topics and sub-topics covered during the instructional period are listed. The major course objectives are also specified, and the instructional objectives defined. The total number of test items is then distributed among the course content and instructional objectives or behaviours.

Example: Terms and concepts in assessment

Content	Behaviour						
	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation	Total
Assessment	1		1	1			3
Test	1	1	1				3

Evaluation	1	1	1				3
Continuous Assessment		2	1	1			4
Test Reliability		2	1				3
Test Validity	1	1		1	1		4
Total	4	7	5	3	1		20

Advantages of the Table of Specification

The table of specifications has a number of advantages.

1) It makes sure that justice is done to all the topics covered in the course.
2) It helps the teacher to determine the content validity of the test
3) It helps to weight the score distribution fairly
4) It avoids overlapping in the construction of the test items.
5) It helps students to determine the content and behavioural areas where they have difficulty. Teachers can also determine areas where the class has difficulty.

Step 4: Write the individual items

In writing the individual items, the specific principles guiding the construction of each Type of test must be followed for the item format chosen. However, the following general guidelines must be considered.

1) Keep the table of specifications before you continually refer to it as you write the items.
2) Items must match the instructional objectives
3) Formulate well-defined items that are not vague and ambiguous and should be grammatically correct and free from spelling and typing errors.
4) Avoid excessive verbiage. Avoid needlessly complex sentences
5) The item should be based on information that the examiner should know.
6) Write the test items simply and clearly
7) Prepare more items than you will actually need.
8) The task to be performed and the type of answers should be clearly defined.
9) Include questions of varying difficulty.

10) Write the items and the key as soon as possible after the material has been taught.
11) Avoid textbook or stereotyped language
12) Write the items in advance of the test date to permit reviews and editing

Step 5: Review the items

- Critically examine each item at least a week after writing the item.
- Items that are ambiguous and those poorly constructed as well as items that do not match the objectives must be reworded or removed.
- Items must not be too difficult or too easy.
- Check the length of the test (i.e. number of items against the purpose), the kinds of test items used and the ability level of the students.
- The test must discriminate between the low achievers and the high achievers.
- Assemble the test in the final form for administration.

Step 6: Prepare scoring key

- Prepare a scoring key or marking scheme while the items are fresh in your mind.
- List correct responses and acceptable variations for objectives-type tests.
- Assign points to the various expected qualities of responses.
- Assign values to each item and ensure representative sampling of content covered.

Step 7: Write directions

- Give clear and concise directions for the entire test as well as sections of the test.
- Clearly state the time limit for the test. Penalties for undesirable writing must be spelt out.
- Directions must include number of items to respond to, how the answers will be written, amount of time available, credits for orderly presentation of material (where necessary), and mode of identification of respondent.
- For selection-type tests, indicate what will be done to guessing.

Step 8: Evaluate the test

Before administration, the test should be evaluated by the following five criteria**: clarity, validity, practicality, efficiency and fairness.**

Clarity:

i. Who is being tested?
ii. What material is the test measuring?
iii. What kinds of knowledge is the test measuring?

iv. Do the test items relate to content and course objectives?

v. Are the test items simple and clear?

Validity:

i. Is the test a representation sampling of the material presented in the chapter, unit, section or course

ii. Does the test faithfully reflect the level of difficulty of material covered in the class?

Practicality:

i. Will students have enough time to complete the test?

ii. Are there sufficient materials available to present the test to complete it effectively?

Efficiency:

i. Is this the best way to test for the desired knowledge, skill or attitude?

ii. What problems might arise due to material difficulties or shortages?

Fairness:

i. Were the students given advance notice?

ii. Have I adequately prepared students for the test?

iii. Do the students understand the testing procedures?

iv. How will the scores affect the students' lives?

UNIT 5: TYPES OF TEST ITEMS (1)

There are two major types of classroom achievement tests. These are;

i. the essay-type test, and
ii. objectives type tests.

ESSAY-TYPE TESTS	OBJECTIVES-TYPE TESTS
i. Requires students to plan their own answers and to express them in their own words	i. Requires students to choose among several designated alternative or write a short answer.
ii. Consists of relatively few items that call for extended answers.	ii. Consist of items requiring only brief answers.
iii. A lot of time spent by students in thinking and writing when taking the test.	iii. A lot of time is spent by students in reading and thinking when taking the test.
iv. Quality of test is determined largely by the skill of the test scorer.	iv. Quality of test is determined largely by the skills of the test constructor.
v. Relatively easy to prepare but rather tedious and difficult to score.	v. Relatively tedious and difficult to prepare but rather easy to score.
vi. Permits and encourages bluffing.	vi. Permits and encourages guessing.
vii. Afford both the student and teacher opportunity to be individualistic.	vii. Afford only the test constructor or the (teacher) the opportunity to be individualistic.
viii. Score distribution varies from one scorer to another.	viii. Score distribution is determined largely by the test.
ix. Less amenable to item and statistical analysis.	ix. Amenable to item and statistical analysis.
x. Scoring is subjective	x. Scoring is highly objective.
xi. Content validity is low	xi. Content validity is high.

xii. Reliability of test scores is low.	xii. Reliability of the test scores could be high.

OBJECTIVE-TYPE TESTS

Description

An objective test requires a respondent to provide a brief response which is usually not more than a sentence long. The tests normally consist of a large number of items and the responses are scored objectively, to the extent that competent observes can agree on how responses should be scored.

There are two major types of objective test. These are the selection type and the supply type. The selection type consists of the multiple-choice type, true and false type and matching type. The supply type has variations as completion, full-in-the-blanks and short-answer.

Strengths and advantages

1. Scoring is easy and objective
2. They allow an extensive coverage of subject content.
3. They do not provide opportunities for bluffing.
4. They are best suited for measuring lower-level behaviours like knowledge and comprehension.
5. They provide economy of time is scoring
6. Student writing is minimized. Premium is not placed on writing.
7. They are amenable to item and statistical analysis
8. Scoring is not affected by extraneous factors such as the likes and dislikes of the scorer.

Weaknesses and disadvantages

1. They are relatively difficult to construct.
2. Item writing is time consuming
3. They are susceptible to guessing
4. Higher-order mental processes like analysis, synthesis and evaluation are difficult to measure.
5. Places premium on student's reading ability.

Multiple-Choice Tests

Description

A multiple-choice test is a type of objective tests in which the respondent is given a stem and then is to select from among three or more alternatives (options or responses) the one that best completes the stem. The incorrect options are called foils or distracters.

There are two types of multiple-choice tests. These are the single 'correct' or 'best response' type and the multiple response type. The single 'correct' or 'best' response' type consists of a stem followed by three or more responses and the respondent is to select only one option to complete the stem.

Examples:

Single correct response

Write 0.039387 as a decimal correct to 3 significant figures

A. 0.394
B. 0.0394
C. 0.0393
D. 0.039

Single best response

In which one of the following sites would you, as a community health worker, advice a community to dispose of refuse?

A. A compost pit
B. An abandoned well
C. An incinerator
D. An uncultivated land

Multiple response type

The multiple response type consists of a stem followed by several true or false statements or words. The respondent is to select, which statement(s) could complete the stem.

