❏   109

Test, measurement, and evaluation: Understanding and use of the concepts in education

Dickson Adom[1], Jephtar Adu Mensah[2], Dennis Atsu Dake[3]
[1]School of Economic Sciences, North West University, South Africa
[2]CESA Higher Education Cluster, Association of African Universities, Ghana
[1,3]Departement of Educational Innovations in Science and Technology, Kwame Nkrumah University of
Science and Technology, Ghana

Article Info	ABSTRACT
	Test, measurement, and evaluation are concepts used in education to explain how the progress of learning and the final learning outcomes of students are assessed. However, the terms are often misused in the field of education, especially in Ghana. The objective of the study was to thoroughly explain the concepts to assist educationists and researchers in the field of education to better apply them in educational discourses. The study also suggests best practices in setting test items in measuring students' learning outcomes while showing policy directions to assist educationists and researchers in the field of educational evaluation.

Corresponding Author:

Dickson Adom,
Department of Educational Innovations in Science and Technology,
Kwame Nkrumah University of Science and Technology,
P. O. Box NT 1, Kumasi-Ghana, Ghana.
Email: adomdick2@gmail.com; 37069888@nwu.ac.za

1. INTRODUCTION

In the teaching-learning environment, there is a constant need to gauge the outcome or the quality of responsiveness of the teaching and learning process [1]. This important symbiotic process generally referred to as assessment, does not only occur after teaching but can also be undertaken before teaching is affected or during the teaching process. More specifically, concepts of test, measurement, and evaluation continue to dominate educational practice around the world. Though several scholars have advanced multiple interpretations, definitions and clarifications to these important educational concepts [2], the temptation to misconstrue one construct for the other have been a regular occurrence for student-teachers, educationists and even academics. In other words, these concepts have more often than not been erroneously used synonymously by practitioners to mean the same thing [3, 4]. As professional educators, this is unacceptable to the extent that our ability to distinguish these concepts and appropriately apply one or more within a given context is an important component of a teacher's professional practice. More so, depending on the nature and stage at which it is conducted, teachers have over the years applied different types of assessments for varied purposes. This study contends that, until classroom teachers have an appropriate appreciation of the nature of tests, measurement, and evaluation, an effective educational assessment will remain a mirage. Thus, this study will attempt to provide an overview of tests, measurement, and evaluation and explain the uses of these key co-dependent concepts in relation to educational practice. To this end, some important concepts

related to educational assessment have been thoroughly discussed in this paper to provide further understanding, draw the fine distinctions and outline the purposes among these concepts.

2.   RESEARCH METHOD

The researchers employed document analysis [5] and desk survey [6] for the comprehensive review of test, measurement, and evaluation in peer-reviewed journals, reports, and newspapers [7]. A systematic search was conducted using the keywords 'test and test item construction', 'evaluation in education', 'measurement', and 'assessment of learning outcomes' from online databases such as Scopus, Springer, JSTOR, PubMed, EBSCO, ProQuest and Google Scholar. A total of 48 articles were thoroughly assessed. The articles were carefully reviewed and scrutinized to wholly comprehend [8] the concepts of test, measurement, and evaluation and how they are used in educational research, especially in the Ghanaian context. The key qualities in the interpretive document analysis that guided the review were authenticity, credibility, and representativeness [9]. The papers were read severally to fully understand their theoretical perspectives [7]. The main ideas in the reviewed materials were summarized and discussed with the main objectives of the study in view. The new understanding was subjected to verification to validate the claims, assumptions, and theories made by scholars [10]. Finally, a captivating discussion on the concepts and how they are used in assessing the academic performances of learners was presented.

3.   RESULTS AND DISCUSSION
3.1.   Explaining the concepts: test, measurement, and evaluation in education

With an illustration of three concentric circles, Lynch [4] provides a conceptual framework as the basis for understanding the inter-related constructs of evaluation, measurement, and testing. Figure 1 is the schematic representation of the constructs of evaluation, measurement, and testing as applied in education. The conceptual framework sought to illustrate the superordinate-subordinate relationship between these concepts and demonstrate the areas of overlap.
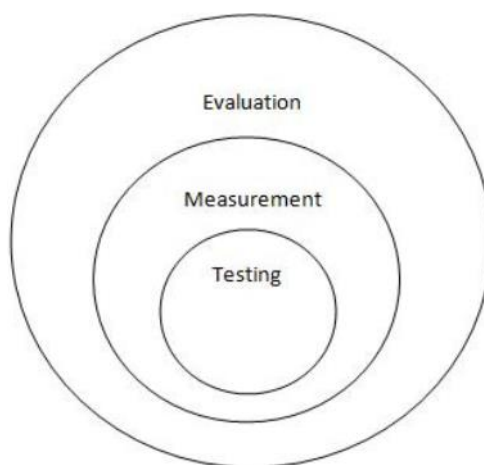

Figure 1. Lynch's model of evaluation, measurement, and testing [4]

From Figure 1, measurement and testing can be seen as a component of the evaluation. Bachman [11] and Lynch [4] in their postulation of evaluation agree that evaluation is the superordinate term to both measurement and testing. Bachman [11] adds that measurement encompasses testing when decision-making is done through the use of a specific sample of behavior. For this study, further exposition to the concepts has been provided in the subsequent section.

