Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.6, No.34, 2015

Procedural Influence on Internal and External Assessment Scores of Undergraduate Vocational and Technical Education Research Projects in Nigerian Universities

Prof. John A.C.
Department of Technology Education, Modibbo Adama University of Technology, P.M.B. 2076, Yola.

S. S. Manabete
Academic Planning Unit, Adamawa State Polytechnic, Yola

Abstract
This study sought to determine the procedural influence on internal and external assessment scores of undergraduate research projects in vocational and technical education programmes in the university under study. A survey research design was used for the conduct of this study. The population consisted of 130 lecturers and 1847 students in the selected Universities between 1995/1996 and 2005/2006 sessions. The study was analyzed using Statistical Package for Social Science (SPSS) version 13.0. The statistical tool used for the analyses were mean, standard deviation and multiple correlation. Cranach Alpha was used via SPSS to compute the internal consistency reliability of the instrument which was 0.86. the results showed that different criteria were applied not only at different universities but change with time. A number of suggestion were proffered to reduce the negative influence using procedure can have on internal and external assessment scores of vocational and technical education projects and enhance its reliability and validity.

Introduction

One of the most important parameters for measuring the validity of internal and external assessment scores of the undergraduate vocational and technical education research projects is the quality of procedure used. It therefore, means that the procedural influence can make or mar assessment process, the final scores awarded to individual students and consequently its reliability and validity.

Assessment procedure according to Ojerinde and Folayajo (1984) tends to ignore the essential process and substantive structure of research project writing. In addition, the scoring rules give little or too much attention to certain skills, concepts and knowledge. This provides an unfair advantage or disadvantage to students who have certain strengths and weaknesses or related to the targeted learning.

The process of assessment does not only differ from one university to another, but changes with time and from one assessor to another. Okpala, Onocha and Oyedeji (1993:20) described the complexity of the process this way: even when two or markers were asked to assess an undergraduate research project work, the perceptions of the assessors as regards the procedure differ resulting in differential scoring. Consequently, this calls to question the reliability and validity such assessment and the interpretation that teachers make concerning such course.

Assessment forms an integral part of the teaching process to the extent that, without it teaching and learning success or failure will be difficult to measure. It is as a result of its lifelong consequences on the students who are assessed that the process is critically reviewed from time to time and carefully examined. Assessment, according to Osokoya (1987), does not operate in a vacuum rather it is used to introduce innovations into instructional methods and improve the learning process. Cronbach (1988) raised four important questions on assessment techniques. What is the purpose of assessment? What do we assess? When should we assess? And what procedures will provide the most reliable and valid assessment?

In actual practice of project writing, a lot of assessments which may be implicit precede the internal and external ones. These are the assessment carried out by individual supervisors about candidates without necessarily assigning marks and the one that is based on the project proposal. It is assumed that assessments preceding the external all contribute to the improvement of the quality of the project at every stage. As the researcher overcomes one level of difficulty, it motivates the learner to move to the next stage (Bray, 1992 & Black, 1993). But, in most cases the reverse is the case because the average scores in external are lower than internal assessment.

In any assessment system the procedure is the pivot on which reliability and validity revolves. If one of the procedures is violated or standard not strictly adhered to, inconsistencies of vocational and technical education research projects become prominent.

Statement of the Problem

One of the requirements for the award of degrees in Nigerian Universities is the completion of research project

work. For this reason, the project work is normally subjected to both internal and external assessment before the student is awarded the final score at undergraduate level. However, Soyinbo and Fasunloye (1984) pointed out the differing opinions on independent assessment by many educators and criticized the measures used to evaluate research projects as not being reliable and therefore not valid. Consequently, the problem becomes glaring when there is a gap between what the internal assessors believe the students should know and the external examiners' perceptions in assessment of vocational and technical education projects (Brown, Bull & Pendlebury, 1997).

Whenever human beings are used in the assessment procedure, there is always a worry about whether the results are reliable. People, according to Bull and Kimball (2000), are notorious for their inconsistencies. Assessors are easily distractible, get tired of repetitive tasks, daydream and misinterpret.