An example is:

Which of the following action(s) contribute(s) to general principles of First Aid?

i. Arrest haemorrhage

ii. Bath the patient

iii. Immobilize injured bone

A. I only

B. II only

C. I and II

D. I and III

E. I,II and III

Strengths and advantages

1. Scoring is easy, accurate and efficient.
2. Highly objective measurement of student achievement.
3. They allow an extensive coverage of subject content.
4. They do not provide opportunities for bluffing.
5. They are best suited for measuring lower-level behaviours like knowledge and comprehension.
6. They provide economy of time in scoring
7. Students' writing is minimized; premium is not placed on writing.
8. They are amenable to item and statistical analysis
9. Scores are not affected by extraneous factors such as the likes and dislikes of the scorer.

Weaknesses and disadvantages

1. They are relatively difficult to construct.
2. Item writing is time consuming
3. They are susceptible to guessing
4. Higher-order mental processes like analysis, synthesis and evaluation are difficult to measure.
5. Places premium on student's reading ability.

Guidelines for constructing multiple-choice tests

1. ***The central issue of the item should be in the stem. It should be concise, easy to read and understand.***

The following are examples of poor and good items

Poor	**Good**
Ghana;	The largest man-made in Africa lake is in
A. became independent in 1960	A. Chad
B. has West Africa's largest population	B. Ghana
C. has the largest man-made lake in Africa	C. Kenya
D. is the world's leading cocoa producer	D. Tanzania
E. is a landlocked country	E. Uganda

2. The options should be plausible. Distracters must be plausibly attracted to the uninformed.

Poor	**Good**
The longest river in Africa is…….	The longest river in Africa is……
A. Benue	A. Congo
B. Congo	B. Niger
C. Gambian	C. Nile
D. Nile	D. Volta
E. Thames	E. Zambesi

In the poor example, rivers Benue and Gambia are not significantly long to attract respondents and Thames is not in Africa.

3. All options for a given item should be homogeneous in content

Poor	**Good**
The first woman Prime Minister/Head of State in the world is/was	The first woman Prime Minister/Head of state in the world is/was
A. Margaret Peil	A. Corazon Acquino
B. Margaret Thatcher	B. Golda Meir
C. Mother Teresa	C. Indira Gandhi
D. Sirimavo Bandaranaike	D. Margaret Thatcher
E. Valentine Terescova	E. Sirimavo Bandaranaike

In the poor example, only one options B and D were Prime Ministers

4. ***All options for a given item should be homogeneous in grammatical structure.***

Example

In constructing multi-choice test items, options to an item should be

A. Arranged horizontally
B. Copied directly from class notes or textbooks
C. Must have a discernible pattern of responses.
D. Homogeneous in content.

5. ***All options must follow syntax and punctuation rules***

Poor:

A nurse observes that a colleague nurse reports to work always drunk. What should be the nurse's first reaction?

A. Ignore the drunk nurse
B. Report the colleague to the union
C. Talk to the colleague
D. Request for change of the colleague from the unit.

Good:

A nurse observes that a colleague nurse reports to work always drunk. What should be the nurse's first reaction?

A. Ignore the drunk nurse
B. Report the colleague to the union
C. Request for transfer of the colleague
D. Talk to the colleague.

6. ***Repetition of words in the options should be avoided.***

Poor:

Which is the best definition of a contour-line?

A. A line on a map joining places of equal barometric pressure.
B. A line on a map joining places of equal earthquake intensity
C. A line on a map joining places of equal height.
D. A line on a map joining places of equal mean temperature.
E. A line on a map joining places of equal rainfall

Good:

A line on a map joining places of equal pressure is called an

A. Isobar
B. Isobront
C. Isochasm
D. Isogeotherm
E. Isotherm

7. ***Specific determiners which are clues to the best/correct option should be avoided.***

Poor	**Good**
The first woman cosmonaut is a……	The first woman to go into space is a/an……
A. American	A. American
B. Englishman	B. British
C. Irish	C. French
D. Italian	D. Italian
E. Russian	E. Russian

In the poor example, the article 'a', gives a clue that the correct option is Russian. In addition, it is only Russians who use the term, cosmonant. Also, Englishman does not belong to the group.

8. ***Vary the placement of the correct options. No discernible pattern of the correct/best responses should be noticed.***

9. ***Sentences should not be copied from textbooks or from others' (colleagues, friends etc) past terms. Original items should be made. This builds capacity in item writing.***

10. ***The responses/options in agreement must be in alphabetical/sequential order. This reduces unnecessary searching on the part of the respondents.***

For example:

A. Arranged horizontally
B. Copied from textbooks
C. Heterogeneous in content
D. Homogeneous in content

11. Items measuring opinions should not be included. One option should clearly be correct or the best

Poor

A. Charlotte Gardiner
B. F. I.D. Konotey-Ahulu
C. Mary Grant
D. Mohammed Mustafa

Good

A. I.D. Konotey-Ahulu
B. F.O Acheampong
C. K. G. Korsah
D. M. K. Mustafa

12. The responses in agreement must be itemized vertically and not horizontally.

Poor

In constructing multiple-choice test items, option to an item should be

a. Arranged in horizontally **b.** Copied directly from class notes or textbooks

c. Must have a discernible pattern of responses **d.** Homogeneous in content

Good

In constructing multiple-choice test items, options to an item should be

a. Arranged horizontally

b. Copied from textbooks

c. Heterogeneous in content

d. Homogeneous in content

13. The responses in agreement must be parallel in form i.e. sentences must be about the same length

Poor

In constructing multiple-choice test items, options to an item should be

a. Arranged horizontally

b. Copied from textbooks

c. Heterogeneous in content

d. Homogeneous in content

14. Each option must be distinct. Overlapping alternatives should be avoided.

Poor

In a healthy adult, the liver weights about

A. 3.0kg
B. 2.5kg
C. 2.0kg
D. 1.5kg

Good

In a healthy adult, the liver weighs between

A. 6.5 – 7.5kg
B. 4.5 – 6.0kg
C. 3.0 – 4.0kg
D. 1.0 – 2.5kg

15. Avoid using "all of the above" as an option but "None of the above" can be used sparingly. It should be used only when an item is of the 'correct answer' type and not the 'best answer' type.

Poor

The following are local signs and symptom of inflammation except

A. Rashes
B. Redness
C. Restoration of function
D. Sleeplessness
E. All of the above.

Good

In administering intramuscular injection, the needle is inserted into the muscle at an angle of

A. 30o
B. 45
C. 60
D. 90
E. none of the above.

In the poor example, there are other signs and symptoms not included whereas in the good example there is one and only one answer.

16. *Stems and options should be stated positively. However, a negative stem could be used sparingly, and the word not should be emphasized either by underlining it or writing it in capital form.*

An example is:

Which of these inserts has not been incriminated to transmit diseases?

a. Bed-bug

b. Blackfly

c. Body louse

d. Housefly

e. tsetsefly

17. Create independent items. The answer to one item should not depend on the knowledge of the answer to a previous item.