3.1.1.   Tests

One of the most commonly used assessment tools in education is to conduct tests. Beyond being considered as an instrument, tests can also be seen as standard procedures used to systematically measure a sample of behaviour by posing a set of questions [12]. Tests are designed to measure the quality, ability,

skill or knowledge of a sample against a given standard, which usually could be deemed as acceptable or not. In educational practice, tests are methods used to determine the students' ability to complete certain tasks or demonstrate mastery of a skill or knowledge of content. Tests can take the form of multiple choices or a weekly spelling. Manichander [13] adds that, although tests have been interchangeably used to mean assessment or even evaluation, the distinguishing factor of a test is the fact that is a form of assessment.

Braun et al [14] conjecture testing as the process of measuring single or multiple concepts, under a set of predetermined conditions. They are used to measure the level of students' learning. Tritschler [15] explains a test to mean administering a given tool or undertaking a procedure to solicit students' responses as information, which provides the basis to make judgement or evaluation regarding some characteristics such as skills, knowledge, and values. Three types of tests have been identified by Skinner [16], which can be used in determining a student's progress against the set objective(s). Tests can take the form of standardized tests, diagnostic tests, and teacher-made tests. Diagnostic tests (also referred to as analytic tests) are tests used by the teacher to get evidence detailing the learners' progress about a given subject. To undertake this, the teacher approaches this during the learning process by breaking the subjects into units. Since teachers adapt their teaching methods in their schemes of work, teacher-made tests are made by teachers. Consequently, teachers are at liberty to customize these tests. The advantage of a teacher-made test over standardized tests is that it allows further specific and individualized evaluation. However, a downside to teacher-made tests is its ineffectiveness in determining certain parts of objectives like skills of speaking and reading. Deducing from the preceding explanations, a test can be understood as a method or tool administered to measure the levels of knowledge, ability, and skills of learners. This means that there is some performance or activity required of either the learner or the teacher or both. Moreso in formulating tests, there is the need to attach the approach to the method whereby deliberate efforts must be directed towards striking the fine balance so that the items are neither too difficult nor too simple. That way, learners will be motivated to participate.

3.1.2. Measurement

Just like tests, multiple definitions are ascribed to the concept of educational measurement. Generally, measurement has to do with the assignment of quantifiable data by using one or more instruments such as a test or rating scale. When contextualized within education, a measurement can be referred to as a process used to glean the degree of an individual's competence in numerical terms. In other words, measurement is undertaken to quantify the level of knowledge or skills acquired by a learner. Tripathi and Kumar [17] in quoting the definition provided by James M. Bradfield state that measurement "is a process of assigning symbols to the dimensions of a phenomenon to characterize the status of the phenomenon as precisely as possible". This means that measurement entails subjecting a phenomenon or variable to some precise and quantifiable yardstick(s). Scriven [18] similarly avers that measurement is undertaken to determine the magnitude of a quantity. This determination typically is carried out on either a criterion-referenced test scale or on a continuous numerical scale. These measurement instruments can take a variety of forms such as a questionnaire, a test or any piece of apparatus. The observer in certain situations can be used as the measurement instrument which will need to be calibrated or validated [18]. Scriven further notes: "Measurement is a common and sometimes large component of standardized evaluations, but a very small part of its logic, that is, of the justification for the evaluative conclusions".

Kizlik [3] conceptualizes measurement as the process of determining the attributes or dimensions of some physical object. The measurement process involves gathering information to monitor students' progress and possibly intervene should the need arise. The concern of measurement is with the application of its findings, thus calls for some judgement on the effectiveness or desirability of a product, process or progress in line with a set of generally acceptable objectives or values. From the expositions this far, an educational measurement can mean the standard procedures and the principles underpinning the application of the procedures used for tests.

3.1.3. Evaluation

In simplistic terms, making judgement or determination of the quality or worth about an object, subject or phenomenon can be referred to as evaluation. Relating the concept to education, Coleman [19] defines evaluation as the "determination of how successful a programme, a curriculum, a series of experiments, etc. has been in achieving the goals laid out for it at the outset".Other terminologies used synonymously as "Evaluation" or other variants of the same include but not limited to: appraisal, analyses, assessment, critique, examination, grading, inspection, judgement, rating, ranking, and review. According to Braun et al. [14], evaluation is the process of reaching conclusions regarding abstract entities. These intangible units can range from curricula to institutions. Thus, evaluation calls for undertaking a process to provide information to be used as a basis for judging a situation. An evaluation has to do with the procedures

