International Journal of Innovative Social & Science Education Research 9(3):200-207, July-Sept., 2021

© SEAHI PUBLICATIONS, 2021 www.seahipaj.org ISSN: 2360-8978

Effect of the Construction Patterns on the Standard of Test Items of Mathematics Teachers in Rivers State Senior Secondary Schools

***Goodluck Uwaks Miwari & Prof. J. B. Vipene**

Department of Educational Foundations,
Faculty of Education, Rivers State University,
Nkpolu-Oroworukwo, Port Harcourt, Nigeria
***goodmiwaks@gmail.com**

ABSTRACT
The study examined the Pattern of test construction as it affects the quality assurance of Mathematics teachers' test items in Rivers State Senior Secondary Schools. It investigated whether mathematics teachers test items attained standard by reflecting and covering the curriculum content, spreading over the cognitive behavioural objectives according to prescribed percentage and whether the item sample size reflect the weight in terms of time allotment for teaching. To achieve the objective of the study, three research questions and three null hypotheses were put forward. Descriptive survey research design was adopted. A sample of 150 state owned senior secondary schools and 350 mathematics teachers were selected using Taro Yamene's method, simple random and where applicable, purposive sampling techniques. Mathematics teachers' most recent past questions and mathematics syllabus/scheme of work were collected for examination. Validity of mathematics syllabus/scheme of work was implied while validity of mathematics test items was one assignment the study sought to establish. Inter-rater reliability index of k=0.80 for the teachers' mathematics test items was obtained using Kohen's kappa formula. Simple percentage was used to answer research questions while Chi-square, and ANOVA were used for the testing of hypotheses. Finding revealed that many mathematics teachers test items were not valid in content, lopsided and were of low standard as a result of poor pattern of construction. Based on the findings, it was recommended that professionals should be engaged in the teaching of mathematics with some intermittent refresher trainings. The establishment of internal examination committees and the enforcement of the use of test blueprints by teachers during items construction and development should be encouraged.
Keywords: Test construction pattern, standard test, quality assurance, quality control, internal assessment.

INTRODUCTION

One major responsibility of the teacher is the evaluation of his children and to accomplish this task he is faced with the challenge of developing evaluation tools which must be in uniformity with prescribed standard. Uniformity and standards in public and internal examinations presupposes quality assurance and quality control (QAQC) in the examinations and their outcomes. Internal assessment is undoubtedly the responsibility of the teachers just as the external is the responsibility of the accredited examination bodies. The National Policy on Education (FRN, 2014) prescribes and emphasizes on continuous assessment that should also ensure some uniform national standards within the area of public examinations as well as in the internal one. This policy targets the teacher training institution, faculties and institutes of education in universities, the Ministry of education and the public examination bodies as the agents to work out and implement the scheme.

Teachers as major stake holders in the business of testing, they should be well groomed with all continuous assessment techniques and in particular the development of instruments for assessing especially the cognitive domain. Unfortunately, according to Nwagu in Ezegwu (2011) very many teachers are not familiar with the techniques of test construction. And any instrument developed

without adopting the appropriate techniques will fall short of standard and any judgment passed based on it will be faulty and dangerous (Okoro, 2010).

The major tools which the teacher uses for the cognitive assessment of his students is the periodic test and examination (Okoro, 2010). A test is a device or instrument used for obtaining data or information concerning an individual's attributes. Measurement uses test to obtain data on an individual's traits in a quantitative manner, assessment ranks the individuals based on the measurement data while evaluation passes value judgment on the individual. Assessment is however used as a generic term to describe all. Any too used by the teacher to obtain scores must meet required standards as it could also serve the secondary purpose of exposing the students to standard or quality questions as would be expected in public examinations. A standard test is one that has adequate content coverage, reasonable difficulty level, accuracy in measurement of academic abilities, and must possess high discriminatory power so that differences between examinees can be obvious.

