

TEST- BLUE PRINT AND ITS EFFECTS ON THE VALIDITY OF TEACHER MADE TESTS OF THREE STATE COLLEGES OF EDUCATION IN EASTERN STATES

Rev (Dr.) A.C Egereonu

Abstract

This paper entitle "Tests Blue Print" and its effects on Teacher made Tests, compares the course outline with Essay and objective tests to know how content valid are both tests, using Taxonomy of Educational Objectives of Benjamin Bloom on cognitive for three college of Education in Eastern States. The Taxonomy coverage for both tests were examined and their relative weights. Two research questions were generated. Sample consisted of six specialists in tests and measurement out of twelve teachers as population. Three provided the course outline and questions and the other three rated the instrument. A rater - reliability coefficient using sum of squares method, and Rank order correlation gave respectively 0.92 and 0.71. Generally from all indications alt schools displayed interest in setting mainly lower aspects of taxonomy levels in Basic knowledge and comprehension than other higher levels of Application, Analysis, synthesis and Evaluation. School C had a better coverage than school A and B. recommendation were made for workshops and entire change using a new system where both Essay and objective are obtained as a new instrument to meet the criteria of content validity and good taxonomy.

Introduction/Background

The Test-Blue Print otherwise called the table of specifications is a by-product from the Taxonomy of Educational objectives of Benjamin Bloom.

Gronland (1976), defines table of specifications as a two-way chart that relates to the instructional objectives to the content and specifics given to each task of learning outcome.

Benjamin Bloom (1956), develops a new taxonomy involving cognitive, affective domains. This concept not only revolutionized the process of lesson planning for classroom teaching, testing and measurement, but also defines goals for achievement by teachers, counselors, and educational administrator..

Such goals as "I want to assist students to realize their full potentials or to become self actualized etc, are narrow and unachievable, according to Blooms, rather Blooms viewed the concept from not only trivial, but also cosmic, that is/he viewed achievements from micro and macro variables. His scheme classified educational objectives to specific classroom procedures. This approach then has healthy effects in making the teacher to specify his goals and means of getting there. It also fits process and materials to instructional strategies. Finally, it specified a sequence of six stages/levels of objectives, that are matched to a sequence of assessment strategies, namely; Basic Recall (knowledge), Comprehensive Application, Analysis, Synthesis and Evaluation.

Test-Blue-Print or the TBP/(TOS) is central to testing preparation for content validity to exist since its objectives is to cover content of instructions. Evidence from professionals and non-professionals; attest that difficulties are encountered in using this concept to validate tests. For instance, Ihekwa (2001), writing on validity of senior secondary certificate Examination Questions in physics between 1994 - 1998 in May - June Exams discovered that SSCE Physics questions lacked content validity due to non-usage of the Test-Blue print. Similarly, Offor (2001), also studied a (Chemistry Test validity between 1994- 1998 using objective questions, inferred from the lest data set by West African Examination Council that the SSCE are neither representative of the curriculum topics, nor the cognitive levels.

Chukwu (1998), associating the Mock Senior Secondary Certificate Examination and WAEC Certificate Exams in Federal Capital Territory Abuja analyzed the curriculum to ensure coverage and difficulty in testing mathematics topics, discovered that coverage was low based mainly on first three levels (basic) neglecting other levels, yet lacked cognitive weight coverage.

Mehrens & Lemann (1975), explained that the purpose of Table of Specification is to define as

clearly as possible the scope and emphasis of test and to relate the objectives to the content. If test experts are impatient, to use it and who use standardized instruments find it difficult, this paper or writer wonders to what extent teachers using teacher made tests, use this taxonomy of educational objectives. In fairness to current test evaluators in the classroom, we are not only dealing on population explosion in enrolment in schools due to knowledge explosion, we are also dealing with poor infrastructure, poor motivation and remunerations. Such extraneous variables could be hampering worthwhile efforts of dedicated and committed teachers or lecturers, vis-a-vis teaching and assessment strategies.