employed for determining whether or not the learner meets a preset criterion. Evaluation in the real sense refers to the process used to determine the merit, worth, or value of a process or the product of the process [18]. Assessment tools such as tests are used during the evaluation process to determine the qualification based on set criteria [20]. This means that the process of making judgments is based on criteria and evidence. Evaluation refers to the systematic acquisition of information and consequent assessment so that some useful feedback is provided regarding an initiative [21]. With a well-undertaken evaluation, learners are enabled to reflect and hence are assisted to identify changes for the future. Although there are countless dichotomies ascribed to the forms of evaluation, there are two main types; formative and summative evaluation. Usually, at the planning and designing phase of an educational programme, the formative evaluation is conducted. This is done for soliciting immediate feedback for the given programme to modify and improve should the need arise. It is on-going and it helps to determine the programme strengths and weaknesses. In agreement with this assertion, The Glossary of Education Forum [22] considers formative evaluation as an in-process evaluation of student learning. They state further that formative evaluations are typically administered multiple times during a unit, course, or academic program. This type of evaluation involves the teacher giving and making a series of tests and exercises, adding, averaging the marks and entering them on a report card.

On the other hand, Baehr [23] states that summative mean "addition of all things" Summative evaluation is concerned with the evaluation of an already completed programme. Evaluation is what is obtained at the end of a course that is used to determine whether students have mastered the course objectives [24], and the evaluations may be based on tests and other assessment procedures. When all that has been planned and done, summative evaluation can be carried out to determine whether the programme has achieved its goals. Simply, it is the kind of evaluation that summarises the strengths and weaknesses of a programme. Singh [25] puts forward the purposes of regularly undertaking educational evaluation. He posits that evaluation in education is purposed for making reliable decisions concerning educational planning, used for ascertaining the worth of time, and to identify students' growth or otherwise in acquiring desirable knowledge, skills, attitudes, and societal values. Other reasons are to enable teachers to determine the efficacy of their instructional techniques and learning resources as well as to motivate learners to discover their progress in accomplishing given tasks. It is crucial to take cognizance of and follow the principles that underlie the evaluation process for meaningful outcomes.

3.2. Best approaches in setting test items to measure and evaluate students' learning outcomes: cognitive, affective and psychomotor areas of development

One core responsibility of a teacher is to assess the amount of learning done by students or their achievement at the end of a course unit, or an instructional period and provide feedback to key stakeholders in the form of grades [26]. In the course of teachers discharging their assessment responsibilities, they provide essential feedback on students' progress and also contribute to improving the learning process [27, 28]. One effective tool that is mostly used by teachers in assessing the quantity and quality of learning done is a test. A test connotes the presentation of a standard set of questions to be answered by the learner [29]. Crooker and Algina [30] further describe a test to be a standard procedure for obtaining a sample of behaviour from a specified domain. In other words, a test is an instrument comprising of well-crafted items that in totality measures realistic learning outcomes that represent expected behavioural trait(s). In the classrooms, students learn varieties of content, and teachers are required to assess students' knowledge on these contents and summarize them in the form of alphabetical or numerical code thus grades [31]. The assigned grades as an outcome give the institutions an independent indication of the achievement/ability level of a given student. At times it guides our choices regarding who to pick for our university, which programme or to determine which ones need extra help to be successful. Considering the influential role that the evaluation of test scores play in decision making among stakeholders, it is crucial to suggest that both test developers and users must make conscious effort to improve the validity and the reliability of the test to get objective information by minimizing errors in measurement [32]. We suggest that in testing what students know or have learned in an area of study, well-crafted test items ought to be used and must match intended learning outcomes. When there is an alignment of assessment with learning, teaching and content knowledge, test scores turn to be valid [33]. Learning outcomes is a practical way of maintaining standards and improving teaching [34]. Etsey [35], suggest that a complete learning objective should include an observable behaviour, conditions under which the intended behaviour must be manifested and the level of performance considered to be sufficient to demonstrate mastery. Learning outcomes help in assessing knowledge and concepts that point to the total development (cognitive, affective and psychomotor) of students [36]. However, assessing learning outcomes on the psychomotor and affective domain maybe of a challenge to some specific courses and to include them may unnecessarily increase the number of outcomes [34].

3.3. Classroom achievement tests

Classroom achievement tests are generally teacher-made tests [35]. These tests are constructed by teachers to test the amount of learning done by students and it is often done formatively or summatively. Teacher-made tests usually measure attainment in a single subject in a specific class or form or grade [36]. Teachers are empowered by institutional policies to assess the amount of learning done after a stipulated period of instruction. In Ghana, the School-Based Assessment (SBA) is used as a guide by basic and Senior High teachers when assessing students' learning. For teachers to be knowledgeable and efficient in their assessment practices, teachers are taken through a full course in educational assessment of which test construction is a key component [37]. Teachers who have received training in the assessment are expected to have the propensity to employ the various assessment techniques correctly when assessing students learning. This will help them in ensuring that teachers can craft their test items to measure students' learning. When teachers are equipped with the relevant content procedures of classroom assessment, it evaluates whether a student's learning effective [38]. Despite the importance of classroom assessment, studies suggest some deficiencies in teacher-made tests [37, 39]. According to Lane et al. [40], most teachers craft flawed items that measure the ability to recall basic facts and concepts. Teachers also have a negative attitude toward test construction practices, which make them, perceive it as a tedious task to undertake in schools [37]. To mitigate errors in the construction of test items, researchers recommend several guidelines that need to be observed. Tamakloe, Atta, and Amedahe [41] and Etsey [42] suggested an eight steps approach to the construction of test items. The test developer should first define the purpose of the test; determine the item format to use; determine what is to be tested; write the individual items; review the items; prepare the scoring key; write directions, and evaluate the test. From the perspective of Quansah and Amoako [38], the construction of test items should follow four broad categorizations thus planning, item construction, review, and assembling.

