



## INVESTIGATING THE SPREAD OF ITEM TYPES IN CONNECTION TO COGNITIVE DEMAND OF TEACHER-MADE MATHEMATICS TEST IN MISAU LOCAL GOVERNMENT AREA, BAUCHI STATE.

### ABSTRACT

This study investigated the spread of item types in connection to cognitive demand of teacher made Mathematics Test in Misau local government area, Bauchi state, Nigeria. The study adopted Causal comparative (Ex-post factor) design from total population of fifty-seven (57) teachers consisting of 46 males and 11 females. A total of 174 sample of TMT was retrieve from teachers. and is

### BABAYO MUHAMMAD OTHMAN

*Department of psychometrics, AD Rufa'i College of Education, Legal & General Studies Misau, Bauchi State*

### INTRODUCTION

The importance of assessment in the overall quality of teaching and learning in education has been widely acknowledged, White (2012) describes assessment as an engine that drives learning noting that a well-designed assessment has the potential of setting clear expectations, establishing a reasonable workload and providing opportunities for students to self-monitor, rehearse, practice and receive feedback, while a poorly designed assessment has the potential of hindering learning.

Teacher's skill in classroom assessment enhances student achievement; with these teachers are very important stakeholders in the educational system, as they drive whatever takes place within the classrooms in the school system. Apart from teaching, they generate assessments which are used to evaluate teaching efficacy and learning achievement within their classrooms. Teacher-made tests (TMT) can be seen as more than assessment devices; they are fundamental parts of the educational process. Thus they can define instructional purposes, influence what



restricted to mathematics teachers from public senior secondary schools, all the population was used for this research. The researcher developed a Test Investigation Guide (TIG) for TMT. One research question was answered and two hypotheses were formulated and tested at 0.05 level of significant. The data collected were analyzed using descriptive statistic; frequency counts and percentages to answer research questions, while Chi-square (cross-tabulation) was used to test the two hypotheses. The study revealed that teachers' in Misau Local government frequently used problem exercise, representing 32% followed by true/false and context dependent type is the list having only 7%, similarly 56% of the teachers set their test item to measure higher cognitive demand whereas about 44% of their items measures low cognition respectively, the finding also indicated that the hypotheses on the types of test items (Questions) mostly found on TMT did not differ significantly by the teachers' working experience as well as qualification in Misau local government is retained, The frequency of cognitive level of test items found on TMT between experience and in-experience teachers is retained but rejected in-relation to qualification. The study recommended that Items dealing with the same content may be grouped together; doing this will help the examinee to concentrate on a single domain at a time rather than having to shift back and forth among areas of content. Similarly, teachers' level of awareness should be raised to the level that they expand the status of their test items to another type such as matching, short answers, context dependent etc. so that other aspect of cognition would be tested using a variety of items type.

**Keywords:** Teacher-made Mathematics test, items type, Cognitive demand, Bloom's Taxonomy, Mathematics test items.

students study, and help teachers to gain perspectives on their course. To determine and achieve these purposes, the tests should be qualitative. Qualitative inform of selecting item type based on cognitive demand in relation to curriculum based objectives, doing this will make the measurements effective and good, but when the measurement is poor, then there will be an inaccurate data-based inference, which in turn leads to bad decision-making. The use of poorly designed



Mathematics achievement test is a major problem as it affects students' interest and achievement in Mathematics. It has already been pointed out that poorly designed tests could make the students loose interest in a particular subject (Osadebe, 2001).

To facilitate students' learning effectively, teachers need the knowledge regarding test items cognitive demand. Bloom's six levels of cognitive demand if properly digested, will help in retrieving the required responses from students; this will enhance learning diagnoses in class.

It is pertinent that TMT is more frequently used in educational settings than any other type of test. One is then tempted to wonder if the frequent and regular use of TMT automatically implies that they are of high quality, if that is the case, then there is a need for competency in test development regardless of being professional or not. Ololube (2008) evaluated competencies of professional and non-professional teachers in Nigeria. He reported that professional teachers tend to construct various effective evaluative instruments more than the non-professional teachers.