For example:

Item 1: The perimeter of a rectangular field is 60 meters. If one side is 20 metres long, what is the width of the field?

a. 10 metres

b. 20 metres

c. 30 metres

d. 40 metres

e. 60 metres

Item 2: Find the length of the diagonal of the rectangular field in item 1 above.

a. 10.0 metres

b. 20.0 metres

c. 22.4 metres

d. 30.6 metres

e. 40.0 metres

18. The expected response should not be put at the beginning of the stem.

Poor

…….. printing device transmit output to a printer via radio waves.

a. Infrared

b. Laser

c. Bluetooth

d. Large Format

Good

What printing device transmits output to a printer via radio waves?

a. Bluetooth

b. Laser

c. Cartridge

d. Infrared

19. Check each item to make sure that there is only one correct or best response to the item.

Poor

Amsterdam is the capital city of

a. Holland

b. Hungary

c. Luxemburg

d. Netherlands

Both Holland and Netherlands provide the answer

20. Be consistent in the number of options used. Four or five options are good for higher education students.

21. Read through all items carefully to ensure that the answer to one question is not revealed in another.

Example:

1. What do you use to test for sugar in urine?

a. Albustix

b. Clinitest

c. Ketostix

d. Uristix

2. This can also be used to test for sugar in urine if clinitest is not available.

a. Albustix

b. Ictostix

c. Ketostix

d. Uristix

Levels of Test Items

I. Knowledge: (the recall of specific facts, methods and processes)

1. To what percentage is all continuous assessment scores currently converted at the end of each term in Ghanaian basic schools?
    A. 30%
    B. 40%
    C. 60%
    D. 70%
2. Test reliability refers to the degree of
    A. Consistency of assessment results
    B. Content representativeness of assessment tasks
    C. Relationship among assessment tasks
    D. Soundness of assessment results.

II. Comprehension: (the ability to grasp the meaning of material)

1. A group of examiners are engaged in producing a taxonomy of educational objectives. In ordinary English, what are these persons doing?
    A. Classifying teaching goals
    B. Constructing learning exercises
    C. Evaluating classroom tests
    D. Preparing teaching materials
2. A standard error of measurement of 4 points for a test indicates that a student's observed score is
    A. Greater than or less than the true score by 4 points
    B. Greater that the true by 4 points
    C. Less than the true score by 4 points

D. Less than the true score by 16 points.

III. Application: (the ability to use learned material in new and concrete situations)

1. Which of the following situations is a norm-referenced interpretation of a test score?
    A. Annete scored 100% in her final examination and was promoted.
    B. Justina obtained Grade A in her WASSCE Social Studies examination.
    C. Mathias won the first prize for his performance in the statistics competition.
    D. Patrick passed his end-of term examination in Educational Measurement
2. The steps leading to a swimming pool appear bent where they enter the water. Which one of the following gives the best explanation of this phenomenon?
    A. Diffraction of light by the suspended particles on the surface of the water.
    B. Dispersion of light due to difference in velocity of light on entering water.
    C. Reflection of light because light does not travel in straight lines in water.
    D. Refraction of light due to difference in speed of light in air and water.

True and False Tests Items

Description

A true and false test consists of a statement to be marked true or false. A respondent is expected to demonstrate his command of the material by indicating whether the given statement is true or false.

Variations/Types of True-False Tests Items

There are 4 variations/types

1. Simple True-False
2. Complex True-False
3. Compound True-False
4. Multiple True-False

1. **Simple True-False:** This consists of only 2 choices; True, False

Example:

Sir Gordon Guggisberg was the governor who built the Takoradi Harbour. True or False

2. **Complex True-False:** This consists of 3 choices; True, False, Opinion

Example:

Adrenaline can relax uterine smooth muscles. True, False, Opinion

3. **Compound True-False:** This consists of 2 choices, True and False plus a conditional Completion response

Example;

A nurse who values equality demonstrates honesty to patients. True, False

If this statement is false, what makes it false?

4. **Multiple True-False:** This consists of a stem with three, four or five options and the respondent indicates if the options are True or False

Example:

The factors that reduce the reliability of classroom tests include:

a. Scoring test items objectively

b. Using homogeneous groups

c. Setting essay questions

d. Using unidentifiable pattern of answers

Strength and advantages

i. Scoring is easy, accurate and efficient
ii. Highly objective measurement of student achievement.
iii. They allow an extensive coverage of subject content.
iv. They do not provide opportunities for bluffing
v. They are best suited for measuring lower-level behaviours especially knowledge, comprehension application.

Weaknesses and limitations

i. They incorporate an extremely high guessing factor.
ii. Higher-order objective and behaviours like application, analysis, synthesis and evaluation are difficult to measure.
iii. Often results in ambiguous statements due to difficulty in writing statements that are absolutely true or false.
iv. Often includes more specific determiners or irrelevant clues.
v. Students memorize facts without understanding them.
vi. Students memorise facts without understanding them.
vii. Reliability is now unless there are many items.

Guidelines for constructing true and false tests

1. For simple, compound and Multiple types, statements must be definitely true or definitely false

   Poor: The value of 2/3 as a decimal fraction is 0.7. **True or False**.

   Good: The value of 1 expressed as a decimal fraction correct to two decimal places is 0.66. **True or False**.

   Poor: the value of 2/3 as a decimal fraction is 0.7. **True or False**

   Good: the value of 1 expressed as a decimal fraction correct to two decimal places is 0.66. **True or False**

2. Avoid words that tend to be clues to the correct answer.

   Words like **some, most**, **often**, **many**, **may** are usually associated with true statements and **always**, **never**, **none** are associated with false statement. These words must therefore be avoided.

3. For simple true-false type, approximately, half, (50%) of the total number of items should be false because it is easier to construct statements that are true and the tendency are to have more true statements.

4. Statement must be original. They must not be copied directly from textbooks, past test items or any other written materials.

5. Statements should be worded such that superficial logic suggests a wrong answer.
   Poor: A patient took one tablet of a prescribed medicine and was healed in 24 hours. 8 tablets would therefore heal him in 3 hours. **True or False**
   The true case is that 8 tablet would constitute an overdose.

6. Statements should possess only one central theme
   Poor: Akropong Teacher Training College, built in 1900, is the first teacher training institution in Ghana. Two main themes are in the statement.

7. State each item positively. Negative item could however be used with the negative word, 'Not' emphasized by underlining or writing in capital letters. Double negatives should be avoided.

8. Statements should be short, simple and clear. Ambiguous as well as tricky statements should be avoided.
   Examples: (a) Abedi Pele is the best Ghanaian footballer. **True or False**
   (b) Margaret Thacher was the British Prime Minister in 1989. **True or False**
   Item (a) is ambiguous because best is relative while the trick in item (b) is the spelling of Thatcher

9. Statement should measure important ideas nor trivial.
   Poor: Dr. Kwame Nkrumah, had artificial teeth. **True or False**
   Good: Dr. Kwame Nkrumah was the first President of Ghana. **True or False**

10. Arrange the items such that the correct responses do not form a discernible pattern like TTT FFFF TTTT FFFF.
11. To avoid scoring problems, let students write the correct options in full.

12. Double-barreled statements should be avoided. These statements have one part true and one part false.

    Poor: The Bond of 1844, signed by Governor Commander Hill declared the Northern territories of Ghana a Protectorate.

    Good: The Bond of 1844 was signed by Commander Hill but did not achieve the stated purpose.

13. Avoid the use of unfamiliar vocabulary.

    Poor: According to some politicians, the raison d'etre for capital punishment is retribution.

    Good: According to some politicians, the justification for the existence of capital punishment is retribution.

14. Avoid using extreme items. Words such as; all, no, always, never, the very most or very least, usually make a statement false.

    Example: The weather in Northern Ghana is always hot. **True or False**

Matching-Type Tests

Description

The matching type of objective test consists of two columns. The respondent is expected to associate an item in Column A with a choice in Column B on the basis of a well-defined relationship.

Column A contains the premises and Column B the responses or options.

Example.

Match the vitamins in Column A with the diseases and conditions which a lack of the vitamin in column A causes in column B.