3.4. The planning stage

Developing a good test is like target shooting. Hitting the bull's eye requires much attention and planning; you must focus on the target, select an appropriate arrow, and take careful aim. In simple words, developing a good test requires comprehensive planning. The planning stage provides a systematic framework that highlights major activities that emphasizes test security and quality control procedures from the onset [39]. Hence, the planning stage is very crucial and should be given the needful time and attention. Teachers should not be in a haste to construct test items without any kind of planning because for constructed test items to relate in a meaningful fashion with intended learning outcomes, it required extensive planning. According to Lane et al. [40], the fundamental questions to be addressed in this phase are: What is the construct to be measured? What is the population for which the test is intended? Who are the test users and what are the intended interpretations and uses of test scores? What test content, cognitive demands, and format will support the intended interpretations and uses? Fairness should also be considered in the overall test plan because it is a fundamental validity issue [42]. In determining the purpose of the test, a test can be used to serve several purposes, such as judging the mastery level of intended skills and knowledge, measuring progress over time, diagnosing pupil difficulties and misconceptions about a course as well as ascertaining the effectiveness of the curriculum [29]. Decisions on the construct domain and degree to be assessed thus the Knowledge, Skills, and Attitudes (KSAs), are considered when preparing a table of specification. The table of specification is a two-way chart which maps instructional objectives with the course or subject contents [43]. It helps in ensuring that instructional objectives and the test items are congruent which increases the likelihood of obtaining more valid test scores. Test scores are considered to be valid when there is enough evidence to support their interpretations and use. When teachers ensure that there is a marriage between what is taught and what is been tested, it helps in gathering validity evidence based on test content. Content validity is the degree to which test items are considered to be a representative sample of topics considered during the instructional period [44].

3.5. Constructing a table of specification

The most widely used method in obtaining validity based on content evidence is through the construction of a table of specifications. The construction of a table of specification helps in improving the degree of domain representation [45]. It serves as a crucial guide for item development and showcases the level of educational domain been assessed. The purpose of a *table of specification* is to identify the achievement domains being measured and to ensure that a fair and representative sample of questions appears on the test. It thereby provides the link between teaching and testing [46]. After considering, the total test items, the preparation of a specification table helps to avoid overlapping in the construction of the test items, helps to determine the weighting of learning outcomes regarding content areas and ensures that justice is done to all parts of the course. Although a table of test specifications is no guarantee that the errors in test

items will be corrected, such a blueprint help improve the content validity of teacher-made tests [38]. One simple method to go about a table of specification is to create a table with the content areas along the side with the domain levels covered by the test on the top. Each cell in the table corresponds to a particular task and subject content. By specifying the number of test items you want for each cell, you can determine how much emphasis to give each task and each content area. Table 1 presents a sample of the table of the test specification.

Table 1. Table of test specification

Instructional Objective Contents	Categories of Cognitive domain						
	Knowledge	Comprehensive	Application	Analysis	Synthesis	Evaluation	Total
Total							

In preparing a table of test specification, the test developer, in this case, the teacher must first list all content taught in the unit/course; assign corresponding numerical weighting to each topic; decide on the item format; decide on the number of items to be constructed for each topic; decide on the type of question under the different cognitive learning domain. In assigning a numerical weighting to each topic, the instructor must consider how relevant the topic is and the volume of its content in terms of teaching. Table 2 presents a developed sample of a test specification table.

Table 2. Sample of table of test specification

Instructional Objective Contents	Knowledge (25%)	Comprehensive (10%)	Application (25%)	Analysis (15%)	Synthesis (10%)	Evaluation (15%)	Total
Water	2			1		1	6
Electrical Energy	1	1	2		1		5
Force & Pressure		1		1		1	4
Machines	2		3	1	1	1	8
Total	5	2	5	3	2	3	20

Table 2 has knowledge 25%, comprehension 10%, Application 25%, Analysis 15%, Synthesis 10% and Evaluation 15%. The moment instructional objectives have been identified, a test blueprint is developed linking both the content and behavioral objectives as shown in the table above. A table of specifications of this kind helps to ensure that the test has content validity in terms of covering all the objectives of instruction.

