Research on Humanities and Social Sciences
ISSN 2224-5766 (Paper) ISSN 2225-0484 (Online)
Vol.8, No.1, 2018
www.iiste.org

Attitude of Senior High School Teachers Toward Test Construction: Developing and Validating a Standardised Instrument

Frank Quansah Isaac Amoako
Department of Education and Psychology, University of Cape Coast

Abstract
Test construction is an essential part of teachers' responsibility. Teachers are therefore supposed to craft well-functioning items in ensuring effective teaching and learning. This study seeks to develop and validate a standardised instrument in measuring teachers' attitude towards test construction. The study further explores the attitude of teachers towards test construction. The instrument was developed based on literature as well as personal experiences of the researchers. The developed instrument was administered to 432 Senior High School teachers in the Cape Coast Metropolis. Through an exploratory factor analysis, four dimensions were obtained which include: planning, item construction, item review and assembling. A confirmatory factor analysis was then conducted to examine the factor loadings of the items. After critical evaluation, the items on the instrument remained 32 which was on a four point Likert scale. Further analysis revealed an overall negative attitude of SHS teachers towards test construction. It is recommended that Ghana Education Service (GES) together with headteachers of various SHS should ensure effective supervision of teachers in constructing test for students.
Keywords: Item construction, Item review, Testing, Assembling, Test construction

1. Introduction

The competency in test construction is an essential tool needed by every teacher if learning and instructional objectives are to be effectively attained. The importance of tests in the educational system is enormous. Test provides a platform by which any significant educational objectives can be achieved (Hamafyelto, Hamman-Tukur & Hamafyelto, 2015). The effectiveness of learning goals, entrenched in the curricula of a school continues to be the most fundamental sign pole for institutional superiority, educational development and individual goals. Teachers are therefore required to have adequate knowledge in achieving these learning objectives in an accurate and precise manner. Teachers must, thus, have the capability in the science and art of test constructing (D'Agostino, 2007).

A number of studies have explored teacher's classroom test construction skills (Hamafyelto et al., 2015; Kazuko, 2010; Onyechere, 2000). Ololube (2008) also evaluated test construction skills of professional and non-professional teachers in Nigeria and reported that professional teachers tend to construct effective evaluative instruments more than the non-professional teachers. It was also found in Ololube's study that professional teachers have the propensity to employ the various assessment techniques correctly, which is unlikely to happen in the case of non-professional teachers. Onyechere (2000) found that some teachers craft poor tests while others continue to use replica of test items because they seem to have inadequate skills in test construction. Hamafyelto et al. (2015) discovered that Senior High School (SHS) teachers in Borno State, Nigeria, constructed items which focused on lower cognitive operations. In Ebinye's (2001) view, test construction has been found to be a major source of anxiety among many teachers in Nigerian schools, especially, less experienced ones. This anxiety stems greatly from lack of skills in test construction among these teachers. The problem of test construction was made clear in a typical example:

> *A classroom teacher taught her pupils in second grade a lesson on 'magnet' and asked them, on the following day, to write a six letter worded object which picks things. She expected almost the whole class to return the word –'magnet' as their response. To her chagrin, the answer given by more than 50% of the class was 'mother' (Daily Bread, 2011, p.23).*

The teacher must have wondered what actually went wrong. Was it that she did not teach well or that the pupils did not understand what was taught? The problem stems from neither the teaching nor the pupils' learning but from the way the test item was written. The question given by the teacher was not perfectly clear, thus, giving room for more than one possible correct response.