Column A:	Column B
Vitamin A	a. beriberi
Vitamin C	b. kwashiorkor
Vitamin D	c. pellagra
	d. Poor eyesight
	e. Rickets

f. Scurvy

Strengths and Advantages

i. Scoring is easy and objective
ii. They allow an extensive coverage of subject content
iii. They do not provide opportunities for bluffing
iv. They are best suited for measuring lower-level behaviours like knowledge and comprehension.
v. They provide economy of time in scoring
vi. Student writing is minimized. Premium is not placed on writing.
vii. Item and statistical analysis can be done easily.
viii. Scores are not affected by extraneous factors such as the likes and dislikes of the scorer.
ix. They encourage the integration of information.
x. Assesses students understanding of relationships.
xi. Reliability is often high because of scoring objectivity.

Weaknesses and disadvantages

i. Item writing is time consuming
ii. It provides opportunities for guessing
iii. Items are relatively difficult to construct
iv. Higher-order mental processes like analysis, synthesis and evaluation are difficult to measure. Often restricted to factual knowledge.
v. Premium is placed on students reading ability.
vi. It is difficult to generate sufficient number of plausible premises.
vii. May overestimate learning due to the influence of guessing.
viii. It is often difficult to obtain sufficient homogeneous items for a set.

Guidelines for constructing matching-types tests

1. Do not use perfect matching. Have more responses than premises. There should be at least three more responses than premises.

2. Arrange premises and responses alphabetically or sequentially. This reduces the amount of unnecessary searching on the part of the person who the answer.
3. Column A (premises) should contain the list of longer phrases. The shorter items should constitute the responses.
4. Limit the number of items in each set. For each set, the number of premises should not be more than six per set with the responses more than ten.
5. Use homogeneous options and items.

Poor

Instruction: Select an option from list B to match list A.

A	B
1. The Battle of Dodowa	a. 1824
2. built Korle Bu hospital	b. Gordon Guggisberg
3. longest river in Africa	c. Nile
	d. 1826
	e. Lord Listowel
	f. Congo

Good:

Instruction: Select a river from list B to complete the description in list A. write the answer against the number in list A.

A	B
Description of river	Name of river
1. Aswan Dam is built on it.	a. Niger
2. The longest river in Africa	b. Nile
3. It is a tributary of River Congo	c. Orange
	d. Ubangi
	e. Volta
	f. Zambezi

6. Provide complete directions. Instruments should clearly show what the rules are and also how to respond to the items.
7. State clearly what each column represents.
8. Avoid clues (specific determiners) which indirectly reveal the correct option.

9. All options-must be placed (and typed) on the same page.
10. Avoid using multiple correct choices for one premise.

Constructed-Response Type

Short-Answer Type Tests

Description

This type of objective test is also known as the Supply, Completion, and fill-in the blanks. It consists of a statement or question and the respondent is required to complete it with a short answer usually not more than one line.

Examples:

a. Modern nursing was introduced into Ghana in the year?
b. What is the name of the first Ghanaian Prime Minister?
c. The environment has three component parts: Name them.

Strengths and Advantages

1. Scoring is easy
2. They allow an extensive coverage of subject content.
3. They do not provide opportunities for bluffing
4. Minimize guessing.
5. They are best suited for measuring lower-level behaviours especially knowledge, comprehension application.
6. Encourage students to study in a deeper and more integrated manner.
7. Can assess students' reasoning skills
8. Makes cheating more difficult and reduces its incidence as compared to multiple-choice and true/false types of tests.
9. Discourages last minute rote learning.
10. Gives some practice in writing.

Weaknesses and Limitations

1. They are difficult to construct so that the desired response is clearly indicated.
2. Higher-order objectives and behaviours are difficult to measure.
3. Often includes more specific determiners
4. Time consuming to score
5. Difficult to score since more than one answer may have to be considered.
6. Panelises students who write slowly or have poor writing skills.

Guidelines for constructing short-answer tests

1. Keep the number of missing words or blank spaces low. Preferably use one blank per item. There should not be more than two blanks in one item.
   Poor: The ……. of ……tool place in …….
   Good: The battle of Dodowa took place in the year…….
2. Use original statements that are carefully constructed. Statements should not be lifted from textbooks or past items or any written materials.
3. Avoid specific determiners which provide clues to the correct option.
4. Blanks must be placed at the end or near the end of the statement and not at the beginning.
   Poor: ……. Is an instrument used for measuring temperature.
   Good: An instrument used for measuring temperature is called……..
5. Items should be so clearly written that the type of response required is clearly recognized.
   Poor: The battle of Nsamankow was fought in ……..
   Good: The battle of Nsamankow was fought in the year……
6. Avoid lengthy and tortuous statements
   Poor: A specific disease in which acute glomerular damage occurs following distant infections, particularly with certain streptococci and usually affects children and young adults and which clinical picture is commonly one of a dramatic onset of oedema and haomaturia is……
7. Think of the intended answer first before constructing the item.
8. Missing words must be important ones. Avoid omitting trivial words to trick the student. Only test for important facts and knowledge.

Poor: The…… of the June 4, 1979 revolution in Ghana was Flt. Lt. J. J. Rawlings

9. Specify the degree of precision and the units of expression required in computational problems.

   Poor: The value of 2.6 ÷ 0.07 is ……..

   Good: the value of 2.6 ÷ 0.07 correct to 3 decimal places is…….

10. Aim at providing items that belong to the correct answer type and not the best answer type.

    Poor: The best audio-visual materials to use in the classroom is …….

    Good: Radios and tape recorders are regarded as……. audio-visual aids.

11. Keep all blanks the same length, and in a column to the right of the question.
12. A direct question is generally more desirable than an incomplete statement.
13. Allocate marks/scores fairly to each item where sub-items are used.

    Example:

    i. State the function of a retaining wall

    ii. State 3 design principles of a retaining wall

UNIT 6: TYPES OF TEST ITEMS (II) (ESSAY –TYPE TESTS & COMPUTATION ITEMS)

ESSAY-TYPE TESTS

Description:

An essay type test is a test that gives freedom to the respondent to compose his own response using his own words. The tests consist of relatively few items, but each item demands an extended response.

Types of Essay-Type Tests

There are two types of essay-tests. These are;

i. the restricted response type and
ii. the extended response type.

The restricted response type limits the respondent to a specific length and scope of the response. For example, in not more than 200 words explain the causes of the Yaa Asantewa War of 1900.

The extended response type does not limit the student in the form and scope of the answer. For example, discuss the factors that led to the overthrow of the D. Kwame Nkrumah's government in Ghana in 1966.

Strengths and Advantages

1. They provide the respondent with freedom to organize his own ideas and response within unrestricted limits.
2. They are easy to prepare.
3. They eliminate guessing on the part of the respondents.
4. Skills such as the ability to organize material and ability to write and arrive at conclusions are improved.
5. They encourage good study habits as respondents learn materials in wholes.
6. They are best suited for testing higher-order behaviours and mental processes such as analysis, synthesis and evaluation.
7. Little time is required to write the test items.