3.6. Deciding on item format

The decision on the ideal item format to used is influenced by several factors. Among them include the purpose of the test, content coverage, ease of scoring, the number of students to be tested, the skills to be tested, the difficulty level desired, the physical facilities available for reproducing the test, the age of the students and the teacher's skill in writing the different types of items [38]. The most recognized item format in classroom achievement testing is the essay and the objective types. Most teachers in Colleges of Education often use objective type tests in assessing students [47]. However, Etsey [42] avers that it is sometimes necessary to use more than one item format in a single test. The rationale has been that certain item formats are more suitable than others in measuring specific learning outcomes. For example, an essay question will allow a student to demonstrate in-depth knowledge and measure outcomes such as critical writing. On the other hand, essay questions are relatively more time consuming to score and difficult to control subjectivity hence greater efforts are needed to ensure inter-scorer and intra-scorer reliability [48]. In summary, when planning an achievement test, a teacher has to consider the feasibility of a specific item format taking into consideration the surrounding practical constraints.

3.7. Item construction stage

The process of developing well-crafted items is indeed a complex task. A great deal of decision needs to be made to increase the likelihood of meeting the criteria of a good test. The test developer has the responsibility of developing test items that measure the intended construct. Though experts in educational

measurement over the years have developed basic principles and suggested guidelines that need to adhere to when constructing test items for teacher-made tests [35, 40,], effective item writing has become a skill that must be learned and practiced by test developers [39]. Most novice teachers (item writers) create flawed items that measure the ability to recall basic facts and concepts [39]. Since test items are the building block for all tests, the methods and procedures used to produce effective items are a major source of concern when determining the psychometric properties of a test. The process of writing good test items is not simple. It requires time and effort [38]. Therefore, teachers must strive to match test items to the desired instructional outcome. Regardless of the item format used, there are basic principles that need to be adhered to when constructing test items [35]. Mehrens and Lehmann [39] and Etsey [35] suggested the following guidelines:
- The table of specifications must continually be referred to when writing test items.
- The test items must be related to and match the instructional objectives.
- Well-defined items that are not vague and ambiguous must be formulated.
- Grammar and spelling errors must be checked.
- Textbook or stereotyped language must be avoided.
- Excessive verbiage and complex sentences must be avoided
- The test items must be based on information that students should know.
- More items than are actually needed in the test must be prepared in the initial draft. Mehrens and Lehmann [39] suggested that the initial number of items should be 25% more.
- The items and the scoring keys must be written as early as possible after the material has been taught.
- The test items must be written in advance (at least two weeks) of the testing date to permit reviews and editing.

Adhering to these recommended principles will enhance the validity and reliability of test scores by minimizing errors.

3.8. Item review stage

After the items have been written, the next stage is to evaluate them. At this stage, Etsey [35] suggests that the items must be critically examined at least a week after writing them. The evaluation of written items can also be done by allowing fellow teachers or colleagues in the same subject area to review the test items. The evaluation of test items can also be done statistically. The statistic approach involves using statistical analysis to determine how good an item is in terms of difficulty, how it discriminates among test takers and the strength of its distractors. Item analysis is one statistical analysis often used in evaluating test items. It refers to the process of collecting, summarizing and using information from students' responses to decide on each assessment task or item [49]. To do this, the crafted test items after a series of review and editing is given out to a representative sample of students who possess similar characteristics to the intended test takers for them to respond to each test items, to help judge the quality of the item. The purpose is to determine if items function as intend; difficulty level and how distractors of each item function. It should be noted that the use of statistical item difficulty or item difficulty indexes by the classroom teacher seems impracticable to a large extent [40, 50]. This is because statistical item difficulty data are always gathered after test administration or test try-outs and teacher-made test items are usually not pre-tested. However, Mehrens and Lehmann [39] recommended that subjective judgement must be relied on to determine the difficulty level of items. This could be done by categorizing the test items as difficult, average or easy. In brief, the item review stage serves the purpose of removing or rewording poorly constructed items, checking for technical errors and irrelevant clues. After reviews and editing, the test items can now be assembled.

3.9. Assembling stage

After the evaluation of crafted test items by considering both statistical and grammatical errors, the approved test items are assembled and prepared for administration. In assembling test items, the following points must be considered [35, 38, 40]:
- The items should be arranged in sections by item formats.
- The items must be spaced and numbered consecutively so that they can easily be read.
- A definite response pattern to the correct answer must be avoided.
- Within each section or format, the items must be arranged in order of increasing difficulty.

One way of achieving this is to group items in each format according to the instructional objectives being measured and make sure that they progress from simple to complex. The categorization of test items according to topics has the advantage of helping the teacher to ascertain which learning activities appear to be most readily understood by students, those that are least understood and those that students have a misconception on [38]. Experts in educational measurement and evaluation recommend that test items of

lengthy or timed tests should progress from the easy to the difficult, if for no other reason than to instill confidence in the examinee, especially at the beginning [35, 38].