In Ghana, a number of studies have indicated that teachers do not follow testing principles and consequently, have poor testing practices (e.g., Anhwere, 2009; Amedahe, 1989). Amedahe (1989) revealed that SHS teachers in the central region of Ghana have inadequate skills in testing. In a similar study among Junior High School teachers in Ghana, teachers were found to have limited competencies in the management in the assessment practices (Curriculum, Research & Development Division [CRDD] of Ghana Education Service, 1999). A critical examination of literature indicates poor test construction skills of most teachers in all levels of education across diverse subjects globally, and in Ghana to be specific (Anhwere, 2009; Amedahe, 1989; Ebinye, 2001; Hamafyelto

et al., 2015; Kazuko, 2010; Onyechere, 2000). This is really a great problem as students achievement are likely to be reported with errors because poor items are used to measure achievement. Is it that teachers are not well trained in test construction? Is it that teachers are trained well but feel reluctant in using what has been taught them? It is important to state that these previous studies examined teachers test construction skills by asking them what they actually do when crafting test items for students. However, these studies do not provide a comprehensive picture of what teachers know. This is because teachers might have the knowledge but would be reluctant in practising what he/she know. This is seen in the study of Ebinye (2001) who found that crafting test items appeared to be a burden on teachers. Therefore, irrespective of the knowledge the teacher has, it is likely to construct poor questions, or perhaps, repeat already existing questions (Onyechere, 2000).

In Ghana, teachers are trained in assessment of which test construction is an important component. In the Colleges of Education, for instance, students are taken through a full course in educational assessment. The course content allows these students to have a practical knowledge on test construction and assessment, in general. Similarly, universities in Ghana who train teachers (e.g., University of Cape Coast, University of Education, Winneba and Valley View University) also have a course in assessment for potential teachers to be trained in assessment. This course also enlightens students on the construction of test items. Our personal observations have confirmed earlier studies (Anhwere, 2009; Amedahe, 1989; Ebinye, 2001) that even though teachers are trained in school assessment which includes test construction, most of them do not adhere to the rules governing these practices which leads to poorly crafted questions. From our interaction with some teachers in some SHS (in the Cape Coast Metropolis) during an educational out-programme, it appears that teachers attitude towards test construction is nothing to boost of and this, to a greater extent, contributes to the construction of poor items. This study seeks to empirically examine the attitude of SHS teachers through the development and validation of a standardised scale in order to provide a standard measure of attitude towards constructing tests.

2. Development of the Instrument

The instrument was developed based on the behaviours exhibited by teachers in various schools. These behaviours were observed by the authors of the instrument. Literature was, further, reviewed to obtain information on the test construction behaviours of teachers (e.g., Allen & Yen, 2002; Nitko, 2001). Items were then cautiously crafted based on literature and observations made by the researchers. Initially, 41-items were crafted but only 32-items remained after the instrument had gone through several review and factor analysis. The items were on a four point Likert scale of agreement (SD- strongly disagree, D-disagree, A-agree, SA- strongly agree). After the items were crafted and reviewed, a pilot testing was conducted among 100 teachers from some selected SHS in the Sekondi-Takoradi Metropolis. This was done to establish the validity and reliability of the responses which will be elicited by the instrument. Some items were modified after the pilot testing of the instrument. Items like "Learners decide item format to be used" was changed to "I prefer the item format of a classroom test to be decided by the learners". In all 4 items were reworded after the pilot testing. The instrument was then administered to 432 teachers in some selected SHS in the Cape Coast Metropolis.

2.1 Ensuring Validity

The development of the ATC scale was carefully done to ensure the validity of responses solicited. Efforts were made to ensure that the questions crafted represented attitudinal behaviours of teachers (Nitko, 2001). After the items were crafted, they were also reviewed by experts (PhD students) in the Measurement and Evaluation field to validate the instrument. This was done in line with Anim's (2005) assertion that content and construct validity is determined by expert judgement. Results from the factor analysis revealed that the Kaiser-Meyer-Olkin (KMO) test of sampling adequacy and Bartlett;s test were not violated (See Table 1) based on Crocker and Algina's (2008) criteria. An exploratory factor analysis using Principal Component Analysis Method was, then, conducted to determine the factors involved in the scale. The scree plot was used to determine the factors. The exploratory analysis revealed four factors (See Figure 1).