For a standard test, according to Okoro (2010), it is advocated that at least 60% of the items must be structured to cover the higher level cognitive behavioural objectives which include "analysis, synthesis and evaluation" of Bloom's Taxanomy, while "knowledge, comprehension and application" are at the lower level with 40%, (FRN, 2014). Nwagu and Okoye cited in Ezegwu (2011) have argued that test items prepared by most secondary school teachers are of low quality (low standard) because most teachers are not familiar with the principles of test construction. Test construction pattern may be described as the procedures followed to develop test items such that there is adequate spread over subject content, language structure and appreciably difficult. The modern approach requires more technical stages of design which include planning, items development, item analysis and marking scheme development.

There have been arguments that some teachers do not apply the required test construction techniques when developing test items for assessing students, Nwagozie (2012). He concluded that there is poor orientation and lukewarm attitude towards the proper testing techniques, safe-keeping and storage of assessment scores by some administrators. This conclusion is reasonable against the backdrop of quality control and quality assurance (QAQC). Test items, scores and indeed everything that has to do with assessment of the individual must be securely kept and away from unauthorized interferences. Ordinarily, test items can be drawn from an existing pool but this can question the reliability and validity of the result if not properly protected.

Ifamuyiwa (2008) revealed from a study that teachers assess their pupils only on the cognitive aspect as they frequently use test/examination, short quizzes and homework/assignment. Here again the point of concern is the reliability and validity of the test/examination questions. Ezugwu (1999) decried from his study on the content validity of teacher-made physical education tests items in Enugu State, that teachers questions do not reflect and cover the content area and are lopsided to the low level objectives. If the questions do not reflect and cover the content, then content validity has failed and it is needless to question about reliability and other qualities of the tests.

Statement of the Problem

It is a common knowledge that students of secondary schools when faced with external examinations show anxiety and panic to the extent that some get involved in one form of examination malpractice or the other. Yet the schemes of work and syllabuses followed during instructions, are the same as those used by the external examination bodies when drawing items, the teachers that taught them are qualified by their certifications, the students themselves have been passing their internal examinations to have attain promotion to subsequent classes till the final level that qualify them to write external examinations. When all these conditions were met, nothing less than courage and confidence should be expected from students during external examinations. What then is the gap in the examining process or procedure that warrants the panic and anxiety? Speculations have it that the standard of items teachers used for internal assessment is low compared to those of external exams. These speculations could be true if the teachers' pattern of test construction is an issue since the pattern is a determinant of the standard and quality. This study therefore presupposes a missing link in "standard" between the internal (teachers) test items and those of the external examinations bodies.

Purpose of the Study

The purpose of this study was to investigate the construction pattern as it affects the standard of test items of Mathematics teachers in Rivers State Senior Secondary Schools. Specifically, the study examined:

1. whether the test items used by Rivers State secondary schools' mathematics teachers reflect and cover the content of the syllabus.
2. if mathematics teachers' test items cover the cognitive behavioural objective and in prescribed percentage
3. whether mathematics teachers test items reflect the emphasis in each topic in terms of time allotment for teaching.

Research Questions

The following research questions were raised to guide the study:

1. To what extent do Rivers State Secondary School Mathematics teachers test items reflect and cover the subjects' curriculum content?
2. What is the distribution of Rivers State Senior Secondary School Mathematics teachers test items between the higher and lower levels behavioural objectives of the cognitive domain?
3. To what extent do test item drawn by mathematics teachers in Rivers State secondary schools reflect the emphasis in each topic in terms of time allotment for teaching?

Hypotheses

The following hypotheses tested at 0.05 significant level guided the study:

1. There is no significant difference between the content of mathematics teachers' test items and the subject curriculum content in Rivers State Senior Secondary School.
2. There is no significant difference between the number of items at the lower level and higher level cognitive domains of the test item of senior secondary school mathematics teachers in Rivers state.
3. There is no significant difference between the emphasis in teaching in terms of time allotment and emphasis in testing by senior secondary school mathematics teachers in Rivers State.

METHODS

Research Design

The study adopted the descriptive survey research design

Sample and Sampling Techniques

A sample of 150 out of 235 state owned schools and 350 out of a total of 701 mathematics teachers was used for the study. The 23 local Government Areas of the state were stratified into groups on bases of geopolitical contiguity and language affinity. The number of schools and mathematics teachers in each group were noted and the proportionate, simple random and (where applicable) purposive sampling techniques were used to select the samples.