According to Grunlurd (1987), using taxonomy is exerting and consuming and most teachers avoid it. It needs skill, encouragement, motivation, training, and commitment.

A poorly designed testing mechanism can impede a most dedicated efforts, while a well design system can be the bad rock for educational development.

Problems of the Study

In cognizance of the above facts and taking an academic ride backwards into test concepts on essay and objectives as instruments of measurement. It is a well-known fact that essay has few questions making it subjective while objective questions has more but lacks dept in reasoning and expositions. Other definable problems arise from unskilled teachers in the field and uncommitted skilled teachers who produce poor taxonomy including poor weighting of the content and sampling of good items.

As we said earlier, using Ihekwoaba, (2001); Offer (2001), Chukwu (1998), statement as based to buttress our arguments; using also (WAEC) opinions, they all agreed that Blooms concepts are not properly used.

GRUNLUNG (1987), expressed the same opinion of the difficulty of the test blue print. This paper intends to find out to what extent this problem exist. Since 'habits die hard', a revolution to change this evil habit or an entire new system as a remedy is needed.

Research Questions

To guide this paper, and actualize our objectives, two research questions were generated

To what extent are the essay questions weighted using the taxonomy of educational objectives?

To what extent are the objectives questions weighted using the taxonomy of educational objectives?

Review of Related Literature

Validity refers to the extent to which an empirical indicant measures what it purport to measure. We are not measuring or validating the indicant parse, rather, the purpose, and objective for which (he indicant is being used, that is subjected to validation procedures. For instance, Crombach (1971), agrees that "one validates not a test, but an interpretation of the data from a specified procedure.

There are types of validity namely: face, criterion (pre-dative and concurrent), construct, and *content validity*.

Anocha and Okpala (1995), defines content validity as the systematic examination of the degree to which a research instrument covers a representation sample of the universe of content, that may be cognitive, effective or psychomotor.

In other words, it must adequately sample all topics or concepts in the universe (course outline). Chase (1978), in his efforts to establish the obvious meaning of content validity took the render to a geography and historical-analysis, where the map of California, San Fransisco in U.S.A. is clearly represented in bold type, while Los Angeles was a mere dot, identified by a very small type. Also Health valley was pictured as encompassing a large segment of the Southern part of the State. Yose Mile Valley, was not even listed. Sacramento mountains appear huge and conspicuously labeled.

According to Chase, utterances of amazement will follow by a glance by any school child in the United States of America. This implies laying emphases on less relevant areas, and neglecting some areas. Even in Nigeria such amazement and exclamation should follow in an Essay Test (Examination) where few questions are asked out of many. Some areas that are very relevant neglected and weighing lopsided.

Both examination whether in U.S.A. in Nigeria are very subjective, yet we do nothing about it.

Competence of Teachers

Another relevant factor is the setting of Essay Questions or even Objective type by the teacher. His expertise, experience, skill, commitment, skill and training are factors crucial to the objectivity of setting questions for credible scores.

Say (1974), defines evaluation as a "process through which a value judgment or decision is made from a variety of observations and from the background and training of the evaluator (p 3). The evaluator becomes a sine-qua-non in making the test construction objective, reliable and valid.

Ugodulunwa (1977), in his paper argued that educational assessments at all levels in Nigeria needs to provide a valid and reliable results on which important decision should be based. He highlighted the threats to quality educational assessment, when he stressed that:

Teachers need to possess necessary competence in instrument development, in addition to knowledge of the subject matter and competence in art of teaching. The teacher requires adequate provision of necessary facilities support staff, equipment and materials Teachers should also be properly motivated and remunerated, if they are to perform the roles expected of them (p. 11).

The above reference speaks for itself, to meet the evaluation roles of the teachers for credible results and good testing strategies.