3.10. Bad practices in test item construction in educational institutions in Ghana

One core responsibility of the classroom teacher or an instructor is to determine the extent to which learning outcomes have been achieved. For one to effectively quantify the attained instructional objectives on the part of the learner, competency in test construction cannot be overlooked. The competency in test construction is crucial for effective evaluation of learning and instructional objectives [37]. Unfortunately, scholars have argued that test construction practices among teachers in Ghana, is not encouraging [26, 35, 40, 51]. The implication is that teachers may end up taking inaccurate information about student learning which may be misleading. When test items are poorly crafted in the sense that it does not accurately measure the intended learning outcomes and is not aligned to teaching activities, it possesses a great challenge as students' achievement scores are likely reported with errors. Challenges in testing practices have been an issue across countries of which Ghana is not exempted. In Ghana, Amedahe [51] in a study of the assessment practices of secondary school teachers in 18 secondary schools in the Central Region found that teachers lacked the skills and principles of test construction. Hence, their proficiency in assessment practices was not adequate to meet classroom needs. The study revealed that test items constructed by teachers were prone to error and mostly measures the knowledge level of cognitive processes. For instance, where test items solely focus on recall, it encourages students to engage in rote learning. In a similar vein, Anywhere [52] also revealed that teacher training college tutors do not follow the basic principle of testing in the construction of teacher-made tests or classroom tests and that they perceived the management of assessment in the colleges as a workload to their teaching activities. When teachers have such negative perceptions, they are likely to consider test construction as a major source of anxiety. Anywhere further identified no significant difference in test construction, practices between teachers concerning their teaching experience.

On the other hand, Quaigrain [53] found out that the majority of teachers do planning when constructing essay-type tests. However, teachers did not comprehensively adhere to the basic prescribed principles in classroom test construction. He further suggested that most teachers do not review constructed items. Hence, the majority of the items look ambiguous. A test item is considered to be ambiguous when a statement or word has two or more meanings. For example, in essay tests, words such as discuss or explain may be ambiguous in that different students may interpret these words differently. The ambiguous question has the possibility of affecting the reliability of a test. The use of excessive wording contributes to difficulty in teacher-made tests. Too often teachers think that the more wording there is in a question, the clearer it will be to the student. This does not always happen. The more precise and clear-cut the wording, the greater the probability that the student will not be disorganized. Sasu [54] in the study of assessment practices of basic school teachers in TEN Junior High Schools in the Central Region found that teachers did not consider the meaning of words against different ethnic backgrounds of their students when constructing test items. When teachers fail to consider the meaning of words against the different ethnic backgrounds, the interpretation made from the test may lead to faulty conclusions [55]. The possible cause of teachers not considering the meaning of words against the ethnic background of students may be as a result of the limited time and excessive workload on teachers, which may lead them to give less attention to the wording of test items with little consideration to students' ethnic background. It could also be that teachers do not consider the evaluation of test items. The study further revealed that teachers often asked colleagues who are not in the subject area to help them construct test items. This attitude might have a great deal of implication on the validity of test results. This is because a test constructed by such a teacher might not appropriately measure the real achievement of the students since the test items are likely not to cover the content and thinking processes required.

Moreover, the Curriculum Research and Development Division [56] studied student assessment procedures in Junior Secondary Schools across 11 districts in the country and found that teachers did not have adequate training in the management of assessment practices. This limitation in skills was due to their inability to receive training in assessment practices. It was reported that the majority of the teachers were not confident enough when it comes to the assessment of students' achievement hence, replicating assessment practices they experience when they were students. Conversely, Quansah and Amoako [37] found that SHS teachers in the Cape Coast Metropolis irrespective of their knowledge in classroom assessment have a negative attitude towards test construction. Such a negative attitude could be another factor that accounts for the poor construction of test items among teachers. Teachers likely know test construction but their attitude prevents them from utilizing the knowledge they have. Test construction, we might say, is a difficult and rigorous task if teachers are supposed to do it effectively [49]. Hence, a negative attitude by teachers

toward test construction could explain why teachers irrespective of their knowledge in classroom assessment practices still construct poor test items or perhaps, replicate already existing test items.

3.11. Policy direction: workshop and training on sound practices in test, measurement, and evaluation in education

Testing in education is an invisible thread that cannot be underscored when evaluating teaching and learning. In the absence of testing, there is no teaching and learning. Considering the integral role that test plays in the teaching profession and as an assessment tool that is widely used in Ghanaian schools. We suggest that in-service training should be organized in the form of workshops and seminars on a school basis for teachers within an educational circuit for their testing practices. The rationale is to assist teachers to improve their testing practices. This could be achieved through the collaboration of the Ministry of Education, the Colleges of Education and other stakeholders of education. From the office of the Educational Directorate, a sensitization program should be roled out on regular basis by Lead teachers, Circuit supervisors and staff from the Education office, on the importance of their testing practices regarding test construction. Teachers should be educated on the implication of their testing practices and their effect on the validity and reliability of test scores.

A guideline on test construction should be designed by the Curriculum Research and Development Division in collaboration with tertiary institutions such as the University of CapeCoast and the University of Education, Winneba which could be made available for teachers to guide them in the test construction. As more training programmes through seminars and workshops are organized for teachers, stakeholders should be aware of the fact that the training alone does not bring about the application of competencies gained but also their attitude towards constructing the test. It is recommended that teachers should not only be trained in constructing test items but should also be enlightened on the need to adhere strictly to testing procedures. Ghana Education Service (GES) together with headteachers of various SHS should ensure effective supervision of teachers in constructing tests for students.

