

Teachers' Test Construction Skills in Senior High Schools in Ghana: Document Analysis

Frank Quansah ^{1,*}, Isaac Amoako¹, Francis Ankomah¹

¹ Department of Education and Psychology, University of Cape Coast, UCC, PMB, Cape Coast, Ghana.

ARTICLE HISTORY

Received: 29 September 2018

Revised: 05 December 2018

Accepted: 11 December 2018

KEYWORDS

Content relevance,
Reliability,
Validity,
Representativeness,
Test specification

Abstract: Assessment, specifically test construction, forms a critical part of the teaching and learning process. This aspect of teachers' responsibility has been questioned by several authorities in contemporary times. The study explored the test construction skills of Senior High Schools (SHS) teachers in the Cape Coast Metropolis. Using a qualitative document analysis, samples of End-of-Term Examination papers in Integrated Science, Core Mathematics and Social Studies in three selected SHS in the Cape Coast Metropolis were randomly (Lottery method) selected. The assessment tasks on the sampled instruments were critically examined by experts in the area of Educational Measurement and Evaluation. The results revealed that the teachers have limited skills in the construction of end-of-term examination. This was evident as issues were found with the content representativeness and relevance of the test, reliability, and fairness of the assessment tasks which were evaluated. It was recommended that head teachers should take up the challenge of inviting resource persons from recognised academic institutions to organise workshops for teachers on a regular basis to sharpen their skills on effective test construction practices.

1. INTRODUCTION

In the management of schools in Ghana, teachers, schools' management and policymakers in the course of or after teaching and sometimes before classroom teaching need to make decisions concerning teaching and learning. These decisions are made based on information gathered from the students' learning. Generally, this information gathering procedure denotes assessment. Nitko (2001) explained assessment as a process of obtaining information which is used for making decisions about students, curricula and programmes, and educational policy. Assessment, therefore, involves the utilisation of empirical data on students' learning to improve programmes and enhance students' learning (Allen & Yen, 2002). Scholars have pointed out that assessment systems have a significant effect on learning characteristics and personalities as children become young adults and then adults (Crooker & Algina, 2008;

CONTACT: Frank Quansah ✉ fquansah99@gmail.com 📍 Department of Education and Psychology, University of Cape Coast, UCC, PMB, Cape Coast, Ghana.

ISSN-e: 2148-7456 / © IJATE 2019

Ecclestone & Pryor, 2003; Nitko, 2001). Thus, the effects of assessment can continue through a learner's life of formal learning.

In the school setting, a test is generally used as an assessment tool for obtaining information about students' learning. It should be made clear at this point that testing is a key component in educational assessment. In testing what students know or have learnt in an area of study, well-crafted test items should be used. Tamakloe, Amedahe and Atta (1996) described a test as a device or procedure for measuring a sample of an individual's behaviour in a specific learned activity or discipline. Crooker and Algina (2008) further gave a description of test to be a standard procedure for obtaining a sample of behaviour from a specified domain. These tests are normally administered to students after a period of instruction, if for achievement purposes. Considering the sensitive role that information from a test play in making educational decisions for students as well as management, it is important to say that both test developers and users must make conscious effort to improve the validity and the reliability of the test in order to get objective information that approximate the individual's true characteristic, which the test developer seeks to estimate.

Unfortunately, test construction role of teachers has been reported as a main source of anxiety, especially with teachers with few years of teaching experience (Ebinye, 2001). This anxiety, according to Ebinye (2001), largely stems from inadequate test construction skills of these teachers. Scholars have also argued that test construction among teachers is not encouraging (e.g., Amedahe, 1989; Hamafyelto, Hamman-Tukur & Hamafyelto, 2015; Kazuko, 2010). The implication is that teachers may end up taking inaccurate information about student learning. For instance, Ololube (2008), which assessed the test construction skills of teachers in Nigeria, found poor test construction skills among non-professional teachers. Another study by Onyechere (2000) found that most teachers construct poor items which actually failed to function as it was supposed to. Some teachers, acknowledging that they have weak test construction skills resort to past or already existing questions to assess students (Onyechere, 2000). Teachers in the Borno State, Nigeria, were also found to construct items with lower levels of cognitive operations.