Table 1: Sample Size of Teachers per Group of Local Government Areas

Groups	LGAs	No of Schools	Sampled Schools	No of Maths. Teachers	Sampled Maths. Teachers
Group 1	Abua/ Odual, Ahoada East, Ahoada West Ogba/ Egbema/ Ndoni	54	34	106	53
Group 2	Asari Toru, Akuku Toru, Degema, Bonny	29	18	29	15
Group 3	Emohua, Etche, Ikwerre, Obio/ Akpor Omuma, Oyigbo Port Harcourt	82	54	388	193
Group 4	Andoni, Ogu/Bolo, Opobo/ Nkoro, Okrika	21	13	35	18
Group 5	Eleme, Gokana, Khana & Tai	49	31	143	71
Total		**235**	**150**	**701**	**350**

Data Collection Instruments

Include the senior secondary mathematics curriculum and most recent promotion examinations test items in mathematics. These were collected from the vice principals and mathematics teachers directly by the researcher.

Validation of the Instruments

The validity of the mathematics curriculum is implied since it is a document generally accepted by the state and federal government as a working document (FRN 2012). The validity of the teachers' made test is part of what this study intends to establish or otherwise.

Reliability of the Instruments

Experts in mathematics matched the teachers test items into topics and classified them under high and low level cognitive domains. Cohen's Kappa inter rater reliability method was applied and an index (K) of 0.80 was found, which establishes the reliability of the instrument.

Method of Data Analysis

Research questions were answered using frequency counts and simple percentages, hypotheses were tested using chi-square (x^2) statistics. F-ratio test was conducted to ascertain pattern of teachers testing on lower, higher or equally on both levels of cognitive objectives.

RESULTS

Research Question 1

To what extent do Rivers State secondary schools' mathematics teachers test items reflect and cover the subject curriculum content?

Table 2: Distribution of Teachers Test Items at various Level of the behavioural Objectives in each Content Area.

Content Area	Lower Level			Higher Level			Total
	Knowl.	Compre.	App.	Analy.	Synth.	Eval.	
Number / Numeration	1073	851	675	110	77	47	2833
Algebraic process	925	763	739	74	23	10	2534
Mensuration	695	565	203	118	41	83	1705
Plane Geometry	300	400	89	154	236	238	1417
Trigonometry	908	879	713	63	0	9	2572
Statistics/ Probability	1150	1100	500	135	36	79	3000
Total	**5051**	**4558**	**2919**	**654**	**413**	**466**	**14061**
Order Total		**12528**			**1533**		
		(89.10%)			**(10.90%)**		

Table 2 gives a summary of the number of items per content area. Out of the total of 14061 items drawn by the 350 teachers, only 1705 and 1417 reflected on mensuration and plane geometry respectively. All the schools seem to concentrate on number and numeration, algebraic processes, trigonometry and statistics which had 2833, 2534 and 3000 questions respectively. On the whole, more than 85% of the questions focused on lower level behavioral objective while below 15% only tested on higher. Therefore, the extent of coverage is very low.

Research Question 2

What is the distribution pattern of Rivers State Senior Secondary School Mathematics teachers test items between the higher and lower levels behavioural objectives of the cognitive domain?

Table 2 reveals a sharp discrepancy between the number of items in the lower and higher order behavioural objectives with totals of 12528 (89.10%) and 1533 (10.90%) respectively. This is against the 40%:60% ratio for a standard/quality test as prescribed (FRN, 2014) and (Okoro, 2010). A critical view of the table shows that over 80% of the questions in each content area tested knowledge recall and development of mental ability, while only a little above 10% tested the more complex order requiring abstract reasoning and discoveries.