4. CONCLUSION

This study brings to the fore a growing need to constantly reexamine the concept of educational assessment as it has proven over time to be an evolving one. Given that there exist a barrage of scholarly publications regarding the concepts of measurement, testing, and evaluation in the educational context, the concepts remain difficult to be understood by educational researchers and educators. However, there is widespread agreement that evaluation with its components of measurement and testing is fundamental to the educational practice. This study clarifies the concepts with a detailed explanation of their respective applications from the Ghanaian perspective with the hope that stakeholders in the educational enterprise will be better equipped for effective educational practice.

REFERENCES

[1] Masters, G., Getting to the essence of assessment, 2014. Retrieved on October 2019. [Online]. Available: https://rd.acer.org/article/getting-to-the-essence-of-assessment
[2] Mahmoodi-Shahrebabaki, M., Assessment, "Evaluation, and Testing: What are the Differences?" Unpublished manuscript, 2018.
[3] Kizlik, R. J., Measurement, assessment, and evaluation in education, 2012. Retrieved on October 2019. [Online]. Available: https://www.adprima.com/measurement.htm
[4] Lynch, B. K., "Rethinking assessment from a critical perspective," *Language Testing*, vol. 18, no. 4, pp. 351–372, 2001.
[5] Bowen, G. A., "Document Analysis as a Qualitative Research Method," *Qualitative Research Journal*, Vol. 9, No. 2, pp. 45-65, 2009.
[6] Travis, D., Desk Research: The What, Why and How, 2016. [Online]. Available: https//:www.userfocus.co.uk
[7] Creswell, J.W., *Research Design* (3rd ed). United States: SAGE Publications Inc, 2009.
[8] Denzin, N. K. & Lincoln, Y. S., "Introduction: Entering the Field of Qualitative Research," in: NK Denzin and YS Lincoln (eds.) Handbook of Qualitative Research. Thousand Oaks: SAGE Publications, 1994.
[9] Hefferman, C., Qualitative Research Approach, 2013. [Online] Available: http://www.explorable.com

[10] Peshkin, A., "The Goodness of Qualitative Research," *Educational Researcher*, Vol. 22, No. 2, pp. 23-29, 1993.

[11] Bachman, L.F., *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press, 1990.

[12] Linn, R. L., *Measurement and assessment in teaching*. Pearson Education India, 2008.

[13] Manichander, T., Assessment in Education, 2016. Retrieved on October 2019. [Online]. Available: https://www.Lulu.com

[14] Braun, H., Kanjee, A., Bettinger, E., & Kremer, M., "Improving Education Through Assessment. Innovation, and Evaluation," The American Academy of Arts and Sciences, 2006.

[15] Tritschler, K. A. (Ed.), *Barrow & McGee's practical measurement and assessment*. Lippincott Williams & Wilkins, 2000.

[16] Skinner, C, E., *Educational Psychology*, (4th ed). New Jersey; Prentice-Hall, 2002.

[17] Tripathi, R., & Kumar, A., "Importance and Improvements in the Teaching-Learning process through Effective Evaluation Methodologies," *ESSENCE Int. J. Env. Rehab. Conserv*. Vol. 9, no, 1, pp. 7-16, 2018.

[18] Scriven, M., *Evaluation Thesaurus* (4th ed.). Thousand Oaks, CA: Sage, 1991.

[19] Coleman, A. M., "The Dictionary of Psychology," *Applied Cognitive Psychology*, vol. 15, no. 3, pp. 349-351, 2001.

[20] Overton, T., *Assessing Learners with Special Needs: An Applied* Approach (7th ed.). Boston, M. A.: Pearson Education, 2013.

[21] O'Donnell, A. M., Reeve, J. & Smith, J.K., *Educational Psychology: Reflection for Action* (3rd ed.). N.Y. United States: John Wiley & Sons, Inc, 2011.

[22] Glossary of Education Forum, Assessment, 2015. Retrieved on October 2019. [Online]. Available: https://www.edglossary.org/assessment/

[23] Baehr, M., *Distinctions between Assessment and Evaluation*. Coe College, Faculty Guidebook, pp. 441-444, 2005.

[24] Bondai, B., Towards a comparative analysis of formative and summative evaluation in education. In Chemhuru, O, H. (eds) (2013) Issues and foundations in education. (pp. 141-146). Gweru: Booklove publishers, 2013.

[25] Singh, G., Unit-1 Concept and Purpose of Evaluation. IGNOU, 2018.

[26] Adu-Mensah, J., "The attitude of basic school teachers toward grading practices: Developing a standardized Instrument," *International Journal of Social Sciences & Educational Studies,* vol. 5, no. 1, pp. 52-62, 2018.

[27] Popham, W. J., *Classroom assessment: What teachers need to know,* (5th ed.). Boston: Allyn and Bacon Publications, 2008.