Similar findings have been found in Ghana. Amedahe (1989), in his study, found that SHS teachers in the Central Region of Ghana have inadequate skills in constructing both essay and objective type tests. According to the Curriculum, Research and Development Division [CRDD] of Ghana Education Service (GES) (1999), Junior High School teachers all over Ghana are found to have inadequate competencies in testing practices. Etsey's (2003) supported the views of CRDD (1999) and stated that the Division of Teacher Education of GES should authorize curriculum planners in education within the country to make assessment courses compulsory and as well prioritise these courses in the first 2 years in teacher training colleges.

In Ghana, Quansah and Amoako (2018) found that SHS teachers in the Cape Coast Metropolis have a negative attitude towards test construction. The authors specifically found a poor attitude of teachers in the planning of test, item writing, item review and assembling of the items. Quansah and Amoako concluded that this attitude of teachers had an effect on the quality of test used for assessing students. It is of essence to state that the poor attitude might not be due to their inadequate skills but also from the fact that some teachers see test construction as a burden. Exploring the test construction skills of teachers is significant if objective and accurate information are to be gathered from students in the teaching and learning process.

Moreover, previous studies employed self-reported means to describe teachers' skills in test construction. This measurement procedure does not appropriately estimate the skills of teachers in test construction. Majority of these studies gathered their information through administering questionnaires to the respondents or by interviewing them. The mere asking of questions about how these teachers construct test items do not provide a comprehensive view of the skills

teachers have. It is even likely that these teachers would provide responses which do not reflect their actual practice. In actual sense, previous studies just provide information about teachers' testing or test construction practices through the lens of the same teachers. It is essential to conduct an exploratory study to critically examine some questions crafted by these teachers to find out whether they have the competencies in test construction. These crafted questions serve as the "end-product" of their skills which is being put to use. This paper, therefore, explores the skills of SHS teachers in the Cape Coast Metropolis. The paper, particularly, assessed the content of the documents (samples of examination questions) with regards to five hypothetical dimensions: (a) content representativeness and relevance; (b) thinking processes and skills represented; (c) reliability and objectivity; (d) fairness to different students; and (e) practicality.

2. METHOD

The research methodology for this study is qualitative document analysis. This study seeks to review and evaluate documents (Creswell, 2014). Just like other forms of the use of document analysis is to examine samples of previous examination papers in SHSs in the Cape Coast Metropolis in order to give meaning and understanding of teachers' test construction skills (Corbin & Strauss, 2008). Samples of End-of-Term Examination papers in Integrated Science, Core Mathematics and Social Studies in three selected SHS in the Cape Coast Metropolis were randomly (Lottery method) selected. These papers were used for summative assessment and thus, the questions were crafted by the classroom teachers in conjunction with the examination board of the school. This means that the papers went through some form of evaluation before they were administered. The question papers selected were papers between 2015 and 2018 in three subjects: Integrated Science, Core Mathematics, and Social Studies. Specifically, 5 samples of question papers were selected based on each subject from each school. In total, 15 samples of examination papers were taken from each of the three schools. In all, 45 samples of examination papers were sampled from the three schools.

The assessment tasks on the sampled instruments were critically examined by experts in the field of Measurement and Evaluation. The examination of the papers took four months. Before the questions were examined, the scheme of work of the various subjects selected was taken. There was also an interaction with the teachers on the areas which were covered for the term. We made a lot of effort to ensure the schools' anonymity, confidentiality and privacy in the data gathered. Consent of the teachers together with the examination committee was sought before the data was gathered. We employed the qualitative content analysis to analyse texts in the pre-defined dimensions.

2.1. Description of the Papers

The Integrated Science papers were for Form 1 (first year/grade 10) students in the selected Senior High Schools. All the Integrated Science papers, the assessment tasks were in two sections: paper A and B. Paper A carried 40 points whereas the paper B carried 60 points. The paper A consisted of 40 multiple choice items with four options and students were required to respond to all the questions. The paper B was the essay section which also had two parts: Part I and II. The part I was a practical compulsory question which had five sub-sections. Part II of the paper B had four questions of which the students were required to answer only two. For all the Integrated Science papers, the examinees (students) were required to answer the question within a two-hour duration.