Table 3 below shows the number of teachers whose items fall into both high and low, and those whose items exclusively tested on higher or lower level behavioural objectives

Table 3: Total weight of Questions and Number of teachers testing on Lower, Higher and both levels of cognitive Domain

	School Groups	No of Sampled Schools	No of Sampled Teachers	Lower Order No. of Items	Lower Order No of Teachers	Higher Order No. of Items	Higher Order No of Teachers	H & L Order No of Items	H & L Order No of Teacher
1	(4 LGs)	34	53	1585	45	131	3	380	5
2	(4 LGs)	18	15	86	7	104	6	12	2
3	(7 LGs)	54	193	6458	150	892	16	1000	27
4	(4 LGs)	13	18	322	15	50	1	18	2
5	(4 LGs)	31	71	2119	45	294	6	610	20
	Total	**150**	**350**	**10570**	**262**	**1471**	**31**	**2020**	**56**
	($\bar{x}$)			**52.6**		**6.2**		**11.2**	

From table 3 it can be observed that total number of items under the lower order behavioural objectives outweigh those in the higher order; only group 2 approached a 50:50 ratios, while no group adopted the expected 40:60 at the lower and higher order. All other groups concentrated on lower order. Taking the mean **($\bar{x}$)** of teachers per group testing on the levels yielded 52.6, 6.2 and 11.2 respectively for lower, higher and both. This is a reflection that more of the teachers tested on lower level than higher level.

Research Question 3

To what extent do test item drawn by mathematics teachers in Rivers State secondary schools reflect the emphasis in each topic in terms of time allotment for teaching?

Table 4 Observed and Expected distribution of items to each content area by classes

Content Area SS1	Teaching Time in weeks (T) SS1	SS2	No. of Items (Observed) SS1	SS2	Number of Items (Expected) SS1	SS2
Number/ Numeration	8	9	1292	1541	1631	2120
Algebraic Processes	5	9	1335	1199	1019	2120
Mensuration	5	0	759	946	1019	0
Plane geometry	7	7	657	760	1427	1649
Trigonometry	3	4	824	1748	612	942
Statistics/ Probability	4	3	1656	1344	815	707
Total	**32**	**32**	**6523**	**7538**	**6523**	**7538**

Table 4 reveals that time allotted for teaching with corresponding number of questions set for testing do not match. For instance, 8 weeks and 9 weeks (as reflected in their record of work), were spent in teaching number and numeration, but only 1292 and 1541 of question instead of 1631 and 2120 reflected for SS1 and SS2 respectively. It is also observed that questions are drawn from content areas even when no teaching was evident to have been done, as in the case with content area of mensuration in SS2. In both classes the discrepancies are glaring indicating that test items developed by mathematics teachers in Rivers State Senior Secondary School do not reflect emphasis as indicated by time allotted for teaching.

Test of Hypothesis

Hypothesis 1

There is no significant difference between the content of mathematics teachers' test items and the subject curriculum content in Rivers State Senior Secondary School.

Tables 5: Chi-Square (x^2) Analyses for number of test items for SS1 and SS2.

Class	Content Area	1	2	3	4	5	6	x^2_{cal}	x^2_{crit}	Df	α	Rmks
SS1	f_o	1292	1335	759	657	824	1656	1587	11.07	5	0.05	Reject
	f_e	1631	1019	1019	1427	612	815					Ho
	f_o	1541	1199	760	760	1748	1344	2301	11.07	5	0.05	Reject
SS2	f_e	2120	2120	0	1649	945	707					Ho

The expected frequencies were computed based on the time allotment for teaching and the chi-squares goodness of fit whose values for both SS1 and SS2 are presented in table 5 was computed. In both cases of SS1 and SS2 the calculated chi-square (x^2) were far more than the table values hence the null hypothesis was rejected meaning there is significant difference between the observed and expected frequencies distribution of mathematics teachers test items against the curriculum content prescription.

Hypothesis 2

There is no significant difference between the number of items at the lower level and higher level cognitive domains of test items of senior secondary school mathematics teachers in Rivers state.

The observed percentage frequencies were 89.10% for Higher and 10.90% for Lower (table 2), while the expected is the 60%, 40% required for a test to be of standard (Okoro, 2008 and NERDC 2014). Below is the result.