On this regard, teacher education training programs in Nigerian Colleges of Education and Universities, offered prospective teachers with courses on measurement and evaluation, methods of test construction, types of test, etc. (Azuka, 2014). The students on graduation are expected to put these techniques into practice in assessment of students in classroom

Item type has a basic consideration in term of test item construction, since it involve a test of knowledge and skills. Under test construction, emphasis is placed among others, on the students' ability to understand the test items and the potential levels of students' performance in such a test. Item types are either selective or constructive, Rana & Murtaza, (2011). Selective response items are questions which require the test-taker to choose the correct answer from a number of options such as the multiple choice questions, true/false, matching items etc. Constructive response items involve tasks which require the test-taker to provide constructive solutions and ideas, such as the essay-type questions, short answer and completion items in an examination. Teachers' knowledge as well as experience is determined by the quality of item types they produce,



especially on choosing the appropriate type based on cognitive, affective or psychomotor skill.

It is against this background that the researcher will investigate the quality or otherwise of TMT, bearing in mind Teachers qualification and working experience in relation to tests' development at public secondary schools level in Misau Local Government Area of Bauchi State.

### **Objectives of the study**

The study is aimed to determine:

1. the type of test items (questions) teachers frequently used when developing TMT in Misau local government area of Bauchi State
2. whether there is difference in the type of test items (questions) teachers frequently used when developing TMT with regards to working experience in Misau local government area of Bauchi State
3. whether there is difference in the type of test items (questions) teachers frequently used when developing TMT with regards to qualification in Misau local government area of Bauchi State

### **Research Questions**

The research findings sought to answer the following questions:

What type of test items (questions) do teachers frequently used when developing TMT in Misau local government area of Bauchi State?

### **Research Hypotheses**

The following hypotheses were tested at 0.05 level of significance:

HO<sub>1</sub>- The types of test items (questions) frequently used on TMT do not differ significantly between experience and in-experience teachers in Misau local government area of Bauchi State

HO<sub>2</sub>- The types of test items (questions) frequently used on TMT do not differ significantly between qualified and un-qualified teachers' in Misau local government area of Bauchi State



## **Literature Review**

### **Test construction, Item type and Students' performance in Mathematics**

The quality of a test given by a teacher is closely linked with its ability to provide the kind of information needed regarding students' performances. A well-written test allows the teacher to accurately and consistently measure students' mastery of specific contents taught in class. On the contrary, poorly designed test items may lead to inaccurate measurements of learning and provide false information regarding students' performances as well as instructional effectiveness with regards to cognitive demand. It is observed that any characteristic of a test item which distracts the examinee from the major point or focus reduces the effectiveness of that item.

Emafo and Oribhabor (2016), in their study, meant to assess the quality of test constructed by mathematics teachers' in Egor local government area of Edo state, Nigeria. A sample of Mathematics tests constructed by SS2 Mathematics teachers 'was used. The instruments used for the collection of Data for the study were copies of multiple choices Mathematics achievement test constructed by mathematics teachers and their students marked answer script at the end of 2012/2013 academic session, they found that items mainly focus on lower level of the cognitive domain of Bloom's taxonomy of educational objectives, similarly they noted that a major reasons why TMT is poor are because of lack of preparation in the development of the test,

Ozuru et al. (2007) deduced that "the processes underlying open-ended and multiple choice test items answering in the with-text condition are likely to share less similarity [than those in the without-text condition]" (p. 426). The studies outlined generally found that multiple choice items in reading assessment seem to be easier than open-ended items.

According to Chan and Kennedy (2002), however, a high correlation between scores on multiple-choice and constructed-response items could be misleading. For example, scores could be strongly correlated, but one score could still be significantly higher than the other score. Upon hypothesizing that students' scores on a multiple-choice test would be higher than scores on a constructed-response test, Chan and Kennedy (2002) conducted a study to determine the extent of the correlation between test scores. The types of constructed-response



items used in the study were questions that required an answer in the form of a brief sentence or phrase. The basis of this decision was that this type of constructed response item was most prevalently used in the literature.