One of the approaches to assessment is assessing the learner. This approach is often used by school personnel especially lecturers whenever the student is referred for evaluation because the learner is experiencing either academic or behavioural problems. This type of assessment focuses more on what is wrong with the students. It is an effort made to identify what Arter and Jenkins (1989:542) referred to as "within-student deficits" or disabilities or strengths and weaknesses in the development of specific skills. Students who experience such academic problems are assumed to suffer from internal conditions called deficits, disorders, disabilities or dysfunction. It is assumed that assessment can be used to reliably and validly identify student academic deficiency but this approach seldom leads to effective instruction (Newland, 1999).

One of the procedures that has given persistent rise to inconsistent scores is the absence of detailed marks schemes in some of the universities studied. The assessors have many students to cope with and cannot read through all the projects within the short period given to complete the assignment thereby making reliable and valid grading difficult. Gashua (2001) attributed the problem involved in external and internal assessment to improper coordination of criteria and procedure used which resulted in low or no relationship between the two modes. He maintained that something must be wrong and Nigerians will continue to question the disparity.

Procedure that will produce reliable and valid assessment include a clear criteria of the guideline on assessment procedure that is made to stabilise for sometime before any change is introduced, emphasise on quality of written and orals in assessment of research projects. IBO (2004) criticized the process of absence of analytical mark scheme and suggested specific instruction to assessors regarding how to break down the total mark available for the various segments of vocational and technical education research project.

Purpose of the Study

This study sought to determine the procedural influence on internal and external assessment scores of undergraduate research projects in vocational and technical Education programmes in the university under study. The study sought to accomplish the following objectives:

(i) Determine the procedures employed in assessing undergraduate vocational and technical education research projects in the three Universities that may result in inconsistencies;

(ii) Ascertain the strategies that can be used to reduce inconsistencies in undergraduate vocational and technical education research projects.

Research Questions

The study sought answers to the following research questions:

(i) What are the procedures employed in assessing vocational technical education research projects in the three universities?

(ii) What are the strategies that can be adopted to reduce inconsistencies in vocational and technical educational research projects?

Hypotheses

The following null hypotheses were tested at 0.05 level of significance.

Ho_1: There is no significant difference between the mean responses of internal and external assessors on the procedures employed by the universities to assess vocational and technical education research projects.

Ho_2: There is no significant difference between the mean responses of internal and external assessors on strategies that can be used to reduce inconsistencies.

Significance of the Study

This study is significant because it will help to identify good procedures that will enhance reliable and valid assessment and eliminate poor ones.

Assessment Design Model

According to Mislery, Wilson, Ercikan and Chudowsky (2001), the assessment design model is based on three premises: student model, task model and evidence model. This model is similar to the one proposed by Okpala,

Onocha and Oyedeji (1993) as shown in Figure 1.

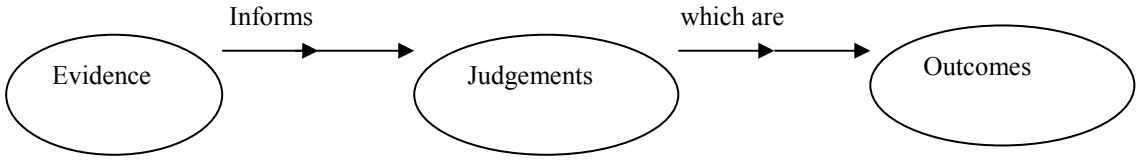

Figure. 1: Assessment Design Model
Excerpts from Okpala, Onocha and Oyedeji (1993:52)

Assessment process was seen as a three-stage process where there is evidence which informs the assessor to make judgment and it is translated into outcomes called grades. The student model is based on what complex knowledge, skills or other attributes should be assessed. The values of student-model variables represent selected aspects of the infinite configuration of skill and knowledge students have, based on a theory or a set of beliefs in the domain. The variables are the vehicles through which a student's progress, decision and plan of instruction is determined. A claim about what a student knows, can do, or has accomplished is expressed in terms of values of student-model variables. The student model provides a language for expressing claim about students.