Table 1: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy	.602
Approx. Chi-Square	7119.354
Bartlett's Test of Sphericity: df	628
sig.	.000

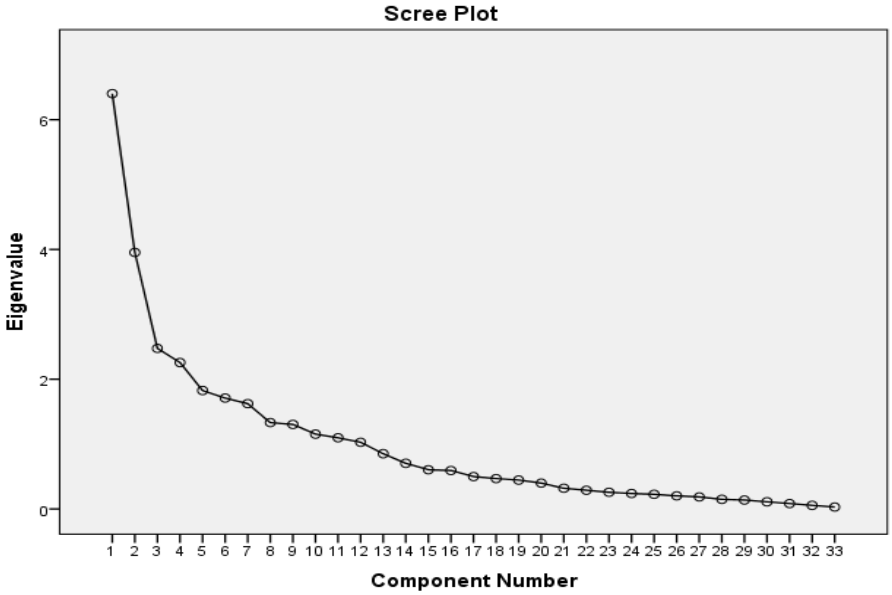

Figure 1: Result on Scree Plot

A confirmatory analysis using Minimum Likelihood Method was further conducted to explore the factor loadings of each item. Items with factor loading of .3 and below were rejected (See Table 2). After the factor analysis, 32-items remained. Item like "I believe good items cannot be crafted without considering the learning objectives" had factor loading of .235 and thus, was rejected. Based on the results from the confirmatory factor analysis, the four factors were labelled as: planning, item construction, item review and item assembling.

Table 2: Factor Rotation

Items	1	2	3	4
To be honest, it is a waste of time trying to outline the purpose of a test when planning for the test.	.633			
I just need my textbook to start writing test items.	.532			
I believe good items cannot be crafted without considering the learning objectives.	.235*			
I mostly do not prefer using test specification table in crafting questions.	.632			
I prefer to finish crafting the test before considering the thinking skills those items measure.	.642			
Since I am the classroom teacher, I do not need to specify the content area I want to test.	.692			
Planning a test is needless if I am the teacher	.772			
I prefer writing items based on what learners are expected to know whether taught or not.	.610			
As a teacher there is nothing wrong with crafting items without considering the learning objectives.	.767			
I prefer the item format of a classroom test to be decided by the learners.	.451			
It is not possible to always craft new questions for learners.	.492			
Crafted items do not necessarily have to match learning objectives.	.463			
I like to write tricky questions to test my students understanding.	.589			
Arranging of the options to multiple-choice items alphabetically is not compulsory	.491			
I always refer to test specification table when constructing items.		.475		
I like to always write items with the same difficulty level.		.739		
There is the need to take items verbatim from textbooks used in teaching.		.403		
I usually construct test items few days for the paper to be written.		.496		
It is optional to review constructed items before it is administered		.620		
Checking for the item difficulty and discrimination after the test has been constructed is not too necessary		.714		
It is essential to present more difficult items before less difficult items in assembling crafted items		.654		
It is optional to number all the items on a test		.443		

Items	1	2	3	4
It is optional to provide clear directions for examinees on the test instrument		.630		
It is right to arrange options of test items horizontally		.621		
It is better to rely on past questions when constructing a test.			.677	
I like to prepare marking scheme after the test have been administered.			.513	
It is necessary to check for the clarity of crafted items			.416	
I prefer preparing marking scheme two or more days after constructing the test			.438	
I always like to arrange questions into sections based on their nature or type			.486	
I select questions from topics I think students have understood.				.414
I think test specification table should be prepared by test experts and not the classroom teachers.				.553
It is essential to identify behaviours to represent a construct when crafting test items.				.654
I do not think it's necessary to craft more items than actually needed.				.654