Table 6: Chi-square Analyses Between Higher and Lower Levels Cognitive Behavioural Objectives of Test Items of Mathematics Teachers in Rivers State Senior Secondary School

	Behavioural Objectives		df	α	x^2_{calc}	x^2_{crit}	Rmks
	% at higher Level	% at Lower Level					
Observed	10.90	89.10					Reject
			1	0.05	100.45	3.84	
Expected	60.00	40.00					Ho

From table 6 the calculated x^2 is far above the critical value hence the hypothesis is rejected. The implication is that the difference observed in the pattern of questions skewing to the lower level behavioural objectives (table 3) is significant and cannot be by chance or sampling error. To further investigate the level where teachers concentrated their testing contributing to any observable difference, an F-test of difference among the means of the number of teachers testing on the 3 levels was computed and the result displayed in table 7

Table 7: Analysis of variance for Number of Teachers Testing Higher, Lower and both Levels of Cognitive Behavioural Objectives

Source of variation	Sum of Square	Df	Mean Square	$F_{-calc.}$	$F_{-critical}$	P
Between group	6486	2	2343.27			
				89.32	4.71	<0.05
Within group	13670.80	347	39.40			
Total	20157.34	349				

F(2,347) at P< 0.05 = 4.71.

With the calculated F-value of 89.32 higher than the critical F-values of 4.71 it implies that there is significance difference among teachers' pattern of testing. A post hoc test conducted showed F(1,2) = 68.30 > 4.71, F(1,3) = 54.38>4.71, and F(2,3). = 0,79<4.71. This result indicates that category one (those testing on lower level) is more significant, since for any comparison involving group 1, the F value is significant.

Hypothesis 3

There is no significant difference between the emphasis in teaching in terms of time allotment and emphasis in testing by senior secondary school mathematics teachers in Rivers State.

Data from table 4 were used to compute chi-square (x^2) with the result shown on table 8 below.

Table 8: Chi-Square test for Percentages in testing by Teachers of Mathematics in SS1 and SS2 (Ref table 4)

	x^2_{cal}	x^2_{crit}	df	α	Decision
SS1	24.41	11.07	5	0.05	Reject Ho
SS2	189.19	11.07	5	0.05	

In each case null hypothesis was rejected as value of computed chi-square was greater than the critical, meaning the difference observed was too significant to be attributed to sampling error. The implication is that mathematics teachers test item do not reflect the emphasis in terms of time allotted for teaching, and the percentage of test items do not agree with the amount of time spent in teaching given topics as revealed in table 4.

DISCUSSION OF FINDINGS

In this study the quality assurance and quality control of mathematics teachers' test items were deeply investigated using the coverage of content, pattern of spread and reflection of question on weight of content in terms of time allotted for teaching and corresponding number of items drawn, as yardsticks. A major finding of the study was that there is inadequate spread of mathematics teachers' test items in the entire content areas, some of the content areas, were actually neglected. Most of the schools that attempted to reflect the entire content area do not spread the questions relatively and some teachers avoided certain content areas. A finding similar to this was made by Ezegwu (2011) who found out that the distribution of the physical education questions to the topics was lopsided as they clustered to some content area while some content areas were not reflected. In this study similarly, the analysis revealed significant difference among mathematics teachers pattern of setting questions, meaning most of the mathematics teachers do not follow a uniform pattern in developing test items. Majority of the teachers' test items are not reflected in most of the content areas and the issue of lopsidedness is not ruled out.

A second major finding is that questions are not spread all over the levels of the behavioural objectives. Putting knowledge, comprehension and application under lower level, which tested knowledge recall and rote memorization on the one hand; and analysis, synthesis and evaluation under higher level (table 4) testing the more complex ideas requiring abstract reasoning on the other hand; it was discovered that question are concentrated on the lower order. The ratio of teachers testing lower order also far out-weighs those testing on higher level. This was further confirmed by the result of the chi-square (x^2) analysis whose calculate value was higher than the critical value, leading to the rejection of the null hypothesis. Ezegwu (2011) had similar findings in his study of teachers made physical education test items. He found out that the behavoural objective tested by teachers of physical education were lower order cognitive domain. The current study revealed an observed F-score of 89.32 that was higher than the table value of 4.71, confirming a significant difference among teachers pattern of testing in terms of covering the cognitive behavioural objectives. The pattern is seen from tables 2 and 3 to be skewed to the lower order indicating that the test items used by mathematics teachers in Rivers State secondary schools are concentrated in the lower order cognitive behavioural objectives and therefore fall below standard.