Essay Objective

Encyclopedia Britannica (1997), exposed that Montague, writing on Essay published in 1958, explained that Francis Baco (1997), Charles Camb; Thoreau and Emerson's Essays within same period, were all effective in nature by "feeling their way" towards (the expression of personal thoughts and experiences, which were lofty and austere; fully of utterances and intellectual energy. From the above discourse, one can glean the subjectivity of the essay through deception by words, expression, language, beauty and strength that can make the unwary examiner trapped by scoring students arbitrarily. Added to this are the few questions reminiscent of the essay. If students are given freedom, exposure, experience to express their feelings and thought, it may improve the essay, not requiring students to the hours of few questions under duress, maybe, without application of test blue print

Mastery tests, using power and speed tests instrument could help, yet they are rarely used. These contain large number of questions that cover the cause domain (outline). Do they use taxonomy of educational objectives as base to set question is another question.

It is well recognized by all researchers that objective tests are content valid unlike Essay tests with few questions, since objective can generate more questions to cover the outline of universe, but Essay lacks coverage of items.

A combination of both are relevant, yet the old system of isolation of each has not solved the problem, due to non-usage of taxonomy of educational objectives.

Taxonomy of Educational Objectives

The Test-Blue-Print (Table of specifications) is a veritable tool, relevant from primary up to Higher Institution (Tertiary) in a varied institution. For instance, if a structural engineer intends to build a house, he draws a plan to the specifics in fine details or micro level that reflects his aims and objectives. These details are in correct dimensions in ratio and measurement to void lopsidedness. An automobile engineer in building a car produces a prototype or sample from a diagram (a blue print). Similarity, tailors produce patterns (graphical), samples before the main sowing is done. But in educational milieu, they find it difficult to apply this blue print. Should there be a modification due to such difficulties encountered in constructing?

The Lagos Department of Research Division (2002), produced a research report on Taxonomy of West African Examination Council in cognitive domain. The objective was to find out cognitive levels of objective and Essay coverage and spread of five subjects between 2000 and 2003 (4 years) using a survey design. They asked the opinions of testing experts on the taxonomy of WASSCE question papers on five subjects using a panel of four experts for each subject. Subjects/(sample) includes WAEC Examiners, Secondary school teacher, lecturers of higher

institutions, and WASCO researcher officers. Sampling was stratified. Findings from all indications firstly proved that many WASSCE questions reviewed, assessed lower cognitive levels, than higher levels.

In some areas WAEC was fair, but to a large extent failed. For example, all subjects assessed some aspects of higher level, and more of lower levels, of cognitive objectives. The researcher further blamed teachers for inadequate knowledge of item generation, due to poor selection of item, during the compilation stage of test papers. The researcher inferred that this is the by-product of what obtains at instructional levels of teaching and learning process, (Ausebel, 1967).

One does not give out what it does not have. So teachers should promote higher order cognitive abilities such as critical thinking, problem solving, and novel applications to their students or else our future evaluators cannot practice correct testing methods to promote taxonomy.

Methodology, Analysis and Results

The Design of this study was survey, (ex-post factor), comprising of course outlines and examination questions from three colleges of Education in Eastern States of Nigeria.

Population, Sampling and Sampling Strategies - This constitutes all specialists teaching measurement and evaluation in Colleges of Education in Eastern States, who are twelve in number. The accessible population consisted of three out of six Colleges selected systematically taking the odds. The sample was made up to three Independent, judges who are also currently teachers, lecturers

in higher institutions and selection was systematic sampling where out of six schools in Eastern State, three were picked as accessible population & was used as base to sample one expert from each school as rater. The schools are Alvan Ikoku College of Education, Aro College of Education Abia State, Nsugbe College of Education.