Task model explains what task or situations should elicit certain behaviours. The task model provides a framework for constructing and describing the situations in which examinees act. The term task was used in the sense proposed by Haertel and Wiley (1993:3) to refer to a "goal-directed human activity to be pursued in a specified manner or context". A task can, therefore, be an open-ended problem in a computerized simulation, and a long-term
project. A task specifies the environment in which the student will say, do, or produce something. Characteristics of stimulus material, instruments, help and tools are examples of various tasks in which the students will act. The task model also specifies the work product, or the form in which what the student says, does or produces will be captured. But again, it is experience that determines the kind of situation that evokes behaviour that provide clues about targeted knowledge and skill and the forms in which those clues can be expressed and captured (Mislery, Steinberg & Almond, n.d). For a particular task, the values of its task model variables constitute data for the evidentiary argument, characterizing the situation in which the student said, did and made something.

It is useful to distinguish between task models from evidence models so as to attend to the resulting performance and know how to evaluate what the examiners see. Distinct and possibly quite different evaluation rules could be applied to the same work product from a given task. Evidence models reveal what behaviours or performances the students should reveal the constructs and what is the connection? An evidence model according to Messick (1994) laid out the part of the evidentiary argument that concerned reasoning from the
observation in a given task situation to revising beliefs about student model variables. There are two parts to evidence model (scoring and measurement component). The scoring component contains "evidence rules" for extracting the salient features of whatever the student says, does or creates in the task situation i.e. the "work product" can be a response to a problem (observable variables) which the assessment designer has determined are key aspects of the performance. For example, a short impromptu demonstration which contains information about a student's subject matter knowledge in Technical Drawing could be the basis of one or more observable variables. As a facet of fairness, however, the student should be informed of which aspects of performance are being evaluated and by what criteria. The measurement component of the outcome tells how the observable variables depend, in probabilistic terms, on student model variables another essential link in the assessment argument. From the forgoing it is clear that the assessment model has clearly supported the argument that a student must demonstrate the knowledge and task attributes both in internal and external examination. The assessor makes judgment by the evidence provided by these attributes which are translated into grades in vocational and technical education research projects.

Research Design

A survey research design is also appropriate because, according to Borg and Gall (1979), it seeks to establish prevailing conditions at a point in time. As a survey research, this study investigates the procedural influence on assessment of undergraduate vocational and technical education research projects.

Population of the Study

The population for this study consisted of 132 lecturers and 1847 students in the Department of Vocational and Technical Education in Federal University of Technology, Yola; Abubakar Tafawa Belewa University, Bauchi and Kaduna Polytechnic, Kaduna (affiliated to Federal University of Technology, Minna) between 1995/96 and 2005/2006 sessions. These consist of students who were admitted into these departments to study Agricultural Technology Education, Construction Technology Education, Mechanical Technology Education and Electrical

Technology Education. Since the entire population constitutes the sample, this research is not constrained by any particular type of sampling technique.

Instrument for Data Collection

A survey instrument was developed using a five-point Likert scale. The instrument contains 14 assessment procedures that can result in differential scoring. 7 items on responses of internal and external assessors on ways to improve undergraduate vocational and technical education research projects. The respondents were asked to indicate their level of agreement or disagreement to the factors on a 5- point Likert rating scale. The rating scale ranged from strongly agreed (5) to strongly disagreed (1). The response category and their assigned numerical values are as follows:

Strongly Agreed (SA) 5 points,
Agreed (A) 4 points,
Undecided (U) 3 points,
Disagreed (DA) 2 points and
Strongly Disagreed (SD) 1 point

Reliability of the Instrument

The reliability of the survey instrument was established through a pre - testing . The instrument was administered to 8 lecturers in the three universities under study. After a few weeks, the same instrument was re-administered on the same group of lecturers. The Spearman Rank order correlation was used to obtain the coefficients of the instrument. The correlation coefficient obtained for the whole instrument was 0.84. The result of the pre- test was used to compare internal consistency reliability of the various sub-sections of the instrument. The Cronbach Alpha was used to determine the extent of homogeneity of the items in measuring attributes addressed in each of the sub-sections of the instrument which was adjudged reliable and used for the study.