A third finding in this study was that the time allotted for teaching each content area and the number of test items used for testing the content area do not match. This confirms Okwu (2004) findings while validating objective CRS test items for JSS3 students in Nigeria. Okwus result revealed a

serious discrepancy between the times allotted for teaching certain topics of CRS and the number of questions drawn during assessment. In this study, a glaring discrepancy was noticed as indicated in table 3. Expectedly though, when much time is spent in teaching, the number of questions during testing should accordingly be more (Okpara, Onocha and Oyediji, 2006) and Ubulom (2010). In this study, it was discovered that in some cases it was totally the reverse. When eight weeks was spent in delivering a content area and four weeks for another, it is equally expected that more questions should come from the content area consuming eight weeks and not in this case where the content area delivered less than four weeks took more questions. Where two different content areas spent the same amount of time (weeks), the number of questions should at least be near equal if not equal. The result in this case revealed lopsidedness where one area takes almost twice of the number of questions of the other. The finding may not be unconnected with the qualification of teachers most of who do not have professional qualifications. Their lack of professional background implies lack of knowledge of test construction techniques.

CONCLUSION

Based on the findings of this study, itt is concluded that secondary school mathematics teachers lack the requisite skills in testing. The schools and teachers whose questions did not reflect adequately may have not been able to cover the syllabus during the teaching period and reasonably they need not give questions from areas not taught. It is also possible that some of the content area have been tactically avoided due to incompetence by the teacher; or that some teachers do not have in-depth knowledge of some topics hence students are tested only superficially. Majority of mathematics teachers do not use test blue print if they are actually aware of its existence. The high discrepancy between percentage of time spent in teaching and relative percentage of question is an indication of non-usage of table of specification. Finally, the test items used by mathematics teachers in their internal assessment are of substandard, not meeting the requirement for standard of test items.

RECOMMENDATIONS

Based on the findings of this study, the following recommendations are offered:

1. Teachers should intermittently be made to have compulsory refresher trainings.
2. There is need for every school to have an internal examination team whose members must have sound knowledge of testing so that they can vet every test/instrument before the subject teacher is allowed to use.
3. Teachers should be encouraged to use test blue print in order to reduce the lopsidedness of questions in favour of some content areas at the expense of other.

REFERENCES

Ezegwu, G. G. (2011). Content Validity of Teacher-made Physical Education Test used In Secondary School In Enugu State. *Journal of Teacher Educational, 7 (1), 143 – 151*.

Federal Republic of Nigeria (2014). *National policy on Education.* Abuja: NERDC.

Federal Republic of Nigeria (2013). *Universal Basic Education*. Lagos: NERDC.

Federal Republic of Nigeria (2012). *Senior secondary Education Curriculum*. Abuja: NERDC.

Gronlund, N.E. (1977). *Constructing Achievement test.* New Jersey: Prentice- Hall, Eaglewood Cliff.

Ifamuyiwa, A. S. (2008). An Assessment of the level of Awareness and usage of continuous assessment techniques among Primary School Teachers. *Nigeria Journal of Educational research and Evaluation,* 4(2), 55-56

Nwagozie, J. I, (2012): Relationship between continuous assessment scores and junior secondary certificate examination final scores in Imo State. *Unpublished M. Ed. Dissertation. University of Port Harcourt*

Nworgu, B. G. (2011): *Educational Research: Basic issues and methodology*. Nsukka: University Trust Press

Okoro, M. O. (2010). *Measurement and evaluation in education*. Oruowulu- Obosi, Anambra State: Pacific publisher ltd.

Ubulom, W. J. (2010). *Principles and practice of Test Construction.* Port Harcourt: Celwill Publishers.