The Seybold Report

ISSN 1533-9211

DOI: 10.5281/zenodo.8405151

ENSURING THE CONTENT VALIDITY OF TEACHER-MADE TESTS THROUGH THE USE OF A TABLE OF SPECIFICATIONS

OKE JAMES AJOGBEJE

Department of Counselling Psychology, Bamidele Olumilua University of Education, Science & Technology, Ikere-Ekiti. E-mail: ajogbeje.oke@bouesti.edu.ng, ojajogbeje@gmail.com

Abstract

The validity of the teacher-made tests created by classroom teachers for educational evaluation has been discussed for a long time and the problem does not seem to be resolved. This study was designed to examine the content validity of teacher-made tests using table of specifications. 250 classroom teachers from 10 secondary schools in the Akure South and Akure North Local Government Areas of Ondo State were chosen at random to participate in the study. A survey research design was used. All relevant information was gathered using a survey research instrument called the Content Validity Survey Questionnaire (CVSQ), which was developed, validated, and used. Obtained data was analyzed using percentage, mean and standard deviation. According to the study's findings, most of the teachers involved in the evaluations do not know what the Specification Table (TOS) is. Second, the vast majority of the sample class's teachers have never attended a TOS related seminar or in-service training, therefore they are unable to develop TOS for the subjects they teach. Despite the fact that the content validity of the teacher-made tests was found to be extremely low and reasonable, the majority of classroom teachers did not use TOS when creating the assessment instruments.

Keywords: Classroom Teacher, Content Validity, Teacher-Made Tests, Table of Specifications, Educational Assessment Process.

1. INTRODUCTION

It is an incontrovertible reality that teachers frequently use test scores to judge students' performance in the classroom. The outcomes of several assessments of the knowledge a student acquires during the teaching and learning process are typically combined to produce the scores a student receives on a particular exam. Each learner's future career is highly influenced by all of these evaluations. Teachers now have the challenging task of ensuring that the assessment instruments are appropriate and valid (Griswold, 1990). With the emergence of alternative modes of evaluation that are continuous and formative in nature, decentralizing assessment is actively being sought on a global basis (Chan and Gurnam, 2006). To satisfy the ambitions of its expanding population in a world that is becoming more competitive, Nigeria's educational system needs to be repositioned and transformed. In order to replace test-based assessment of student performance and an education system that relied solely on tests, School-Based Assessment (SBA) was being implemented in primary and secondary schools in Nigeria. Despite all the modifications, students still have concerns about the reliability of exam results (Feron, Schils, & Ter Weel, 2016; Ugwu & Mkpuma, 2019).

According to the Federal Government of Nigeria (2014), it is important to evaluate pupils' abilities in the cognitive, emotional, and psychomotor domains. This is so that the student's potential can be developed in a balanced, holistic manner, which makes the development of these areas crucial and incredibly vital. However, the teacher's role in the SBA system is crucial

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8405151

for effective lesson planning, where the teacher-made test (TMT) plays a significant role (Ajogbeje, 2012a; 2012b; Majid, 2011). According to Mehrens & Lehmann (1987), Grisword (1990), Ajogbeje & Alonge (2012), and Omoruan (2018), classroom teachers develop the majority of the tests that are utilized in classroom evaluations. As a result, the outcomes of these tests are often used to draw conclusions or make judgments about students. For this reason, teachers in each class should make sure that the tests they administer in class are valid, relevant, suitable, and dependable.

2. LITERATURE REVIEW

Assessment

Classroom assessment includes all the different methods and tools that teachers and educators in a class measure, assess and report. This includes student academic readiness, teaching and learning progress, competence acquisition, and educational needs. Educators, policy makers, parents / guardians, employers, and the general public are all interested in knowing whether students are successfully learning in school and progressing academically. They want to know how assessments are used, how often they are administered, and whether assessments are beneficial or detrimental to students and the educational process. The definition of assessment has varied depending on the function it serves in the teaching and learning process. According to Erwin (1991), assessment serves as a methodical foundation for inferring information about the growth and learning of students. In order to improve student learning and growth, he continued, assessment is the process of identifying, choosing, planning, gathering, analyzing, and using information. According to Palomba and Banta (1999), assessment is the methodical gathering, reviewing, and utilization of data about educational programs carried out to support students' learning and growth. In order to acquire a deeper understanding of what students know, comprehend, and can accomplish with their knowledge as a result of their educational experience, Huba and Freed (2000) also employ assessment to collect information from numerous and diverse sources. When assessment results are applied to support subsequent learning, this process is complete. According to Allen (2004), evaluation entails using empirical facts about students' learning to enhance programs and boost students' academic performance.