8. They are practical for testing a small number of students.

Weaknesses and Disadvantages

1. They are difficult to score objectively. Starch and Elliott (1912, 1913) reported that inter-rater variability could be as high as 68%.
2. They provide opportunities for bluffing where students write irrelevant and unnecessary materials.
3. Limited aspects of student's knowledge are measured as students respond to few items only.
4. The items are an inadequate sample of subject content. Several content areas are omitted.
5. Premium is placed on writing. Students who write faster, all things being equal are expected to score higher marks.
6. They are time-consuming to both the teacher who scores the responses and the student who writes the responses.
7. They are susceptible to the halo effect where the scoring is influenced by extraneous factors such as the relationship between scorer and respondent.
8. A critical teacher as well as a competent scorer can only effectively score responses.

Guidelines in constructing good classroom essay tests

1. Plan the test: Give adequate time and thought to the preparation of the test items. The test items must be constructed from a test specification table and well in advance (at least two weeks) of the testing date. This allows editing to be done.
2. The items should be based on novel situations and problems. Be original. Do not copy directly from textbooks or colleagues/others' past test items.
3. Test items should require the students to show adequate command of essential knowledge. The items should not measure rote memorization of facts, definitions and theorems but must be restricted to the measuring of higher mental processes such as application, analysis, synthesis and evaluation.

Examples of items include:

a. **Application**:

You are in charge of a youth camp of 100 campers. Prepare a menu chart which shows a balanced diet taking into consideration cost and nutritional value. Here the student uses knowledge learnt in school to deal with a concrete situation.

b. **Analysis**:

A form 1 student girl was severely and unfairly punished. Describe the feelings such treatment aroused in her.

c. **Synthesis**:

You are the financial secretary of a society aimed at raising money to build a fish pond in your community. Plan and describe a promotional campaign for raising the money.

d. Evaluation:

e. Evaluation the function of the United Nations Organization as a promoter of world peace.

4. The length of the response and the difficulty level of items should be adapted to the maturity level of students (age and educational level).

   An item like:

   "Discuss the implications of the Lome II Convention on the economy of Ghana" would be too difficult for a first year senior high school student.

5. Optional items should be provided when content is relevant. These may be necessary only for large external examinations and when the purpose of the test is to measure writing effectiveness. If students answer different questions, an analysis of the performance on the test items is difficult.
6. All items should be of equal difficulty if students are to select from a given number of items.
7. Prepare a scoring key (marking scheme) at the time the item is prepared. Decide in advance what factors will be considered in evaluating an essay response. Determine the points to be included and the weights to be assigned for each point. The preparation of a model answer will help disclose ambiguities in an item.
8. Establish a framework and specify the limits of the problems so that the student knows exactly what to do.

The following item for example does not establish any framework for the student to operate in.

Write brief on the following:

i. United Nation Organization (UNO)

ii. African Union (AU)

iii. European Union (EU)

9. Present the student with a problem which is carefully worded so that one ONE interpretation is possible. The questions/items must not be ambiguous or vague.

   For example: Family planning in Ghana is a "missed bag". Discuss.

   Different interpretations could be given to the term 'mixed bag' if it was not mentioned in class.

10. Indicate the value of the question and the time to be spent in answering it.
11. Structure the test item such that it will elicit the type of behavior you really want to measure.
12. The test items must be based on the instructional objectives for each content unit.

    An item like:

    Discuss the factors which in your opinion, contributed to the escalation of the Persian Gulf War in 1990.

    This item elicits student opinions which might be different from the behaviour desired.

13. Give preference to a large number of items that require brief answers. These provide a broader sampling of subject content and thus better than a few items that require extended responses.
14. Statements and sub-questions for each item should be clearly related.

    - Poor: The facilitator/tutor is central in the teaching/learning process in the professional educational programme. Evaluate the roles and responsibilities of tutors in distance education programmes in Ghana.
    - Schizophrenia is an illness in which the functioning of the brain becomes disorganized leading to disturbed feelings and emotions, confusion of thought, abnormal behavior and withdrawal from the reality of the environment into unreal world of fantasy Sainsbury 1987. Discuss.

15. Start essay test items with words that are clear and as simple as possible and which requires the student to respond to the stimulus expected. Avoid words such as: what, list, who, as much as possible.
    - For example: What can you as a teacher do to promote professionalism in the teaching service in Ghana. This item requires only a statement as the response and not an extended answer.

Commonly used words to start essay questions

1. **Explain**: To make plain or clear; to make known in detail. To tell what an activity, process, is and how it works and why it works the way it works
2. **Describe**: to tell or depict (a picture) in written words.
3. **Analyze**: to determine elements or essential features; examine in detail to identify causes, key factors, possible results.
4. **Assess**: to estimate or judge the value, character etc of
5. **Examine**: to inspect or scrutinize carefully; to inquire into or investigate.
6. **Discuss**: to consider or examine by argument, or comment; give points for and against the content of the question.
7. **Evaluate**: to judge or determine the significance worth or quality of, it involves discussion and making a judgment.
8. **Give an account of** –to describe a process/activity, and giving reasons, causes, effects etc.

Scoring essay tests

Essay tests can be scored by using the **analytic scoring rubrics** (also known as the point-score method) or **holistic scoring rubrics** (also call global-quality scaling or rating method).

Analytic scoring

In analytic scoring, the main elements of the ideal answer are identified and points awarded to each element. This works best on restricted response essays.

Example 1

Discuss five reasons why there is the need for counseling in our tertiary institutions **(30marks)**

Introduction:

Definition/Explanation of counselling **(5 marks)**

Counselling is the skilled and principled use of relationship to facilitate self-knowledge, emotional acceptance and growth and the optimal development of personal resources of client(s) towards living more satisfying and resourcefully and reaching self-determined goals through meaningful and well-informed choices.

Content

(5 reasons x maximum 5 mark each. Total = 25 marks).

Consider expression, organization and mechanics of writing in addition to discussion for each reason.

Reasons

1. Upsurge in institutional related problems
2. Need to make informed choices
3. Need to make informed decisions
4. Inability of home to cope with all problems of students
5. Promotion of attitudinal change
6. Need for coping and adjustment skills
7. Demands for modern world.
8. Other reasons that are relevant are accepted.

Example 2

Introduction:

Definition/Explanation of counseling (5 marks)

Counseling is skilled and principled use of relationship to facilitate self-knowledge, emotional acceptance and growth and the optimal development of personal resources of client(s) towards living more satisfyingly and resourcefully and reaching self-determined goals through meaningful and well-informed choices.

Content

(5 reasons x maximum 4 marks each. Total = 20mks).

1. Upsurge in institutional related problems
2. Need to make informed choices
3. Need to make informed decisions
4. Inability of home to cope with all problems of students

5. Promotion of attitudinal change
6. Need for coping and adjustment skills
7. Demands for modern world.

Others reasons that are relevant are accepted. Definition/Explanation of counseling

Organization, Mechanics of writing and expression (5marks)

Holistic scoring

In holistic scoring, major points are written. Sometimes 4 levels of quality are described, and marks awarded. E.g.

A. Excellent 26-30
B. Very good 21-25
C. Good 16-20
D. Fairly good 11-15
E. Fail Below 11

For each level of quality, descriptions are made related to the question. A model answer serves as a standard. Each response is read for a general impression of its adequacy as compared to the standard. The general impression is then transformed into a numerical score.

To check the consistency of the scoring, a first reading is done to sort the responses into several piles (mostly five A, B, C, and D, E) according to the different levels of quality. A second reading of each pile enables the actual grade or score to be given.

Example

Discuss five reasons why there is the need for counseling in our tertiary institutions. **(30 marks)**

Main Reasons

1. Upsurge in institutional related problems
2. Need to make informed choices
3. Need to make informed decisions
4. Inability of home to cope with all problems of students
5. Promotion of attitudinal change
6. Need for coping and adjustment skills
7. Demands for modern world.