Using Cronbach Alpha, the internal consistency reliability coefficient of the instrument was computed to be 0.86. It was therefore, adjudged reliable and used for the study.

Method of Data Collection

The validated questionnaire was personally administered to and collected from the respondents by the researcher. The researcher therefore, visited the three universities several times and had the opportunity to see the respondents. This was done to minimize the difficulties that might be encountered by the respondents in completing the questionnaire and ensure a high rate of return of the questionnaire. Out of the 132 questionnaires that were administered to the respondents, 130 were collected back from them after completion.

Method of Data Analysis

The data for this study were analyzed using the Statistical Package for Social Science (SPSS) version 13.0. The statistical tools used for the analyses were mean, standard deviation and t-test. The mean and standard deviations were used to take decisions on each item for answering research questions while t-test was used to find compare differences in the respondents' opinions.

The decision for answering the research questions was based on real limit of numbers. Any item with the mean value of 2.50-3.49 falls in the category of undecided while items with the mean of 1.0 – l.49 and 1.50-2.49 falls in the category of disagreed and strongly disagreed. Also, items with the mean value of 3.50-4.49 and 4.50-5.00 falls in the category of agreed and strongly agreed respectively. All computed items with the mean values of 3.50 and above are presented as agreed while the ones below are regarded as disagreed in the remarks column.

The hypotheses were tested by comparing the critical values with the calculated values for each hypothesis at 0.05 level of significance. A null hypothesis was accepted when the critical value is greater than the calculated value of the statistic; otherwise the null hypothesis was rejected and its alternative accepted.

Table 1: Procedures Employed by Assessors of Undergraduate VTE Research Projects

N=130

S/No.	Items	Mean $\bar{x}$	Standard deviation $s\delta$	Remarks
(1)	Assessment criteria of VTE research projects do not only differ from one university to another but change with time.	3.98	0.80	Agreed
(2)	Assessment procedure emphasises the quality of written work rather than both oral and written assessment of VTE research project.	3.92	1.11	Agreed
(3)	Assessors always brief the assessees on the salient points to expect in the scoring process.	3.16	0.79	Disagreed
(4)	To obtain evidence of validity, assessors first form learning objectives which will translate into appropriate scores.	3.06	0.99	Disagreed
(5)	Students are often asked to explain their thinking and reasoning as part of assessment task.	3.74	1.09	Agreed
(6)	Assessment procedures are standardized and all students are assessed and re-assessed to complete the exact tasks under the same condition and ensuring the same score when repeated.	3.96	1.02	Agreed
(7)	For a valid assessment, internal assessors who taught and worked with students for at least one session are better positioned to allo marks.	3.69	1.12	Agreed
(8)	When the marking scheme is not detailed enough the same assesso scores the VTE research projects differently at different times.	3.89	0.95	Agreed
(9)	Reading through the entire work before scoring all the sections in a VTE research project usually result in inconsistent scores.	4.38	0.95	Agreed
(10)	Reading through each section of every chapter of VTE projects and scoring is one of the methods of achieving reliability and vali scores.	4.61	0.85	Agreed
(11)	When the number of project works to be examined is too many, it i better to moderate marks rather than going through the project one by one.	3.87	1.06	Agreed
(12)	Reliable grading is very important because carefully written project can be ruined by improper grading procedures and standards.	3.76	1.30	Agreed
(13)	The scores of the internal and external assessors are inconsistent because of the short time given for assessment.	3.90	0.93	Agreed
(14)	Assessment scores of VTE research projects are difficult, time consuming and sometimes boring.	3.97	1.19	Agreed

Table 1 shows that all the items except 3 and 4 are procedures employed by assessors of undergraduate vocational and technical education research projects in the universities under study. The mean ratings of the 12 items that indicate agreement ranged from 3.69 to 4.90; while items 3 and 4 varied from 3.06 to 3.16. It, therefore, means that all the procedures stated in table 9 are employed except items 3 and 4.