Instrument

Instrument used was supplied by three expert teachers in Measurement and Evaluation in the three schools under study. The instruments were course outline, Essay questions for 5 years and objective questions for one year. The instrument has inbuilt reliabilities because Mehrens and Lehman (1978) and Anastesi (1979), confirmed this due to cognitive nature. An attempt to collaborate this fact was done called Rater's reliability coefficient. Using sum of squares methods and Rank order correlations, coefficient of correlations for both was 0.91 and 0.71 respectively. The responses from Raters received were put in belter perspective School/by School for better response and ratings.

Data Analysis Procedures

The Course Outline for each school and five years questions were analysed to know content and weight coverage using Benjamin Bloom's taxonomy of educational objectives in measurement and Evaluation for Essay and Objective while frequency and percentage were highlighted. The three-sampled rater's consensus opinion were tabulated below thus:-

Table 1: Consensus of 3 Ratters For 3 Schools In M/E Essay

Basic Recall	Comp.	Appli.	Analysis	Synthes	Eval.	Total
f%	f%	f%	f%	f% %	f%	f
School	9(29.0)	2(6.5)	7(22.6)	0(0)	1(3.2)	31(100)
School	7(22.6)	13(41.9)	6(19.4)	5(16.1)	0(0)	31(100)
School	4(15.4)	8(30.8)	1(142.3)	2(7.7)	1(3.8)	31(100)

Analysis

Within the 5 years under study for school A, Basic Recall knowledge and comprehension featured prominently with 38.7% (12), and 29% (9) items respectively out of 31. This was followed by Analysis Items with' 7 (22.6%) within this period. For school B comprehension featured prominently with 41.9% or 13 items followed by basic Recall 7 (22%). Application followed with 6 items (19.4%). Others objectives were not prominent. For school C, Application Objective were not

prominent with 11 (eleven) items representing 42.3%. this followed by comprehension with 8 items (30.8%). Further, there were not items for synthesis for school A and B no evaluation items for schools B and C. only one items for evaluation for school A and one synthesis for school C.

Research Question 2 Taxonomy
weighting of objective Questions

Table 2: Taxonomy Objectives Qucestiuo For One Year In 3 Schools Under Study

Basic Recall	Compr.	Appli.	Analysis	Synthes	Eval.	Total
f%	f%	f%	f%	f%	f%	f%
School A 13(93)	0(0)	0(0)	0(0)	0(0)	1(7)	14
School B 313(65)	(0)	7(35)	0(0)	0(0).	0(0)	20
School C 47(34)	16(11)	57(41)	0(0)	0(0)	0(0)	140

Analysis

From the data, objective items which are believed would cover most objective did not (see manuscript), for instance, prominent coverage came from school "C" with 57 (41%) on applications alone, while basic Recall came second with 47 (34%). Analysis featured 3rd with 20 (14%), while comprehension had 16 (11%). No items covered synthesis and Evaluation. School C had a better content validity than other schools showing (100%) content coverage validity.

For school B, its Taxonomy was poor. More features were mainly in Basic Recall with 13 (65%). Followed by Application with 7 (35). No other cognitive level was covered. Similarly, for school "A" it covered basic recall with 13(93%), followed by evaluation with 1 (7%), validity was 0.82 but school 'A' items were not much but sub-standard, so its validity is deceptive. So schools A and B have poor validities while School C has high validity.

Discussion of Results a Findings

Observation on the three expert's raters responses had far reaching results.

Firstly, within five years span prominence from Essay questions was given to basic knowledge for school 'A' and 'B' with 12 (38.7%) and 7(27%) respectively. While school C had 4 (12.4%) better than the other.

For comprehension level, school C topped with 13 (41.9%) followed by B with 8 (30.8%) and C with 9(29%).

For Application level, school C was excellent with 11 (42.3%) followed by B & A with 30% and 39% respectively

For Analysis, A was better than the rest.

The above analysis is reminiscent of what experts and teachers in tests and measurement have observed. For instance, Odukoya (1994), asserted that most tests in SSCE questions paper were only on lower cognitive levels. One wonders what non-experts will do, particularly in an unmotivational environment.