Data collected during the study consisted of students' test scores (Chan & Kennedy, 2002). The multiple-choice questions were graded objectively, with a point given for each correct answer. The constructed-response questions were graded subjectively by one person

Chan and Kennedy (2002) used a statistical procedure to correct scores on the multiple-choice items for the effect of guessing, Recognizing that students could also guess on some of the constructed-response questions, Chan and Kennedy conducted two sets of analyses on scores. Unadjusted test scores were used in one set of analyses, and scores that had been adjusted to correct for the effects of guessing were used in the other set of analyses. Chan and Kennedy had defined expected difference questions as items where students could guess on the multiple-choice version of the item, but could not guess on the corresponding constructed-response item. There were also items of no expected-difference in which the students could guess on both the multiple-choice and the constructed-response items. There were eight expected-difference questions and seven no-expected difference questions. Analysis of the test scores indicated that students scored higher on the multiple-choice questions when there was an expected difference between the multiple-choice and constructed-response items. Students scored comparably on items where there was no expected-difference. Although students scored higher on the multiple-choice items than on the constructed response items for the expected-difference questions, the results were not statistically significant at an alpha level of 0.05.

Chan and Kennedy (2002) suggested that these distracters may have prevented students from engaging in the same thought processes that would have been used in free recall. Analysis of the no-expected-difference questions indicated that there was no significant difference in the scores on adjusted multiple-choice questions compared to scores on the equivalent multiple-choice questions.

Some scholars supported that poor pedagogical content knowledge of teachers in term of students recognition on item demand led to a poor academic performance by pupils, (Ogundele, Olanipekum & Aina, 2014). Teachers'



qualifications, teacher self-efficacy, and teacher effectiveness are all correlated with a term of item cognitive demand. Therefore, teachers' qualifications must have an influence on the quality of item develop cognitively. (Aina, Olanipekum & Garuba, 2015) had observed that teacher effectiveness has an influence on students' academic performance. Teacher personal quality is very crucial to teachers' qualification. Therefore, it is paramount always to consider it as part of teacher qualifications.

Jacobs et al., (2010) examined various levels of teachers' experience and their ability to notice based on teaching experience and a degree in professional development. The findings suggested that those with the least amount of teaching experience were less able to interpret and respond to students' mathematical thinking. Those with teaching experience alone were still limited in their ability to interpret and respond to students' thinking. However, teachers who engaged in sustained mathematics professional development were more able to interpret and appropriately respond to students in the classroom context. This study suggested that there was a continuum of development in teachers' ability to notice students' thinking through applying various item type in order to test different level of student cognition capacity..

### **Methodology**

The research design adopted in this study was the ex-post facto design. The study was carried out in the Misau Local Government located at Central Senatorial District of Bauchi state. In this research work, the population of the study is the senior secondary school's Mathematic teachers from public School in Misau Local Government, Bauchi state.

The populations of the study are all the 57 Senior Secondary School Mathematics teachers that taught at the twelve senior secondary schools within the two districts consisting of day and boarding schools in Misau Local Government Area Bauchi State. Their educational qualifications encompass those with NCE, HND, BSC ED, BSC, BED.TECH and B ENG and few with higher degrees, their working experience varied from one to thirty-two years. Their ages approximately range between twenty-two to fifty-five years, most of them belong to low socio-economic status, and few were of moderate class



### **Population and Sample Size**

The population consist of fifty-three (57) teachers; made up of 46 males and 11 females, one hundred and seventy-four (174) sampled of Teacher-Made Test was collected

### **Sampling Technique**

For the purpose of this research, a non-probability sampling was adopted. According to Creswell (2012), a non-probability sampling (purposive sampling), deliberately avoids representing the wider population. It seeks only to represent a particular group, a particular named section of the wider population, e.g. a class of students; a group of students who are taking a particular examination.

### **Data Collection Instrument**

A face to face interview was used in collecting the demographic information of Teachers for this research work. Similarly, Test Investigation Guide (TIG) which was designed by the researcher ( Othman,2018) was used for investigating the recent test items (question papers), bearing in mind the general information regarding test item format and cognitive level target from individual question papers. At this juncture, three people including the researcher participated as judges for the investigation purpose; they had post-graduate experience in test and measurement.