Table 2
Ways of Improving Reliability of Undergraduate Vocational and Technical Education Research Projects
N=130

S/No.	Items	Mean $\bar{x}$	Standard deviation $s\delta$	Remarks
(1)	Establish clear and common manageable assessment criteria.	4.93	0.96	Agreed
(2)	Use internal moderation where educators meet during and after the assessment to compare scores and interpretation of the criteria set.	4.58	0.23	Agreed
(3)	Use numerical quantification along with description to increase the level of consistency.	3.85	1.02	Agreed
(4)	Use varieties of assessment methods and adopt the most suitable.	4.04	0.56	Agreed
(5)	Evaluate the assessment criteria for efficacy and relevance.	3.87	0.87	Agreed
(6)	Check the marks scheme against actual responses	4.58	0.15	Agreed
(7)	Allow enough time for reading, scoring and grading.	4.62	0.66	Agreed

Table 2 shows that all the items are strategies for reducing inconsistencies and enhancing reliability of undergraduate vocational and technical education research projects in the universities under study. The mean responses range from 3.85 to 4.93 and the standard deviation values range from 0.23 to 1.02 which are indications of unanimity of response. It therefore means that establishing a clear criterion, use of internal moderators where educators meet during and after assessment to compare scores, use of varieties of assessment methods, and evaluation of assessment criteria for efficacy, relevance, check the marks scheme against actual responses and allow for enough time for reading, scoring and grading are strategies for enhancing reliability and validity.

Table 3: t-Test for the Mean Responses of Internal and External Assessors on the Procedures Employed to Score Undergraduate VTE Research Projects

S/No.	Items	Mean Rating Int. Ass	Ext. Ass	Calculated-t	Remarks
(1)	Assessment criteria of VTE research projects do not only differ from one university to another but change with time.	3.97	4.17	-1.45	NS
(2)	Assessment procedure emphasises the quality of written work rather than both oral and written assessment of VTE research project.	3.72	4.00	-1.04	NS
(3)	Assessors always brief the assessees on the salient points to expect in the scoring process.	3.36	3.50	-0.47	NS
(4)	To obtain evidence of validity, assessors form learning objectives which will translate into appropriate scores.	3.16	3.50	-0.75	NS
(5)	Students are often asked to explain their thinking and reasoning as part of assessment task.	3.64	4.00	-1.27	NS
(6)	Assessment procedures are standardized and all students ar assessed and re-assessed to complete the exact tasks under th same condition and ensuring the same score when repeated.	3.94	4.08	-0.57	NS
(7)	For a valid assessment, internal assessors who taught an worked with students for at least one session are bette positioned to allot marks.	3.67	3.83	-0.58	NS
(8)	When the marking scheme is not detailed enough the sam assessor scores the VTE research projects differently at differen times.	3.83	4.08	-1.50	NS
(9)	Reading through the entire work before scoring all the section in a VTE research project usually result to inconsistent scores.	3.89	4.00	-0.49	NS
(10)	Reading through each section of every chapter of VTE project and scoring is one of the methods of achieving reliability an valid scores.	4.01	3.92	0.43	NS
(11)	When the number of project works to be examined is too many it is better to moderate marks rather than going through th project one by one.	3.85	3.92	-0.23	NS
(12)	Reliable grading is very important because carefully writte projects can be ruined by improper grading procedures an standards.	3.46	3.91	-1.32	NS
(13)	The scores of the internal and external assessors are inconsisten because of the short time given for assessment.	3.99	4.08	-0.55	NS
(14)	Assessment scores of VTE research projects are difficult, tim consuming and sometimes boring.	3.81	4.00	-1.44	NS

NS= Not significant at 0.05 level
Critical Value of t= 1.65
N_1= 120 N_2= 10

Table 3 indicates no significant difference between the mean responses of internal and external assessors on the procedures employed to assess students in vocational and technical education research projects at undergraduate level. The calculated t is less than the critical value of t and therefore the null hypothesis is accepted.