For the Core Mathematics papers, the samples were taken from Form 2 (second year/grade 11) students in the selected Senior High Schools. For all the Core Mathematics papers, the test comprised two sections (i.e., A & B). Section A comprised 40 multiple choice items and section B had seven essay type questions with their sub-questions where students were required to

answer five items. Question one of Section B was always a compulsory question, however, students were to choose other four) from the other six questions.

The Social Studies papers were also made up of two sections: Section A and Section B. The first sections comprised of 40 multiple-choice items which the students were required to use 50 minutes in responding to it. The second section had four essay questions which students were required to select three. Each of the essay questions carried 20 points. The paper was for Form 2 (second year/grade 11) students.

3. RESULTS AND DISCUSSION

The results from the examination of the papers are captioned into the following sections: content representativeness and relevance; thinking processes and skills represented; reliability and objectivity; fairness to different students; and practicality.

3.1. Content Representativeness and Relevance

After careful evaluation of the test instruments, it was evident that the test developer who is the subject teacher failed to sample adequately to cover all the content areas listed in the scheme of work for the relevant term. Analysis from the papers revealed that the content of these papers focused on a few of the areas taught. It was evident that the items on the instruments (tests) did not adequately sample the content taught. This implies that a student who attains 90% (Distinction) cannot be addressed as having adequate mastery of the content taught since he/she was not assessed in all the areas taught. Likewise, a student who obtains 35% (Fail) cannot be referred to as lacking mastery of the content taught. This is because the higher scoring student might have specialized in the areas which were sampled while the lower scoring student did not. It is possible that the higher scoring student lacks mastery over the three content areas which assessment instruments did not cover. The result from these assessment instruments can only be interpreted in terms of identifying the strength and the weakness of the students or how much students know in the few content areas assessed. The instruments, thus, lack some degree of content validity.

An assessment task which lacks content validity is likely not to reflect the important content, skills and learning outcomes specified in the school's or district's curriculum framework and content standards (Nitko, 2001). This is reflected in the test papers where emphasis was placed on fewer content areas. This is because the test reflected the learning outcome of those few content areas taught.

Again, some of the questions in the paper A (multiple choice questions) were measuring trivial knowledge. These questions demanded the lowest form of thinking such that the test-wise student who does not have any knowledge of the material can answer these questions correctly. Nitko (2001) argues that most worthwhile learning involves students' using a combination of skills and content rather than using isolated skills or bits of content. This suggests that the assessment instruments do not measure worthwhile learning to some extent. It is evident that the assessment instruments lacked, to some extent, content representativeness and relevance.

The second part of the papers (Section B) required examinees to answer two questions out of four questions provided. It must be indicated that these four questions have different difficulty level and require a different level of cognitive operation to be able to attempt answering them. While some of the questions in this part were measuring knowledge, others were measuring comprehension. Examinees may end up answering different questions. The implication is that as Joan decides to attempt the first two questions, Isaac would be attempting the first and third questions. Francis can decide to answer first and fourth questions. Emmanuel would be tackling the second and third questions while Samuel might also answer the second and fourth question. Therefore, Francis' score cannot be compared to that of his peer who did not answer the same

questions he answered. This affects the soundness of the interpretation and use of the students' assessment result because the performance of examinees who have answered different questions with different difficulty level can never be compared.

3.2. Thinking Processes and Skills Represented

An inspection of the test specification table for the multiple-choice items for all the papers indicate that the majority of the items measured lower-level skills. Thus, most of the items only required examinees to just recall facts. Few of the items measured comprehension and application. The essay part of the assessment instruments covered items on knowledge and comprehension. More specifically, all the sub-items on the compulsory questions were measuring knowledge. Surprisingly, it was found that the items in the essay part were measuring knowledge and comprehension with greater emphasis on knowledge.