For the other second research question on extent of objective questions towards course objectives on spread and taxonomy of weights, proved that school A has 93% (13), has 65% (13) and school C has 34% (47) items. It is a truism that objective items are more content valid than Essay. Although school A and B maintained this objective for a fairly good spread, yet proper observation showed that school A reduced the federal course out line to suit their whims. School B maintained "Status-quo's and school C over-bloated and exceeded the outline without recourse to federal National Commission guidelines.

Further, only school C covered more grounds in items than others. The percentage taxonomy for school A & B are poor, lacking synthesis and evaluational elements.

The stress in preparing test- Blueprint is not easy. Iwuji (1997) Groulund (1976) and Klauseinier (1971), all agreed to the obvious difficulty in preparing test blue print, particular the higher levels. It needs wisdom, skills, commitment, motivation and humane qualities.

Summary/ Conclusion and Implication

From all reviews, opinions facts and analysis, it is clear vivid without equivocation that using Benjamin Blooms taxonomy and their weighting palaver is a process for researcher, pyschometricians, classroom teacher, at instructional level in all institutions of all levels. And since the teacher holds the key to testing, validity and reliability (psychometric properties) and the explosion or knowledge and school enrolment is increasing, with poor motivations, all institutional head at any levels should advocating for training and retraining using workshop.

Secondly, to supplement the Essay whose faults are becoming overwhelming, a new bride device (instrument) to replace this essay, that has inbuilt validities is advocated, and could be a combination of (ES (Essay) and OB/(objective) two words joined. This instrument comprises of twenty or more questions, covering the course domain with resume's short brief half page answers. The reliability of the instrument could come from replicating the tests and correlating both scores at intervals of two weeks. Others can take other forms in parts, like section A, B, and C covering the course domain.

References

- Booms B. (1956). *Taxonomy of educational objective*. Handbook 1 Cognitive Domain. New York Mcnkay Publication.
- Chukwu, G. N. (1998). *An assessment of mock Senior Secondary Certificate Examination (SSCE, WASSCE) in Federal Capital Territory controlled school*. Ibadan: A paper presented at the 40th Annual Conference Proceedings of STAN.
- Cronbach, L.J (1970). *Essentials of psychological testing*. New York 3rd edition. Harper and Coy Publishers.
- Chases, C. (1978). *Measurement and evaluation readings*. 2nd edition. Addison Wesley Pull Coy.
- Encyclopedia of Evaluation (1995). San Francisco Jersey Bass. Inc. P. 458 - 461.
- Groulund N. E. (1976). *Measurement and evaluation in teaching*. New York 3rd edition. Macmillan Publishers Coy Inc.
- Ihekwoba, C.N. (2001). Validity of Senior Secondary Certificate Examination questions in Physics (1994 — 1998). Owerri. *An M. Ed Degree thesis* in measurement and evaluation. Imo State University, Faculty of Education. New.
- Mehrens.W A and Lehman I. J (1970). *Measurement and evaluation in education and psychology*. 3rd edition, New York. Holt Rinehart Wistor.
- Anocha C. O. and Okpala P. N. (1995). *Tools for educational research*. Jatta Uzairue Edo State: Stirling Horden publishers Nig Ltd.
- Offor, B.I. (2001). Content validity of senior school certificate examination (SSCE) in Chemistry Owerri. *An M. Ed Thesis* in measurement and evaluation at Imo University. Faculty of Education April.
- Odukoya, A. (1994) in WAEC (2002). *Lagos Department of Research Division* (2006) Published by WAEC on Taxonomy of WASSCE Questions.
- Say, G. (1984). *Principles of educational measurement and evaluation*. Belmont (California) Wads worth publication. Coy.
- WAEC, (2002). *Taxonomy of WASSCE questions by cognitive domain*. Lagos: Department of Research Division (2006) WAEC.