### **Findings**

The summary of the finding were presented below:

Research Question One: What type of test items (questions) do teachers frequently used when developing TMT in Misau local government area of Bauchi State?

**Table 1.0 Summary of Descriptive Statistics for Test Item Types**

Types of test item	Frequency		Percentage of those using
	Not using	Using	
Multiple choices	23	34	19%
True/False	12	45	26%



Matching type	47	09	05%
Problem exercise	01	56	32%
Context dependent	36	21	07%
Short answers	44	13	11%
Total			100%

Table 1.0 shows the summary for type of test items teachers in Misau Local Government Bauchi State frequently use in their test. From the table, it can be observed that problem exercise is the most frequently used test item having 56 count representing 32% of the entire frequency count. This might be connected with the belief that mathematics is doing not choosing. The table also indicated that true/false is the second test items use by teachers in their tests having 26% and matching type is the list having only 5% as it frequency count respectively.

### Hypotheses Testing

$H_{01}$ : The types of test items (questions) frequently used on TMT do not differ significantly between experience and in-experience teachers in Misau local government area of Bauchi State

Table 2.0: Summary of Chi- Square Cross Tabulation Analysis of Types of Items by Teacher's Working Experience

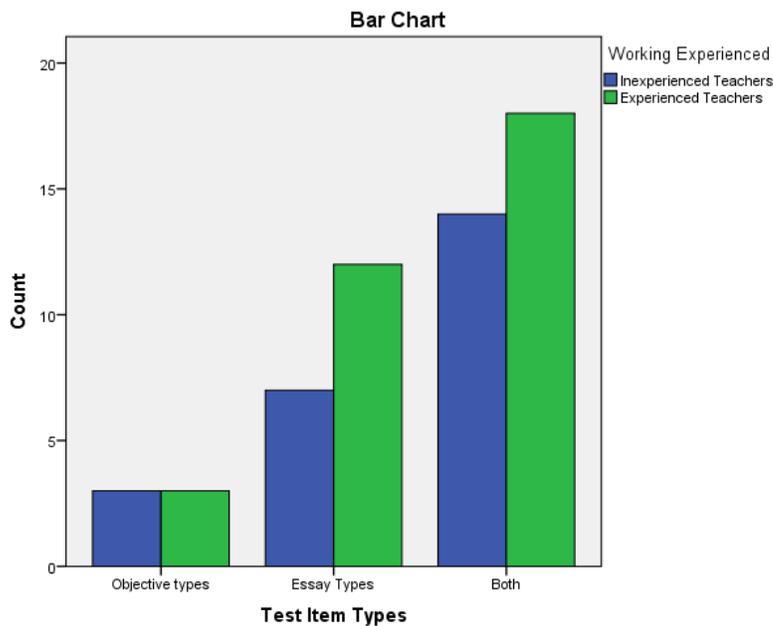
Working Experience	Frequency	Chi square	Df	P-value	Decision
Experienced	32	.405	2	.817	Accepted
In-experienced	25				
N of Valid Cases	57				

Table 2.0 above depicts the analysis for chi-square on test items types in relation to working experience the result shows the calculated  $X^2$  value was found to be .817 with a degree of freedom 2 and p-value of .817. It can be observed that the p-value which is .817 is greater than  $\alpha$  value of 0.05. Thus, the null hypothesis which states that, the types of test items (Questions) mostly found on T.M.T will



not differ significantly by the teachers' working experience is retained. This indicated that the types of test items (Questions) frequently used on T.M.T did not differ significantly by the teachers' working experience. This happen due to the fact that most Mathematics questions are obtained at the end of each topic as an exercise, developed by the authors

The chart below further illustrates the differences.