Based on the definition above, an assessment is an opinion or judgment about someone or something, including the systematic collection of data, the examination and use of information obtained for the purposes of quality improvement, planning and decision-making. However, the assessment tools used for data collection must be reasonably accurate as measurement tools and meet clear and publicly verifiable criteria (Pratt, 1980). Performance-based assessment, according to Brualdi (1998), is a set of techniques required to apply knowledge, abilities, and work habits while completing tasks that are interesting and relevant for students. He added that performance-based evaluation would give teachers the knowledge they require to understand how kids comprehend and subsequently apply the knowledge they have acquired.

Huba and Freed (2000) examined learner-centered assessment in their study. Learner-Centered Assessment on College Campuses: Shifting Focus from Teaching to Learning 1. Creating an Intended Learning Outcome Statement, or ILO, which outlines what the student should know,

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8405151

comprehend, and be able to do with that knowledge after graduating. 2. Design or Selection of Assessment Activities, which refers to the development or choice of data collection activities to determine whether the intended learning outcomes have been attained. This comprises: i. Direct Assessment: Students are obliged to show what they know or can do with their knowledge through projects, products, works or papers, exhibitions, performances, case studies, clinical evaluations, portfolios, interviews, and oral exams. ii. Indirect Assessment: Respondents to surveys or other self-assessment tools offer their opinions about what they believe the graduate understands or is capable of doing with that knowledge. 3. Creating Experiences that Lead to Outcomes: Make sure students have experiences, both inside and outside the course, that assist them in achieving their targeted learning outcomes. 4. Discussion and Use of Assessment Results to Enhance Teaching and Learning: Using outcomes to enhance each student's performance

Teacher-Made Tests (TMT)

Teacher-made tests in Nigerian schools have a number of shortcomings. What most teachers do is to set up a quick test that is easy to grade, or a few questions to complete the assessment in record time. Many of these tests are based solely on recall of factual information and lack validity and reliability in practice. Hopkins (1990) states that this test, written by teachers, is designed to measure student performance and intended goals after they have completed a series of learning tasks in a course. After a student has completed a lesson or set of lessons based on the course content, teacher-made assessments are used to evaluate the number of educational goals that the student has met. It also involves translating well-defined subjects into test tasks to elicit information from learners about the behaviours they are expected to exhibit. To assess student performance in a particular course, teachers must ensure that test tasks adequately assess student learning material and that student learning material is tested at a level relevant to the educational objectives (Ajogbeje & Alonge, 2012). The test should be a precise measurement instrument. In other words, constructing any test should give careful consideration to what the teacher is targeting and what behaviour the teacher is considering (Anthony, 1986). The construction of a table of specifications (TOS) is one method for achieving the aforementioned objectives.

What is Table of Specifications (TOS) or Test Blueprint?

According to Notar, Zuelke, Wilson, and Yunker (2004) and Fives & DoNation-Barnes (2013), the TOS, also known as the Test Blueprint, is a chart that instructs classroom teachers on how to align goals, directions, activities, and assessments. Joshua (2005) compared TOS to blueprints that provide a guide for building objects, taking into account the relative importance of each content area and level of cognition during instruction. Onunkwo (2002) defines the TOS as a two-dimensional diagram, with rows representing the content to be examined and columns representing the educational objectives to be tested. Teachers should make sure when constructing a test that it measures a sufficient sample of the course material at the cognitive level at which the subject was delivered. According to Fives and DiDonato-Barnes (2013), the main objective of TOS is to improve the reliability of teacher evaluations in relation to certain assessments. They added that the TOS links to unit plans and lessons, assisting teachers in

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8405151

drawing clear linkages between the plans, lessons, activities, and assessments. Therefore, as a blueprint or guide for test construction, teachers in the classroom should create a two-dimensional table that summarizes course content in one dimension and instructional or behavioral goals in another (Gronlund, 2006; Hakim & Irhamsyah, 2020; Karim, Sudiro, & Sakinah, 2021; Amelia, Sari, & Astuti, 2021).

The Use of Test Blueprint in Test Construction

Students often complain about teacher test imbalances if the examination emphasizes certain aspects of the topic or pays close attention to detail. Exams either focus on educational topics that have not been addressed or a disproportionate number of items are drawn from areas of study that have gotten little consideration. This is because we do not use TOS. TOS is a test blueprint which helps teachers to align learning objectives, instruction, content, learning activities and test items in assessments systematically (Eweda, Bukhary and Hamed, 2020). However, although TOS cannot provide perfectly fair weighting, it can improve the effectiveness of teacher-made test content (Denga, 2003). TOS can be used in conjunction with lesson and unit planning to help teachers make a clear connection between planning, instruction, activity and assessment.