A: Excellent (26-30)

- Gives an introduction
- Discusses five reasons very well/in depth
- Very few grammatical errors/expressions
- Gives conclusion

B: Very good (21-25)

- Gives an introduction
- Discusses five reasons but not too well or discusses four reasons very well
- Few grammatical errors/expressions
- Gives conclusion

C: Good (16-20)

- Gives an introduction
- Discusses five/four reasons but not in depth

OR discusses three reasons very well

- Many grammatical errors/expressions
- No conclusion

D: Fairly Good (11-15)

- No introduction
- Discusses three reasons but not depth
- Many grammatical errors/expressions
- No conclusion

E: Fail (Below 11)

- No introduction
- Discusses one/two reasons but not in depth
- Many grammatical errors/expressions
- No conclusion

Principles for scoring essay tests

1. Prepare a form of scoring guide, either an analytic scoring rubric or a holistic scoring rubric.

2. Tests must be kept as anonymous as possible. Score tests without knowing the one whose paper is being scored. This reduces the **Halo effect**. Different forms of identification could be used instead of names.
3. Grade the responses item by item and not script by script. Score all responses to each item before going to the next item. This reduces the **carryover effect**. The carryover effect occurs when the mark for a question is influenced by the performance on the previous question.
4. Keep scores of previously graded items out of sight when evaluating the rest of the items.
5. Periodically rescore previously scored papers.
6. Before starting to score each set of items the scripts should be shuffled.
7. Score the essay test when you are physically sound, mentally alert and in an environment with little or no distraction.
8. Constantly follow the scoring guide as you score. This reduces the **rater drift** which is the tendency to either not paying attention to the scoring guide over time or interpreting it differently as time passes.
9. Score a particular question on all papers at one sitting. Break when fatigue sets in.
10. Arrange for an independent scoring of the responses or at least a sample of them where grading decision is crucial.
11. Comments could be provided and errors corrected on the scripts for class tests to facilitate learning.
12. Avoid being influenced by the first few papers read. These could make you either too harsh or too lenient.
13. The mechanics of writing such as correct grammar usage, paragraphing, flow of expressions, quality of handwriting, orderly presentation of materials and spelling should be judged separately from the content.

COMPUTATIONAL TEST ITEMS

They are subjective type test items which present the student with a computation task and require a demonstration of work procedures and a correct solution.

Instructors assign full or partial credit to either correct or incorrect solutions depending on the quality and kind of work procedures presented.

Example:

The following scores are from a class of ten M. Phil. Students in Psychology.

75 76 83 60 80 77 84 81 72 68

i. Calculate the arithmetic mean.
ii. Calculate the quartile deviation.
iii. Estimate the value of the standard errors or the means.

Strengths and Advantages

1. They are quite easy to construct.
2. They measure most levels of learning objectives.
3. Guessing is minimized.
4. Opportunities to bluff are limited.
5. It allows an extensive coverage of content.
6. Allows for partial credit to be awarded.
7. Rote memorization is minimized. Students study in wholes.

Weaknesses and Limitations

1. They are subject to bias on the part of the grader. There is less objectivity in scoring.
2. They are time consuming to score.
3. They are difficult to score since more than one answer may have to be considered.

Guidelines in constructing Computational Items

1. Clearly identify and explain the problem.

   Poor: During a car crash, the car slows down at the rate of 490 m/sec$^{2.}$
   What is the magnitude and direction of the force acting on a 100kg driver?

   Good: During a car crash, the car slows down at the rate of 490m/sec^2
   Using the car as a frame of reference, what is the magnitude and direction of the gram force acting on a 100kg driver?

2. Provide directions which clearly inform the student of the type of response called for:

   Poor: A Ghanaian student in London finds that he weighs 75 kilograms, when he left, he weighed 140 pounds. What was his net change in weight?

   Good: A Ghanaian student in London finds that he weighs 165 pounds after three months stay. When he left Accra, he weighed 63 kilograms. What was his net weight

change in pounds?

3. State in the directions whether or not the student must show all his/her work procedures for full or partial credit.

   Example:

   Answer all questions in Section A. Show all your work/steps

   i. Find the equation of the tangent to the curve $y = 4x^2$ at the point (2, 16).

   ii. The side of a square is 5cm. Find the increase in the area of the square if the side expands by 0.01cm.

4. Clearly separate item parts and indicate their point values.

   Example:

The following scores are from a class of ten M. Phil. Students in Psychology.

75 76 83 60 80 77 84 81 72 68

i. Calculate the arithmetic mean. (5 marks)

ii. Calculate the quartile deviation. (10 marks)

iii. Estimate the value of the standard errors or the means. (5 marks)

5. Use figures, conditions and situations which create a realistic problem.

Poor: You are required to open the asset and liability and capital account and record the following transactions for June, 2012 in the records of C. Wumbei.

2012

June 1, Started business with GH₵ 2 in cash.

June 2, Paid GH₵ 1.80 of the opening cash into a bank account for business.

June 5, Bought office furniture on credit from Betta Ltd for GH₵1.20.

June 8, Bought a van, paying by cheque GH₵ 9.50.

June 28, Took GH₵ 1 out of the bank.

6. Ask questions that elicit responses on which experts could agree that one solution and one or more work procedures are better than others.
7. Work through each problem before classroom administration to double-check accuracy.
8. Allocate separate marks for method, accuracy and correct answer.

Knowledge

The following scores were obtained in a class test, 24, 18, 7, 10, 20, 17, 12, 14, 9 obtain the median score.

Comprehension

The distribution of scores from an end-of- semester examination is normal with a mode of 55 and a variance of 16. What is the value of the mean score?

Application

1. **G**iven that T = 50 + 10z, where Z = (raw score – mean)/std. dev.
   Fati had T score of 90, and the group's mean is 70. Calculate the value of her raw score if it is known that the standard deviation for the group is 4?
2. My daughter is twice as old as my son and half as old as I am. In twenty-two years my son will be half my age. How old is my daughter?

Scoring computation items

1. Prepare the correct response for each item showing the method (M), accuracy (A) results and factual/basic understanding (B).
2. Distribute the total marks for each question for method (M), accuracy (A) and factual basic understanding (B).
3. Provide notes for the observance of the rule that when the method is not correct, accuracy must necessarily be wrong.
4. Indicate marks for partially correct answers.
   - When important step(s) is/are answers.
   - When wrong results are used for a correct method.
   - Only correct answer but no method

UNIT 7

ASSEMBLING, ADMINISTERING AND APPRAISING ACHIEVEMENT TESTS

Guidelines for Assembling Achievement Tests

1. **Review test items and assessment tasks**. Where possible, fellow teachers or colleagues can review the test items or tasks. The following points should be considered.
   i. Test format should be appropriate for the learning outcome being measured.
   ii. Knowledge, understanding or thinking skill required by the item or task should match the specific learning outcome and subject-matter content being measured.
   iii. The item or task should not be excessively wordy.
   iv. The point of the item or task as well as the desired response should be clear.
   v. A scoring rubric or scoring guide should be available.
   vi. The item or task should be free from technical errors and irrelevant clues.
   vii. The item or task should be free from racial, ethnic and gender bias.
2. **Decide on the total number of items and the length of time to constitute the test**. The number of items and time to be used is dependent on institutional policy, number of credits for a course, test format, item difficulty level, examiner's experiences and students' maturity level.
3. **Test items should be typed or written neatly**. Writing items on the chalkboard or dictating them must be done with utmost care since it may cause problems for students especially those with visual, listening comprehension or hearing problems.
4. **Arranging test items**
   - Items should be sequenced (especially objective-type tests) such that they appear in the order of difficulty with the easiest ones placed first.
   - Items should also be arranged in sections by item-type. The sections should progress from easier formats to more difficult formats. Within each section, group items such that the easier ones come first. For example, all true-false items should be grouped together, then all matching items and so on.