**Figure 1: Types of Items VS Teacher's Working Experience**

HO<sub>2</sub>- The types of test items (questions) frequently used on TMT do not differ significantly between qualified and un-qualified teachers' in Misau local government area of Bauchi State

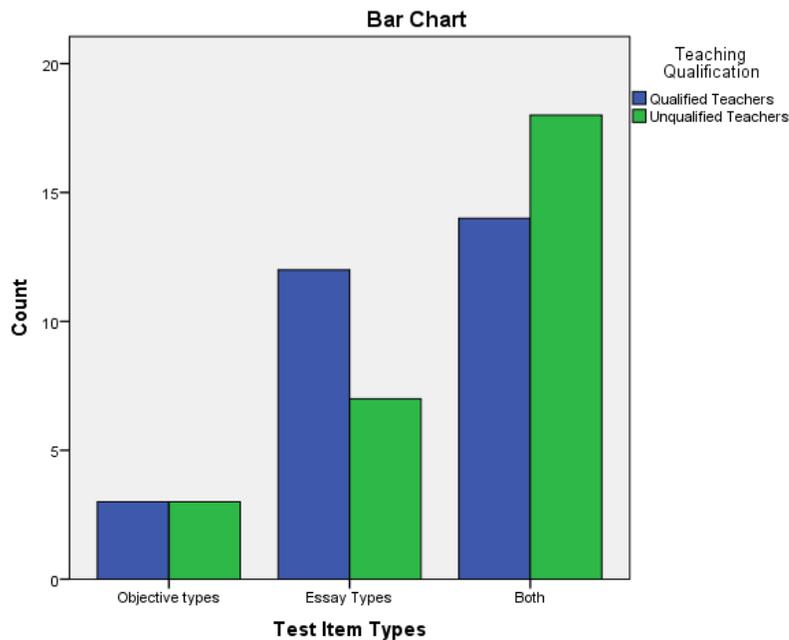
**Table 3.0: Summary of Chi- Square Cross Tabulation Analysis of Types of Items by Teacher's Qualification**

Qualification	Frequency	Chi square	Df	P-value	Decision
Qualified	29	1.799	2	.407	Accepted
Unqualified	28				
N of Valid Cases	57				

Table 3.0 above presented the analysis for chi-square on test items types in relation to teacher's qualification; the result shows the calculated X<sup>2</sup> value was



found to be 1.799 with a degree of freedom 2 and p-value of .407. It can be observed that the p-value which is .407 is greater than *alpha* value of 0.05. Thus, the null hypothesis which states that, the types of test items (Questions) mostly found on T.M.T will not differ significantly by the teachers' qualification is retained. This indicated that the types of test items (Questions) frequently used on T.M.T did not differ significantly by the with regards to their qualification. The chart below further illustrates the outcome of the analysis.



**Figure 2: Types of Items VS Teacher's Teaching Qualification**

### **Discussion**

The first findings indicate that the type of test items type normally used by teachers in Misau Local Government Bauchi state are problem exercise (36%), followed by true/false, (27%), while multiple-choice (20%), then short answers (12%) and matching type represent (5%), as indicated from Table 1.0. The above result indicated that more than half of the type of the item comprises of problems exercises and true/false item while matching type is the list use.

The above findings are contrary to that of (Chan & Kennedy, 2012), they uphold that multiple choice item is the most popular item used because it measures multiple cognition such as comprehension, interpretation, application, analysis,



and synthesis, similarly their finding reveals that it can be scored with ease and accuracy especially, with large classes

The second finding of this study unveils that the types of test items (questions) mostly found on TMT will not differ significantly by the teachers' working experience in Misau local government

Table 2.0 shows the result for chi-square for test items types in relation to working experience. After analysis, the result shows the calculated Exact Sig of .817 tested at a level of 0.05 with a degree of freedom 2. Since the Exact Sig of .817 is greater than a level of 0.05 the null hypothesis which states that, the types of test items (Questions) mostly found on T.M.T will not differ significantly by the teachers' working experience is retained. This indicated that the types of test items (Questions) mostly found on T.M.T did not differ significantly between experience and in-experience teachers in Misau local government

### **Conclusion**

From the findings of the study, it can be concluded that, Teachers-Made-Test developed by senior secondary schools mathematics teachers in Misau Local Government Bauchi state, are normally in a form of problem exercises, true/false and multiple choice, this may be due to the nature of the subject or their level of awareness on test development is limited.