The primary goal of a TOS is to improve the validity of a teacher's evaluation in relation to a given assessment. The casual attitude of classroom teachers to the regular or continuous use of TOS in teacher preparation for academic achievement tests has led to misunderstandings about the assessment of student achievement in our schools. True proficiency could not be determined by such tests without TOS. Gullickson (1986) and Marso & Pigge (1989) suggested that this may not be unrelated to classroom teacher education and training. Most of our teachers are trained in how to teach students, not how to judge them. However, as they are not anticipated to achieve these skills at this point in their academic growth, novices may be excluded from higher categories under the TOS. The domains connected with "retrieving or perceiving knowledge and developing intellectual abilities and skills" (Bloom, 1956) are referred to as the cognitive domains. 1. Indicate the percentage of items for each unit or topic. 2. Calculate the proportion of items that must be prepared for each cognitive domain of the lesson goals. 3. Find out how many items you need to prepare for each level of the lesson's objectives. 4. Find out how many items are prepared for each content topic or unit at various cognitive levels. 5. If any changes are required, make them. According to Mulyani, Tanuatmodjo, and Iskandar (2020), teachers should use the TOS since it outlines both the subject topics covered in class and the performance goals for each level of Bloom's Taxonomy's cognitive domain.

Using Table of Specifications (TOS) to Support Validity

Teaching objectives, cognitive levels of instruction, and the number of test items required to measure each objective are related in a table of specifications (TOS), which also contains interactive charts to assist teachers (Nortar et al., 2004). The TOS provides a structure for compiling data on student experiences with instructional activities. Note that teachers should choose (1) how many test items to include and (2) how to distribute multiple-choice and short-answer questions before drafting the TOS. According to Nortar et al. (2004), TOS gives

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8405151

teachers the freedom to decide what to teach and how much time to devote to each objective. The teacher will professionally choose the amount of items for a given exam based on the unit's objectives, the students' comprehension, the time available for instruction, and the significance of evaluation. Shorter tests may still be reliable if they include enough data to allow teachers to make judgments based on student performance.

A crucial factor for TMT is the validity of an instrument, which has to do with what it measures and how well it assesses it. In order for a test to be considered valid, it must be able to measure everything that it claims to be able to assess (Alonge, 1989; Bandele, 2004; Grimm and Yarnold, 2006; Kolawole, 2010). Bandele (2004) added that the validity of an instrument is typically assessed in connection with the particular purpose for which it is being examined. There are four primary categories of validity. Face validity, content validity, criterion-related validity, and construct validity (Alonge, 1989; Bandele, 2004). The degree to which a test or instrument is evaluated to ascertain what it only visually assesses or is intended to measure is referred to as face validity. Content validity is about test content and test items, and according to Alonge (1989), content in this context refers to the material's substantive, factual, and informational components.

When there is an interest in determining the relationship between results and particular criterion tests, criteria-related validity is evaluated. The validity of an instrument is determined by how closely its measure relates to an external criterion, which is meant to have characteristics similar to those of the instrument whose validity is being assessed (Bandele, 2004). Predictive validity and concurrent validity are the two types of criterion-related validity. A test's capacity to forecast future outcomes is known as its predictive validity. A correlation coefficient between the distribution of test scores acquired at earlier time points and the distribution of scores on the later reference scale is typically calculated to determine this (Alonge, 1989). When an external criterion is obtained concurrently with an instrument result, it is referred to as concurrent validity. The coefficient of concurrent validity is determined by correlating the instrument result with the criterion result. The degree to which a test measures the characteristics of a theoretical construct is known as construct validity. To determine whether the construct is valid, questions from the instrument under construction and questions from similar existing instruments that have been proven to be reliable can be correlated. When there is a strong correlation between the scores, convergence validity is discussed. Convergence validity provides support for construct validity.