1-Planning; 2- Item construction; 3- Item Review; 4-Assembling
*Item rejected

2.2 Estimating Reliability

Estimating reliability of items cannot be overlooked because every investigator consider it necessary in gathering objective and accurate information. There is the need, therefore, to estimate the reliability of responses of the construct of interest (Quansah, 2017). The reliability of the instrument was esyimated using the Cronbach's Alpha Reliability Method. The reliability estimate for each sub-scale as well as the whole instrument were investigated (See Table 3). The overall reliability estimate of the instrument was .85. This reliability co-efficient is sufficient enough to ensure reliable responses as indicated by Pallant (2010) that a reliability coefficient (alpha) of .70 or higher is considered appropriate.

Table 3: Reliability Estimate for Sub-scales

Sub-scale	No. of Items	Coefficient
Planning	11	.81
Item Construction	11	.79
Item Review	3	.70
Item Assembling	7	.71

3. The Use of the Instrument and its Administration

The instrument, after its development and validation was named "Attitude towards Test Construction (ATC) Scale". The ATC scale is designed to provide much knowledge to stakeholders in education on the attitude of teachers towards test construction. Specifically, ATC has been developed to assist headmasters/mistresses, Ghana Education Service (GES), school counsellors and test experts in finding out the attitude of teachers towards test construction in a particular school. This will provide relevant clues of the test construction practices of the teachers. Teachers who have been found to be performing poorly can be administered the ATC scale to find out his/her attitude towards test construction. This is because test construction practices has been found to be significantly related to teacher effectiveness (Hamafyelto et al., 2015). The ATC scale can also be used as a research instrument for students and other researchers who have interest in the area of test construction. These researchers can, thus, adopt or adapt the instrument for their study.

The ATC scale can be administered to individuals or groups. For individual administration, the respondent needs to be educated on the need to respond to the instrument. Effort must be made to establish good rapport with respondent(s) so that accurate responses would be given willingly. The individual should be allowed to independently respond to the instrument. On group basis, the investigator should ensure serene environment for the respondents. Regardless of the individual or group of people who will be given the instrument to respond to, their consent must be sought. It is important to ensure that ethical considerations are followed in the administration of the instrument. In all, 25-30 minutes is appropriate for respondent(s) to respond to the instrument.

4. Scoring and Interpretation

The ATC scale has both positive and negative questions of which responses are measured on 4-point scale. In scoring the items on the instrument, negative items are scored on point score from 1-4. That is, strongly agree is valued 1-point, agree for 2-point, disagree is 3-point, and strongly disagree for 4-point. For positive items, strongly agree is 4-point, agree 3-point, disagree 2-point and strongly disagree 1-points. Apart from items 14, 16, 23, and 32, the rest of the items are negative questions. For the overall attitude, the responses from all the items are added and divided by the number of questions. The same computational method is applicable to the sub-scales (i.e,

calculating the composite score for the responses for a particular respondent or group of respondents). In calculating for the attitude of respondents the mean of their responses is computed for and interpreted. In interpreting the attitude of respondents of a particular item (e.g., item 5), the mean score of the responses is compared with 2.5 ([1+2+3+4]/4=2.5). Mean scores less than 2.5 shows a negative attitude whereas mean scores above 2.5 shows a positive attitude to that particular item. For the interpretation of individual scores, the mean of the obtained scores is also compared with 2.5.

5. Exploring Teacher Attitude towards Test Construction

After the instrument has been validated, the attitude of the teachers were examined based on the validated items.