- Items can also be arranged according to the order in which they are taught in class or the order in which the content appeared in the textbook.
- Sequencing is not necessary for essay-type tests where optional choices are made. all items of this nature should however equal difficulty levels

5. **Provide directions to students**. Directions should include the amount of time allowed to complete the test, where and how answers should be written, number of points for each test item, what should be done about guessing (on selection-type items). Each item format should have a specific set of directions.
6. **Reproducing the test**.
   - Items must be spaced and arranged so that they can be read and scored (for objective-type tests) with the least amount of difficulty. Cramming to many tests onto a page is poor economy.
   - Multiple-choice items should have the alternatives listed vertically below the stem of the item rather than across the page.
   - Items should not be split with parts of the item on two different pages. All items should be numbered consecutively.
   - All illustrative material should be clear, legible and accurate.
   - Proofread the entire test or assessment before it is finally reproduced.

Guidelines in Administering Achievement Tests

1. Prepare students for the test. The following information is essential to students' maximum performance.
   - When the test will be given (date and time)
   - Under what conditions it will be given (timed or take-home, number of items, open book or closed book, place of test).
   - The content areas it will cover (study questions or a list of learning targets).
   - Emphasis or weighting of content areas (value in points).
   - The kinds of items on the test (objective-type or essay-type tests).
   - How the assessment will be scored and graded.
   - The importance of the results of the test.

2. Students must be made aware of the rules and regulations-covering the conduct of the test. Penalties for malpractice such as cheating should be clearly spelt out and clearly adhered to.
3. Avoid giving tests immediately before or after a long vacation, holidays or other important events where all students are actively involved physically or psychologically/emotionally.
4. Avoid giving tests when students would normally be doing something pleasant. e.g having lunch etc.
5. The sitting arrangement must allow enough space so that pupils will not copy each other's work
6. Adequate ventilation and lighting is expected in the testing room.
7. Provision must be made for extra answer sheets and writing materials.
8. Pupils should start the test promptly and stop on time.
9. Announcements must be made about the time at regular intervals. Time left for the completion of the test should be written on the board where practicable.
10. Invigilators are expected to stand at a point where they could view all students. They should once a while move among the pupils to check on malpractices. Such movements should not disturb the pupils. He/she must be vigilant.
11. Invigilators should NOT be allowed to read novels, newspapers, grade papers or receive calls on mobile phones.
12. Threatening behaviours should be avoided by the invigilators. Speeches like 'if you don't write fast, you will fail" are threatening. Pupils should be made to feel at ease.
13. The testing environment should be free from distractions. Noise should be kept to a very low level if it cannot be eliminated or removed. Interruptions within and outside the classrooms should be reduced. It is helpful to hang a "do not DISTURB – TESTING IN PROGRESS" sign at the door.
14. Test anxiety should be minimized.
    - Things that create excessive anxiety are (i) warning students to do their best 'because the test is important' (ii) telling students that they must work fast in order to finish on time, (iii) threatening dire consequences if they fail, and (iv) threatening students with tests if they do not behave.

- Teacher and invigilators should not walk around looking over students shoulders whiles they are responding to assessments.
- Before assessments, teachers should convey a sense of confidence about student's performance in the upcoming assessments.

15. Do not talk unnecessarily before letting students start working. Remarks should be kept to a minimum and related to the test.
16. Avoid giving hints to students who ask about individual items. Where an item is ambiguous, it should be clarified for the test.
17. Expect and prepare for emergencies. Emergencies might include shortages of answer booklets, question papers, power outages, illnesses etc.

APPRAISING ACHIEVEMENT TESTS (ITEM ANALYSIS)

Item analysis is the process of collecting, summarizing and using information from students' responses to make decisions about each test item. It is designed to answer the following questions:

1. Did the item function as intended?
2. Were the test items of appropriate difficulty?
3. Were the test items free of irrelevant clues and other defects?
4. Was each of the distracters effective (in multiple-choice items)?

Benefits of item analysis

1. It helps to determine whether an item functions as intended. It provides information on whether an item assesses the intended learning targets, whether it is of the appropriate level of difficulty or whether it distinguishes between high achievers and low achievers and whether the options are working.
2. Items analysis data provide a basis for efficient class discussion of the test results. Difficult items can be identified and discussed. Misinformation and misunderstanding of distracters can be corrected.
3. Item analysis provides feedback to the teacher about pupil difficulties. It brings to light general areas of weakness that require more attention.

4. Item analysis data provides a basis for the general improvement of classroom instruction. It assists in evaluating the appropriateness of the learning outcomes and course content of the particular students being taught.
5. Item analysis procedures provide a basis for increased skill in test construction. Item analysis reveals ambiguities, clues, ineffective distracters and other technical defects that were missed during the test preparation. Information revealed provided experience for future writing of tests.
6. It helps to create item banks for use in future tests.

Steps in doing item analysis of objective tests

1. Arrange the marked test papers from the highest score to the lowest score.
2. Create three groups – upper, middle and lower groups using the top 27% and the bottom 27% if the total number of students is more than 40. Where the number of students is between 20 and 40, select the top 10 students and the bottom 10 students. For fewer than 20 students, create only two groups.
3. For each item summarize the number of students in each of the upper and lower groups who selected each option.
4. Calculate the difficulty index. I.e. percentages of the total number of students who got the item correct. The difficult index by convention is written as a decimal.
5. Compute the discrimination index. I.e. the difference between the percentages of students in the upper and lower groups who got the item correct. The discrimination index is often written as a decimal fraction.
6. Evaluate the effectiveness of options for multiple-choice tests (distracter analysis).
    - Every distracter should have at least one lower group student choosing it, and more lower group students than upper group students should choose it.
    - Every correct option should be selected by more students in the upper group.
    - Options are ambiguous if upper group students are unable to distinguish between the correct response and one or more of the distractors.
    - If a large number of upper group students select particular wrong responses, check to be sure the answer key is correct.

Examples:

1. **Ideal**

Options	**Upper Group**	**Lower Group**
A	0	2
B	2	4
C	15	5
D	3	9

2. **Ambiguous alternative**

Options	**Upper Group**	**Lower Group**
A	1	4
B	10	5
C	9	5
D	0	6

Options B and C seem equally attracted to the high achiever. Option C should be checked as well as the test item for ambiguities.

3. **Miskeyed Item**

Options	**Upper Group**	**Lower Group**
A	13	7
B	6	6
C	0	3
D	1	4

Majority of the upper group selected A. Option A might be the correct responses and not D.

4. **Poor distractor**

Options	**Upper Group**	**Lower Group**
A	2	6
B	12	6
C	0	0
D	6	8

Option C attracted no student. It is a poor distractor and has to be replaced.