The thorough account of the assessment tasks suggests that almost all the items were "recall" type of questions. This implies that examinees who engage in rote learning are those who will perform well and not necessarily those who have mastery over the material taught. Thus, the assessment instruments did not comprehensively assess different types of thinking skills. For an assessment result to be valid, the tasks should assess a student's ability to use strategies and processes that reflect how scholars in the discipline think (Nitko,2001). These assessment instruments deviated from Nitko's assertion. That is, only one lower-level cognitive process is greatly emphasized.

According to the Ministry of Education [MOE], Ghana (2012), the profile dimensions for an objective test for assessment should be 30% knowledge and 70% for comprehension and application. The assessment instruments did not meet the criteria given by MOE. The tests did not represent the kinds of thinking skills that the state's curriculum framework and performance standards suggest.

3.3. Reliability and Objectivity of the Test Items

The assessment instruments had a longer test length (based on West Africa Examination Council's (WAEC) standard) which is likely to increase the reliability of the results. This is supported by Nitko (2001) who argued that "longer assessments (with more task per learning target) are more reliable than shorter assessments" (p. 41). The paper A of all the instruments had 40 multiple choice questions with a point for a question. This part of the test will be scored objectively which will improve the reliability of the result. The second part of the assessment instruments, where examinees had to select some number of questions out of a lot, is likely to be scored subjectively which might affect the reliability of the test results.

Some of the multiple choice questions had problems in its structure (i.e., syntax error, faulty stem, grammatical errors, and ineffective distractors) and this is likely to affect the reliability of the test results. These flaws are likely to provide clues for the students to get the right answer to the stem. The grammatical error, for instance, might also give the examinees a different understanding of the question. These problems are likely to affect the consistency of the test result because the response to these items would not reflect what the examinees know.

The options to the multiple-choice items in all the papers were arranged horizontally which is likely to affect the reliability of test scores. This is because there is the likelihood that examinees might waste a lot of time reading the options to the questions. This affects slow readers in their attempt to respond to the questions. However, the time allowed was sufficient for an average examinee to answer all the questions required. Again, the options to the multiple choice questions were not alphabetically arranged and this might lead to some identifiable patterns in the key to the questions.

The assessment instruments were not formatted very well. The items were clumsy with poor spacing. In some of the multiple choice questions, options to the same questions had inconsistencies. While some of the options to the same started with a capital letter some of them started with small letters. Again, the font theme and size were not consistent from the instruction to the last question.

3.4. Fairness to Different Students

An evaluation of assessment tasks revealed that the tasks did not contain any form of information which gives a particular group of examinees advantage over others. This suggests that the assessment tasks were fair to the examinees with regards to gender, ethnic group, socio-economic background, among others. However, the clumsy nature of the assessment tasks might bring about unfairness to students who cannot read clearly when assessment tasks are clumsy. This was confirmed by Nitko (2001) that any assessment tasks must be fair to all examinees from all socio-economic background, ethnic group and language as well as students with disabilities who are mainstreamed in one class.

3.5. Practicality of the Assessment Task

A critical evaluation of the assessment papers found that the time allocated was enough and allows the examinees to appropriately respond to the items. Even though time was allocated for the essay part of the test, scores for each item in the essay section were not indicated. This might affect the reliability of the assessment results. This is because the time spent on a particular question depends on the score allocated to it. This is supposed to be done to ensure that examinees do not waste much time on questions with low scores. This was explained by Amedahe and Asamoah-Gyimah (2016) that practicality is concerned with the necessary material and time allotted to the test. They claim that a tester should consider the following questions: Will students have enough time to complete the test and are there sufficient materials such as booklets or answer sheets, tables, chairs etc. available to present the test to complete the test effectively? The critical evaluation of the papers seems to suggest that sufficient answer booklet and time were made available for students to complete the test effectively.

4. CONCLUSION and RECOMMENDATIONS

The evaluations of the tests obtained in the three core subjects revealed that teachers are weak in test construction. Even though some principles were done right, most of the critical issues which are related to validity and reliability were overlooked. This questions the validity of the results which would be awarded to these students. It is important for classroom teachers to be aware of the fact that the measurement of psychological constructs like academic achievement is a difficult thing to do. This is due to the complex and dynamic nature of human beings. However, there is the need for teachers to gather some information about students for decision making about curriculum, students and educational policy. This information is needed not only for teachers but also for parents, schools' management and policymakers. Because the information collected is used for decision making, it must be as accurate as possible. If a test with low validity and reliability are mostly used, then, inappropriate decisions are likely to be made.