[28] Ndalichako, J. L., "Secondary school teachers' perceptions of assessment," *International Journal of Information and Education Technology,* vol. 5, no. 5, pp. 326-330, 2015.

[29] Mehrens, W. A. & Lehmann, I. J., *Measurement and evaluation in education and psychology,* (4th ed.). New York: Holt, Rinehart and Winston Inc, 2001.

[30] Crocker, L., & Algina, J., *Introduction to classical and modern test theory*. Ohio, USA: Cengage Learning Pub, 2008.

[31] Allen, J., "Grades as valid measures of academic achievement of classroom learning," *Clearing House: A Journal of Educational Strategies, Issues, and Ideas,* vol. 78, no. 5, pp. 218-223, 2005.

[32] Quansah, F., Amoako, I., & Ankomah, F., "Teachers' test construction skills in Senior High Schools in Ghana: Document Analysis," *International Journal of Assessment Tools in Education*, vol. 6, no. 1, pp. 1-8, 2019.

[33] Harlen, W., "Teachers' summative practices and assessment for learning: Tensions and synergies," *The Curriculum Journal*, vol. 16, no. 2, pp. 207-223, 2005.

[34] Biggs, J., & Tang, C., *Teaching for quality learning at university* (4th ed.). New York, USA: McGraw Hill, 2011.

[35] Etsey, Y. K., "Assessing performance in schools: Issues and practice," *Ife Psychologia,* vol. 13, no. 1, 123-135, 2004.

[36] Ababio, B. T., & Dumba, H., "Assessment of policy guidelines for the teachers learning of geography," *Review of International Geographical Education*, vol. 4, no. 1, pp. 1-5, 2013.

[37] Mehrens, W. A. & Lehmann, I. J., "National tests and local curriculum: Match or mismatch," *Educational Measurement: Issues and Practices,* vol. 3, no. 3, pp. 9-15, 2009.

[38] Quansah, F., & Amoako, I., "The attitude of Senior High School teachers toward test construction: Developing and validating a standardized instrument," *Research on Humanities and SocialSciences*, vol. 8, no. 1, pp. 25-30, 2018.

[39] Mehrens, W. A. & Lehmann, I. J., *Measurement and evaluation in education and psychology*. New York: Harcourt Brace College Publishers, 1991.

[40] Lane, S., Raymond, M. R., Haladyna, T. M., & Downing, S. M., Test development process. In S. Lane., M. R. Raymond., & T. M. Haladyna (Eds), *Handbook of test development* (pp. 3-18). New York: Routledge, 2016.

[41] Tamakloe, E. K., Atta, E. T. & Amedahe, F. K., *Principles and methods of teaching*. Cantoment, Accra: Black Mask, 1996.

[42] Etsey, Y. K., *Assessment in education*. Unpublished, University of Cape Coast, Cape Coast, Ghana, 2012.

[43] American Educational Research Association, *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association, 2014.

[44] Kolawole, E.B., *Principle of test construction and administration*. Lagos: Bolabay, 2010.

[45] Lennon, R. T., "Assumptions underlying the use of content validity," *Educational and Psychological Measurement,* vol. 16, pp. 294-304, 1956.

[46] Sireci, S.G., "The construct of content validity," *Social Indicators Research,* vol. 45, pp. 83-117, 1998.

[47] Alvares, K. K., Table of specification and test construction, 2013. Retrieved from Amedahe, F. K., *Testing practices in secondary schools in the Central Region of Ghana*. UnpublishedMaster's thesis, University of Cape Coast, Cape Coast, 1989.

[48] Alias, M., "Assessment of learning outcomes: Validity and reliability of classroom test," *World Transactions on Engineering and Technology Education,* vol. 4, no. 2, pp. 235-238, 2005.
[49] Nitko, J. A., *Educational assessment of students*. New Jersey: Prentice-Hall, 2001.
[50] Kubiszyn, T., & Gary, B., *Educational testing and measurement* (7th ed.). USA: John Wiley & Sons Inc, 2003.
[51] Amedahe, F. K., *Testing Practices in Secondary Schools in the Central Region of Ghana.* Unpublished Master's Thesis, University of Cape Coast, Ghana, 1989.
[52] Anhwere, Y. M., *Assessment practices of teacher Colleges of Education tutors in Ghana.* Unpublished Master's thesis. The University of Cape Coast. Cape Coast, Ghana, 2009.
[53] Quagrain, A. K., *Teacher-competence in the use of essay type tests: A study of the secondary schools in the Western Region of Ghana*. Unpublished thesis. The University of Cape Coast, Cape Coast Ghana, 1992.
[54] Sasu, E., O., *Testing practices of junior high school teachers in the Cape Coast Metropolis.* Unpublished Master's thesis, University of Cape Coast, Cape Coast, 2017.
[55] Tom, K., & Gary, D. B., *Educational testing and measurement: Classroom application and practice.* Hoboken, NJ: John Willey and Sons Inc, 2003.
[56] Curriculum Research and Development Division (CRDD), *An investigation into student assessment procedures in public Junior Secondary Schools in Ghana*. Accra: Ghana Education Service, 1999.