Using Difficulty and Discrimination Indices

Points to Note

- A low index of discriminating power does not necessarily indicate a defective item, they could be examined, however, for the possible presence of ambiguity, and other technical effects especially if they are selection-type items.
- Negatively discriminating items should be avoided and not used in test construction.
- Discrimination indexes of 0.30 and above are more desirable in test construction. However, items with high, positive discrimination indices are used mostly by test developers on standardized tests.
- It is sometimes necessary to retain items with low discriminating power in order to measure a representative sample of learning outcomes and course content.
- Items with a 50% level of difficulty make maximum discrimination possible.
- In norm-referenced testing, difficulty indices between 0.16 and 0.84 are used to select items where the test represents a single ability. If performance on the test represents several different abilities, difficulty indices between 0.40 and 0.60 are used.

UNIT 8: INTERPRETATION OF TEST SCORES

Scores obtained in classroom quizzes, tests and examinations are known as raw scores. They give very little information about performance or achievement of students. For example, if John obtained 48 in a test, it is difficult to know his level of performance unless more information is provided. Such types of information include the maximum/total score, mean or median score the variability of the group, the difficulty level of the items, the number of test questions and the amount of time allowed for the test. To interpret and obtain meaning from the scores, they need to be referenced or transformed into other scores.

WAYS OF TEST SCORES INTERPRETATION

There are two popular ways of interpreting test scores so that meaning can be derived from the scores. These are:

1. Norm-referenced Interpretation
2. Criterion-referenced Interpretation

NORM-REFERENCED INTERPRETATION

These describe test scores or performance in terms of a student's position in a reference group that has been administered the assessment. In other words, it compared an individual's performance with others in the group who have taken the same test. The reference group is called the norm group.

In the earlier example, John's score of 48 can be compared with the mean score for the class. If the mean score is 40, then one could say that John's performance was above average. If the median score is 40, then one could also say that John's performance could be placed in the upper half of the class. The score of 40 can appropriately be called the norm and the class that provided the mean or median of 40 is called the norm group.

Types of norm-reference scores

The most popular norm-reference scores are described below.

1. **Class raw score ranks:** Raw scores in a class are often ordered in merit form from the highest score (1st position) to the lowest score (last position). The ranks tell about how a student performs compared with the others in the group.

2. **Percentile and percentile ranks:** A percentile is a point in a distribution below which a certain percentage of the scores fall while a percentile rank is a person's relative position such that a given percentage of scores fall below the score obtained. If a raw score of 48 is the 60th percentile, it means that a student who obtains 48 in a test, has done better than 60 percent of all those in the group that took the test.
3. **Standard Scores:** These are either linear Z or T scores. A linear Z scores is based on the normal distribution such that the mean is 0. Raw scores that are transformed to Z use the formula: $Z = X - \bar{X}$, where X is the raw score, $\bar{X}$ is the group mean and S, the group standard S deviation. Negative values show that performance is below average and positive values means the performances is above average. T-scores are based on Z-scores and use the formula: $T = 50 + 10Z$. Scores above 50 show above average performance and scores below 50 shows below average performance.
4. **Stanines (Standard Nine):** these are derived scores based on the normal distribution with a mean of 5 and standard deviation of 2. It uses the intergers, 1 – 9. The percentage of scores at each Stanine are: 9 (top 4%), 8 (next 7%,), 7 (next 12%), 6 (17%), 5 (next 20%), 4 (next 17%), 3 (next 12%), 2 (next 7%) and 1 (lowest 4%).

Uses of norm-referenced interpretations

1. **Selection decisions:** In selecting students for awards and prizes, a norm-referenced approach if often used. To award the prize for the best student, a ranking is done and the student at the top position is awarded the prize. Also, the ranking of students allows the selection of those who meet a fixed quota. If an educational institution wants to admit 400 students out of a total of 600 based on an entrance examination, an order of merit is done and the top 400 students are selected.
2. **Comparison decisions:** Norm referenced interpretations allow the comparison of performance across subjects. For example, performance in Mathematics and English can be compared by using Z and T scores. It may be necessary to compare performance between two classes, for example, Home Economics and Arts. Norm referenced scores provide the information needed for the comparisons. Mean or median scores would provide information as to which class performs better.

3. **Achievement testing:** Examination bodies such as the West African Examination Council used norm-referenced scores sometimes in interpreting the results of students in examinations such as the BECE and WASSCE.
4. **Monitoring decisions:** Norm-referenced scores are useful in monitoring the general progress of individual students. A student who was at the 10th percentile in Mathematics in the first term but moved up to the 75th percentile in third term has made much progress.

CRITERION-REFERENCED INTERPRETATION

These describe test scores or performance in terms of the kinds of tasks a person with a given score can do. The performance can be compared to a pre-established standard or criterion. For example, a student may be about to solve 8 problems out of 10 concerning fractions. A level of performance can be established at 6.

The criterion or standard can be used as a competency/mastery score so that students who have obtained scores that are greater than 6 are termed competent or have mastered skills in a particular domain. Criterion-referenced interpretations generally indicate what an individual can or cannot do with respect to a specified domain of knowledge attitudes or skill.

Types of criterion-referenced scores

1. **Percent correct scores:** This is the percentage of items that a student got correct. For example, if a student obtained 8 marks out of 10, the percent correct is 80.
2. **Competency scores:** Students who obtained the cut-off scores are believed to have achieved a required level of competency. Cut-off scores should not be arbitrarily set. There should be a support or basis for them.
3. **Quality ratings:** This is the quality level at which a student performs a task. For example, a student can be rated as A for outstanding, B+ for excellent etc.
4. **Speed of performance scores:** These indicated the amount of time a student uses to complete a task or the number of tasks completed within a specified time. For example, a student may type 30 words in a minute or an athlete may run 100 meters in 11.5 seconds.

Uses of criterion-referenced interpretations

1. **Certification decisions:** certificates are needed in several areas of work to demonstrate the acquisition of skills and knowledge. Criterion-referenced scores provide information about

whether an applicant has the required level of skill or not and certificate of achievement attest to this.

2. **Minimum competency decisions:** certain curricula are structured such that a student needs to achieve a certain level of competency before moving on to a higher level. Criterion-referenced scores are used to determine whether a student or a class can move on to a higher level of study.
3. **Diagnostic decisions:** criterion-referenced scores help the teacher to discover the learning difficulties of the pupils. They help the teacher to diagnose and know which topic or learning targets have not been grasped. It helps the teacher to provide individual or class learning activities that will best adapt to students' requirements and thereby maximize their opportunities to attain chosen learning targets.
4. **Placement decisions:** criterion-referenced scores provide information as to whether a student can succeed in a programme or not. For example, to determine whether a person can be a medical doctor or not, a test can be given such that performance on the test can determine whether the individual has the pre-requisite skills to succeed in the medical programme.
5. **Programme evaluation:** criterion-referenced scores provide information about national progress in education. The performance of students can indicate whether a particular curricular is successful in its implementation or not. In Ghana, criterion-referenced scores were used in the 1990s to assess the level of mathematics and English literacy.

COMPATIBILITY OF BOTH MEASURES

1. Both norm-referenced and criterion-referenced scores can be used for guidance and counseling decisions.
2. Scores on a single test can be given both norm-referenced and criterion-reference interpretations. A score of 80 may indicate that the student is the best in the class and may also indicate that the student reached a level of high competency in the material taught.
3. Both scores can be used for summative evaluation. At the end of the instructional unit, a teacher can test to find out if the students have generally grasped the concepts taught. The teacher can also determine which students have performed below the lower quartile and may need remedial actions.