Table 4: t-Test for the Mean Responses of Internal and External Assessors on ways of Improving Reliability of Undergraduate Vocational and Technical Education Research Projects

S/No.	Item	Mean Rating Int. Ass	Ext. Ass	Calculated-t	Remarks
(1)	Establish clear and common manageable assessment criteria.	4.83	4.50	1.56	NS
(2)	Use internal moderation where educators meet during and after the assessment to compare scores and interpretation of the criteria set.	4.28	3.98	1.05	NS
(3)	Use numerical quantification along with description to increase the level of consistency.	3.95	4.36	0.73	NS
(4)	Use varieties of assessment methods and adopt the most suitable.	4.01	3.86	1.25	NS
(5)	Evaluate the assessment criteria for efficacy and relevance.	3.82	3.70	0.59	NS
(6)	Check the marks scheme against actual responses.	3.95	3.86	1.02	NS
(7)	Allow enough time for reading, scoring and grading.	4.35	4.05	0.95	NS

NS= Not significant at 0.05 level
Critical Value of t= 1.65
N_1= 120 N_2= 10

Table 4 shows no significant difference exist between the mean responses of internal and external assessors on strategies for reducing inconsistencies or unreliability of assessment scores in vocational and technical education research projects at undergraduate level. The calculated t is less than the critical level of t and therefore the null hypothesis is upheld.

Discussion of the Findings

The procedure employed by assessors of vocational and technical education research projects is very vital in determining reliability and validity. The findings indicate that twelve out of 14 items were rated agreed. Their mean values ranged from 3.64 to 4.01. The assessors agreed that the following procedures can influence assessment scores of vocational and technical education research projects which corroborated the views of Cronbach (1998) and IBO *w* (2004). These are: different criteria are applied not only in different universities but change with time, emphasis on quality of written work rather than both oral and written research projects; the marking scheme used not detailed enough to enable the same assessor score vocational and technical education research projects consistently. Agbaegbu and Habor -Peters in Ebuoh (2004) agreed with the assessors that reading through the entire work before scoring all the sections usually results in inconsistency and reading through each section of every chapter before scores are awarded is one of the methods of achieving reliable and valid results. This may be due to the fact that when the project is read completely before awarding scores the assessors tend not to remember vividly the points to warrant valid assessment. Conversely, when a score is awarded for each section or sub - section, precise scores is allotted resulting in a more reliable and valid result. Apart from the procedure of reading, some assessors see research projects as time consuming and sometimes boring as shown in Table 1.

The respondents agreed on all the items on ways to improve the procedural influence of undergraduate vocational and technical education research projects in the universities under study. It therefore means that establishing a clear criterion, use of internal moderators where educators meet regularly to compare scores; use of varieties of assessment methods is ways in which inconsistencies can be minimised. This is in line with the views of Rowe, Tunner and Lean (2006) who suggested not only using numerical qualification along with description to increase the level of consistency but check marks scheme against actual responses and allow enough time for reading, scoring and grading.

The finding shows that there is no significant difference between the internal and external assessors opinions on the procedures employed for assessment scores in vocational and technical education research projects. It therefore, follows that the two categories of staff were unanimous in their decision. This finding is in line with Cronbach (1988) and IBO (2004) that different criteria are applied not only in the universities but change with time. In addition, scores not properly allotted to every section to take care of the details in the marking scheme so that inconsistencies can be reduced.

Hypothesis 1 compared the mean responses of the assessors on ways to reduce inconsistencies and increase the level of reliability of undergraduate vocational and technical educational research projects. The

analysis showed no significant difference in the mean opinion of the assessors. This finding is in line with the perceptions of Rowe, Tunner and Lane (2006) that suggested checking of marks scheme against actual response and allow enough time for reading, scoring and grading.