The accuracy of classroom assessment results is very important but difficult to achieve. The complex nature of examinees, examination conditions, problems with test instruments and other factors reduces the validity of classroom assessment results. However, through careful planning of the test as well as adherence to principles in test construction, test assembling, test administration, scoring and result interpretation can help teachers to gather valid and reliable information about students. It, however, appears that some teachers do not have much knowledge in testing practices or do not simply adhere to the principles in testing. Although

Ghana, as a country, does not have a statewide standard in testing, it is important for the Ghana Education Service (GES) to train teachers in assessment (especially, testing practices). Thus, teachers are advised to adhere to the testing practices. It is highly recommended that head teachers take up the challenge of inviting resource persons from recognised academic institutions to organise workshops for teachers on a regular basis to sharpen their skills on effective test construction practices.

The authors of this paper, however, acknowledge that validity and reliability do not entirely rely on the instrument examined. Issues that have to do with the examination conditions such as invigilation, cheating, room ventilation, room lightning, among others also contribute to the variance in test scores. The authors did not adequately probe into some of these issues. We, therefore, recommend that further studies can go further to investigate some of these issues. It is vital, however, to say that the teacher plays a significant role in ensuring proper examination conditions. Again, caution should be taken not to generalise the findings of this study to a wider population.

ORCID

Frank Quansah  <https://orcid.org/0000-0002-4580-0939>

5. REFERENCES

- Allen, M. J., & Yen, W. M. (2002). *Introduction to measurement theory*. Illinois: Waveland Press.
- Amedahe, F. K. (1989). *Testing practices in secondary schools in the Central Region of Ghana*. Unpublished Master's thesis, University of Cape Coast, Cape Coast.
- Amedahe, F. K., & Asamoah-Gyimah, E. (2016). *Introduction to measurement and evaluation*. Cape Coast: University of Cape Coast Press.
- Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative and mixed methods approaches* (4th ed.). London: SAGE Publication Inc.
- Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Ohio: Cengage Learning Press.
- Curriculum Research and Development Division (CRDD) (1999). Investigation into student assessment procedures in public Junior Secondary Schools in Ghana. Accra: Ghana Education Service.
- Ebinye, P. O. (2001). Problems of testing under the continuous assessment programme. *J. Qual. Educ.*, 4(1), 12-19.
- Etsey, Y. K. A. (2003). Pre-service teacher's knowledge of continuous assessment techniques in Ghana. *Journal of Educational Development and Practice*, 1(1), 1-18.
- Etsey, Y. K. A. (2012). *Assessment in education*. Cape Coast: University of Cape Coast Press.
- Hamafyelto, R. S., Hamman-Tukur, A., & Hamafyelto, S. S. (2015). Assessing teacher competence in test construction and content validity of teacher made examination questions in commerce in Borno State, Nigeria. *Journal of Education*, 5(5), 123-128.
- Kazuko, J. W. (2010). *Japanese high school mathematics teachers' competence in real world problem solving*. Keto Academy of New York and Teachers College Columbia University.
- Ministry of Education [MOE] (2012). *Reports for 2012*. Accra: MOE Press.
- Nitko, J. A. (2001). *Educational assessment of students*. New Jersey: Prentice Hall.

- Ololube, N. P. (2008). Evaluation competencies of professional and non-professional teachers in Nigeria. *Studies in Educational Evaluation*, 34(1), 44-51.
- Onyechere, I. (2000). *New face of examination malpractice among Nigerian youths*. The Guardian Newspaper July 16.
- Quansah, F., & Amoako, I. (2018). Attitude of Senior High School (SHS) teachers towards test construction: Developing and validating a standardised instrument. *Research on Humanities and Social Sciences*, 8(1), 25-30.
- Tamakloe, E. K., & Amedahe, F. K. (1996). *Principles and methods of teaching*. Cantoment: Black Mask.