It can also be concluded that the differences existed between teachers working experience on the frequency of cognitive level of item use regarding Blooms' taxonomy, majority of experience teachers measure higher cognition ( analysis, synthesis and evaluation) whereas in-experience measure lower cognition (knowledge, comprehension and application).

It also revealed that the type of test items mostly found on TMT did not differ between experience and in-experience, likewise, qualified and un-qualified teachers. Since the Exact Sig of .817 is greater than a level of 0.05 and also of the Exact Sig of .407 is greater than a level 0.05 respectively, hence the 1<sup>st</sup> and 2<sup>nd</sup> Hypotheses are retained,

### **Recommendations**

Based on the findings two set of recommendation were set, recommendation from the study and recommendation for further study



1. Teachers level of awareness should be raised to the level that in term of test development, they expand the status of their test items to another type such as using matching type, short answers etc. such that another aspect of cognition will be tested using variety of items type,
2. Knowledge on item selection will aid in determining the extent to which content and objectives are correlated for yielding standard, hence, teacher need proper idea on this regard. Developing of test items is a crucial step for the validity of a classroom test, it is determined by the extent to which performance to be measured is called forth by the test items. Having enough knowledge of subject matter, defined learning outcomes, or a psychological understanding of the students' mental processes will serve as a prerequisite for developing high-quality test items.

#### REFERENCES

- Aina, J.K., Olanipekun, S.S., & Garuba, I.A. (2015). Teachers effectiveness and its influence on student learning. *Advances in Social Sciences Research Journal*, 2(4), 88-95.
- Azuka, B.F. (2014). Assessment in Primary School Mathematics Classrooms in Nigeria. *International Journal of Education Learning and Development*, 2 (2). 50-58.
- Chan N. and Kennedy P. E. (2002). "Are multiple-choice exams easier for economics students? A comparison of multiple choice and equivalent constructed response examination questions", *Southern Economic Journal*, Vol. 68, No. 4, pp. 957-971.
- Creswell, J.W. (2012), *Educational Research: Planning, conducting, and Evaluating Quantitative And Qualitative Research* (4th.ed). Boston: Pearson Education, Inc.
- Jacobs, V. R., Lamb, L. C., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education*, 41(2), 169-202.
- N. P. Ololube, (2008). Evaluation Competencies of Professional and Non-Professional Teachers in Nigeria, *Studies in Educational Evaluation (SEE)*, 34(1), 44-51.
- Ogundele, G.A., Olanipekun, S.S., & Aina, J.K. (2014). Causes of poor academic performance in West African School Certificate Examination (WASCE) in Nigeria. *Scholar Journal of Arts, Humanities, and Social Sciences*, 2(5B), 670-676.
- Oribhabor, C. B., & Emafo, O. D. (2016). Determining the Reliability and Content Validity of the Mathematics Tests Constructed by Senior Secondary School Mathematics Teachers in Edo State, Nigeria. *African Journal of Education*, 3(2), 83.
- Osadebe, P.U. (2001). Construction and Validation of Test. A Seminar Paper Presented at the University of Port Harcourt.
- Othman. B. M (2018), Investigating the cognitive demand in Teacher-Made Mathematics Test. Unpublished theses Bayero University Kano



**TIMBOU-AFRICA ACADEMIC PUBLICATIONS**  
**AUGUST, 2022 EDITIONS, INTERNATIONAL JOURNAL OF:**  
**SCIENCE RESEARCH AND TECHNOLOGY VOL. 10**

- Ozuru Y., Best R., Bell C., Witherspoon A. and McNamara D. S. (2007). "Influence of question format and text availability on the assessment of expository text comprehension", *Cognition and Instruction*, Vol. 25, pp. 399–438.
- Rana, R. & Murtaza, N. (2011). *Handbook of Test Construction Techniques: Training held on January 17-18, 2011 at NTS Secretariat Islamabad* Pp. 25-26
- White, E. (2012). Assessment literacy for affective classroom-based assessment. [www.linkedin.com](http://www.linkedin.com). Retrieved on November 2021.