Major Findings of the Study

1. The assessors agreed that procedures change with time, emphasis on written work and standard procedures are used. Other procedures which influence assessment scores of vocational and technical education research projects are reading through the entire work before scoring, reading through each section and allotting scores, moderating marks rather than going through the projects one by one and short time given to assessors to submit scores. However, briefing the assessees on the salient points to expect in scoring and assessors first forming learning objectives were rejected.
2. The assessors accepted all the strategies identified for reducing inconsistencies and enhancing reliability.

Conclusion

(1) There is no significant difference in the mean responses of the internal and external assessors on all the items presented in the instrument for the study.
(2) The assessors agreed that most of the factors can influence assessment scores of
(3) vocational and technical education research projects in the Universities in Northern Nigeria.

REFERENCES

Arter, J.A. & Jenkins, J.R. (1989). Differential diagnosis prescriptive teaching a critical appraisal. *Review of Educational Research*. 49, 517-556

Black, P. (1993). *Testing: friend or foe? Theory and practice of assessment and testing.* In International Baccalaureate Organisation [IBO] (ed.) Diploma programme assessment, principles and practice. Cardiff: International Baccalaureate Organisation [IBO].

Borg, W., & Ball M.O. (1979) *Educational research: An Introduction*. New York: Longman. Inc.

Bray, E. (1992). Fitness of purpose. In R Lloyd-Jones, G. Johnson and R. Currie (Eds). *Guides to assessment in education: Assessment from principles to action* London: Macmillan.

Brown, G. Bull, J., and Pendlebury, M. (1997). *Assessing students learning in higher education*, London: Routledge.

Bull, K.S and Kimball, S.L. (2000). *Basic measurement theory, objectives and needs Assessment.* Oklahoma: Oklahoma State University.

Cronbach, L.J. (1988). Five perspectives on validity argument. Hillsdale, N.J: *File://D:\Education Policy Analysis Archieves Vol.4 No.17 Taylor & No.htm.*

Gashua, A. (2001). Issues between west african examination council and national examination council of nigeria: what type of assessment practices? *New Nigerian,* Monday, January 29th, p.4

Haertel, E. & Wiley, D.E. (1993). *Representations of ability structures: Implications for testing.* In N. Frederikson, R. J., Mislevy and I.I. Bajah (Eds.), Test theory for a new generation of tests.

International Baccalaureate Organisation (2004). *Diploma programme assessment, principles and practice.* Cardiff: International Baccalaureate Organisation [IBO]

Mislery, R.J., Wilson, M.R; Ercikan, K; & Chudowsky, N. (2001). Psychometric Principles in student assessment. Doodrecht, the Netherlands: *Kluwer Academic Press.*

Mislery, R.J., Steinberg, L.S., & Almond, R.G. (in press). *On the roles of task model variables in assessment design*. To appear in S. Irvine & P. Kyllonen (Eds.), Generating items for cognitive tests: Theory and Practice. Hillsdale, NJ: Eribaum.

Newland, T.E. (1999) Psychological assessment of exceptional children and youth. In W. Cruickshank (Eds). *Psychology of Exceptional Children and Youth*. Engelwood Cliffs, NJ: Prencise-Hall Inc.

Ojerinde, D., & Folayajo, W. (1984). *Assessment practices: A new approach.* Ibadan: University Press Ltd.

Okpala, P.N., Onocha, C.O., & Oyedeji, O.A. (1993). *Measurement and evaluation in education*, Ibadan: Strirling-Horden Publishers Ltd.

Osokoya, I.O. (1987). *Strategies, issues and problems of assessment system in Nigeria.* Ibadan: Bisinaike Education Publisher.

Rowe, K. J., Tuner, R., & Lane, K. (2006). *A method of estimating the reliability of assessments that involve combinations of school based tasks and external examinations.* Victoria: Centre for Applied Education Research, The University of Melbourne.

Soyinbo K., & Fasunloye, G.S. (1984). Influence of assessment on science student's performance in selected Lagos State Secondary Schools. *Journal of Science Teachers Association of Nigeria* 22,(2), 65-77