The validity that is used in this study, specifically with reference to TMT, is content validity. Most experts agree that the most important type of validity required for criterion-referenced measures is content validity (Setiabudi, Mulyadi, and Puspita, 2019; Akinboboye and Ayanwale, 2021). The range that the instrument truly measures or is related to the characteristics it is intended to measure is referred to as content validity (Bandele, 2004). How closely a test item—or an item within a test—represents the subject matter that the exam is designed to evaluate (Kolawole, 2001). It involves topic coverage and relevance (Setiabudi, Mulyadi, and Puspita, 2019; zdemir, zdemir, and Gelbal, 2021) and assesses how accurately the scale captures the range of knowledge and abilities that respondents must possess.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8405151

Empirical Studies on Content Validity of Teacher Made Achievement Test

There are several questions about test validity in the educational sector (Aliah, 2020; Akinboboye & Ayanwale, 2021), particularly the content validity of teacher-made achievement tests constructed by teachers. According to Alvares (2013) and Idris and Alfa (2012) in their respective studies, content validity is all about how topics and subtopics are represented in a test instrument for the purposes of formative and summative evaluation. They discovered that a substantial percentage of respondents in the samples they looked at never made reference to or used the TOS when creating evaluation tools. Classroom teachers who don't adhere to conventional test construction standards are bound to fail or fall short of accurately and fairly measuring their students' academic progress, according to Notar et al. (2004) and Sugianto (2020).

Many teachers are confused about how to use assessments since they are unable to build TOS, according to researchers Ing, Musah, Al-Hudawi, Tahir, and Kamil (2015), Idris et al. (2016), and Odiagbe (2016). Researchers Musah, M. B., Tahir, L. M., Al-Hudawi, S. H., Issah, M., Hussein, A. R., & Ibrahim, M. (2022) found that teachers at the selected schools had a clear understanding of how to connect course content with the evaluation tool using process- and evidence-based methodologies. As a result, the TMT's validity was upheld by the majority of their review processes. It is crucial to remember that the validity of TMT was confirmed by the alignment of instructors' assessments with TOS, which took both procedural and evidence-based approaches into account. This demonstrates that when creating instruments for formative learning performance evaluation, the majority of teachers employ the TOS. The results also show that teachers stick to the standards, procedures, and strategies they learned during teacher professional development rather than considering their students' talents when evaluating their performance.

The findings of their study also demonstrated that among the schools under investigation, none of the demographic parameters investigated (job experience, age, or gender) significantly affected TMT practices. As a result, the content validity of TMTs created by both the most experienced teachers and those with less expertise is the same. The gender of the teachers also has little effect on how they create and apply TMT to evaluate the academic success of their pupils. The teachers' ages likewise have little significant impact on the outcomes. As a result, teachers with and without expertise approach TMT exercises in a comparable way. This can be due to the fact that all teachers go through the same professional teacher preparation program. Based on a literature analysis of the variables that have been discovered, this study investigates the content validity of TMT, teachers' understanding of the TOS, and teachers' knowledge of the TOS. The following research questions were put out to guide the investigation:

Research Questions

1. To what extent are teachers aware of the usage of TOS in test construction?
2. To what extent do teacher-made assessment test meet TOS content validity?
3. How extensively do the teachers employ the TOS in test construction?

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8405151

3. METHODOLOGY

This study employed a descriptive survey research design, and the sample included 250 classroom teachers drawn at random from 10 secondary schools in the Akure South Local Government Area of Ondo State. Regardless of their employment experience, gender, or age, respondents were chosen. To gather all the information necessary for the study, the researcher constructed and verified a questionnaire called the Content Validity Survey (CVS). The instrument is divided into two sections: A and B. Section A asks about personal and academic information, while Section B has 18 questions on teachers awareness of the usage of TOS in test construction, extent to which TMT meet TOS content validity, and teachers usage TOS in test construction. The 18-item survey instrument was constructed using a four-point Likert scale, with "strongly agree" (SA) = 4, "agree" (A) = 3, "disagree" (D) = 2, and "strongly disagree" (SD) = 1.

Two senior lecturers in measurement and evaluation subjected the research instrument's validity to face and content validity checks. Two senior vice principals and two senior principals from Akure, Ondo State's secondary schools, were chosen for their assessments on the CVS document. The chosen expert was asked to remark on whether the questions accurately and completely covered the essential components of the TOS and whether they were simple enough for teachers to understand. Otherwise, they must recognize confusing phrases or words and change or remove them. It was determined that the reliability coefficient (Cronbach's alpha = 0.85) was high enough for a research study. The researcher and two more research assistants administered the research instrument to respondents from various schools. The frequency count, percentage, mean, and standard deviation were used to analyze the data. The reference cutoff was a weighted mean of 2.50 (4 + 3 + 2 + 1 = 10/4). Any mean below 2.5 is interpreted as disagreeing with the item, while any mean above 2.5 is interpreted as agreeing with the item.