Table 4: Attitude of SHS teachers towards Test Construction

Sub-scales (Attitudes)	No. of Items	Mean	SD
Planning	11	2.43	.74
Item Construction	11	1.90	.89
Items Review	3	2.03	.66
Assembling	7	2.14	.63
Overall Attitude	32	2.13	.72

Results (in Table 4) indicate that SHS teachers have negative attitude towards the planning of classroom test (*M*=2.43, *SD*=.74), item construction (*M*=1.90, *SD*=.89), Items review (*M*=2.03, *SD*=.66), and Assembling (*M*=2.14, *SD*=.63). Generally, teachers in SHS in the Cape Coast Metropolis were found to have an overall negative attitude towards test construction (*M*=2.13, *SD*=.72).

6. Discussion

The need for teachers to construct good test in assessing their students have been underscored in literature (Hamafyelto, 2015). While some teachers are found constructing poor items, others are found to be repeating already existing questions (Onyechere, 2000). Some authors have attributed this to teachers' limited knowledge and skills in the area of test construction (e.g., Anhwere, 2009; Amedahe, 1989; Ebinye, 2001; Hamafyelto et al., 2015; Kazuko, 2010; Onyechere, 2000). Others have attributed poor questions of teachers to the fact that teachers see test constructions a major source of anxiety and burden (e.g., Ebinye's, 2001). This present study revealed another factor which also accounts for the poor construction of test items among teachers. Teachers were found to have a negative attitude towards test construction. This may contribute to the construction of poor questions among these teachers as indicated in previous studies. It is likely that teachers have the knowledge about test construction but their attitude prevent 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 (Nitko, 2001). This explains the reason why some teachers see test construction as a burden. The findings of this present study implies that even when teachers are given adequate training in the area of test construction, it is unlikely that their skill attained might be put to use if these teachers have negative attitude towards crafting the questions. This presupposes that the attitude of teachers towards test construction is likely to act as a moderator in the relationship between knowledge and practice of test construction.

7. Conclusions and Recommendations

Testing in education cannot be under emphasized because teaching and learning can never be complete without it. Teachers would, thus, be ignorant of how well they are doing as well as how well the students are grasping the concepts being taught. Nevertheless, these measure of teacher effectiveness and students performance can never be seen if test are poorly constructed. Even though, teachers are taken through courses in their training, test construct seem to be a nightmare (Nitko, 2001). It is believed that if attitude influences practice (Ebinye, 2001), then there is the need for attitude of teachers to be explored. This is the foundation for the development of ATC (Attitude towards Test Construction) scale. The instrument, therefore, provides a standardized measure of the attitude of teachers towards test construction. It is important for stakeholders to re-orient teachers on the need to follow test construction procedures and to put to use their skills attained from various training they have had. As more training programmes through seminars and workshops are organised for teachers, stakeholders should be aware of the fact that the training alone do 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 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 test for students.

References

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.

Anhwere, Y. M. (2009). *Assessment practices of teacher training college tutors in Ghana*. Unpublished Master's thesis, University of Cape Coast, Cape Coast.

Anim, M. E. (2005). *Social science research: Conception, methodology and analysis*. Kampala: Makerese University Press.

Curriculum Research and Development Division (CRDD). (1999). Investigation into student assessment procedures in public Junior Secondary Schools in Ghana. Accra: Ghana Education Service.

D'Agostino, J. V. (2007). Quantitative research, evaluation, and management section. Ohio State University. 29 Woodruff Avenue, Columbus, OH 42310-1177. USA.

Daily Bread. RBC Ministries (2011*). Our daily bread: For personal and family devotions*. Michigan, USA: Discovery House Publishers.

Ebinye, P. O. (2001). Problems of testing under the continuous assessment programme. *J. Qual. Educ., 4*(1), 12-19.

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.

Nitko, J. A. (2001). *Educational assessment of students*. New Jersey: Prentice Hall.

Onyechere, I. (2000). *New face of examination malpractice among Nigerian youths*. The Guardian Newspaper July 16.

Pallant, J. (2010). *SPSS survival manual* (4th ed.). New York, NY, McGraw Hill.

Quansah, F. (2017). The use of Cronbach Alpha Reliability estimate in research among students in public universities in Ghana. *Africa Journal of Teacher Education*, *6*(1), 56-64.