4. RESULTS AND DISCUSSION

Part I: Descriptive Analysis

Table 1 details the frequency tabulation of the 250 respondents selected for the survey. There are 178 female teachers and 72 male teachers among them (28.8% male and 71.2 % female). According to the respondents' age distribution, 34.8% of them, or 87 respondents, were between the ages of 41 and 50. This is followed by the 31-40 year old group with 72 (28.8%). Respondents aged 51 and over accounted for 56 (22.4%), while those between the ages of 21 and 30 accounted for only 35 (14%). Table 1 also reveals that 13 respondents (5.2%) have taught for one to five years. Twenty (8%) responders with a teaching background of 6 to 10 years came next. Thirty-one of the respondents (12.4%) said they had taught from 11 to 15 years, and 55 or 22% of respondents said they had taught from age 16 to 20 years. 71 (28.4%) and 41 (16.4%) respondents had teaching experience of 21-25 years and 26-30 years, respectively. Finally, the researcher found that 19 of the respondents (7.6%) spent 31 to 35 years. Please note that civil servants of the Ondo State Government (ODSG) are expected to retire after the age of 60 or 35 years of service.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8405151

Table 1: Gender, age, and employment experience of respondents, in frequency and percentages

	Variable	*Frequency*	*Percentage (%)*
Gender	Male	72	30
	Female	178	70
Total		250	100
Age	21 - 30	35	14
	31 - 40	72	28.8
	41 - 50	87	34.8
	51 - 60	56	22.4
Total		250	100
Teaching Experience	1 – 5	13	5.2
	6 – 10	20	8.0
	11 – 15	31	12.4
	16 – 20	55	22.0
	21 – 25	71	28.4
	26 – 30	41	16.4
	31 – 35	19	7.6
Total		250	100

Part II: Teachers' Understanding of TOS

Item 1 had a mean of 1.95, which was lower than the average value of 2.5 and the standard deviation of 0.89, as indicated in Table 2's results for the individual item means. The results of the individual item measures suggest that the majority of respondents did not agree that the TOS should be incorporated into the test design. Item 2's mean value was 2.72, which was greater than the 2.5 average and 1.27 standard deviation. This shows that the majority of respondents concurred that their college professor taught them TOS throughout their preservice teacher preparation.

Table 2: Teachers' responses to teachers' understanding of TOS

S/N	*Statements*	*SA*	*A*	*D*	*SD*	*N*	*Mean*	*Std. Dev*
1.	I am conscious of the need to use TOS when writing test items.	53	77	65	55	250	1.95	0.j89
2.	I was taught TOS by my college lecturer during my preservice training as a teacher	59	84	57	50	250	2.72	1.27
3.	I am capable of constructing a TOS for the subject I teach.	98	107	23	22	250	1.26	0.70
4.	I discuss with my colleagues while constructing assessment instruments	8	13	121	108	250	1.35	0.95
5.	I attend in-service training, and seminars on how to construct assessment instruments	67	143	25	15	250	1.98	1.01

Item 3: - I am capable of constructing a TOS for the subject I teach with (mean = 1.26; SD = 0.70); Item 4: - I discuss with my colleagues while constructing assessment instruments with (mean = 1.35; SD = 0.95); and Item 5: - I attend in-service training, seminars, and conferences that are related to the construction of assessment instruments with (mean = 1.98; SD = 1.01). All got poor average scores that were below the mean (2.50). This research demonstrates that

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8405151

the respondents were in agreement that they could not develop a TOS for the courses they teach. They do not engage in collegial discussion about how assessment tools are made, nor do they go to conferences, seminars, or training sessions on the subject.

The results of the study showed that there was very little understanding among teachers since the majority of respondents were unaware that the TOS should be utilized to create tests. Most of the respondents did not utilize TOS when creating tests because, in their opinion, the use of a TOS is not relevant in the design of tests, according to the teachers, who also disagree. As a result, the majority of respondents were unable to create a TOS because they lacked the necessary information. The findings of this study are supported by the research findings of Ing, Musah, Al-Hudawi, Tahir, and Kamil (2015), Idris et al. (2016), and Odiagbe (2016), which all demonstrated that many teachers are uncertain about how to administer assessments as a result of their failure to design TOS.

Part III: Content Validity of Teacher-Made Tests

The results of item 6 in Table 3 showed that, with a mean of 2.00, which is below the acceptable criterion of 2.5, and a standard deviation of 0.97, the majority of respondents disapproved of using TOS to prepare the classroom assessments. The majority of respondents said they knew the subject area they teach extremely well, as evidenced by Item 7's mean of 3.52, which is greater than the permitted value of 2.5 and had a standard deviation of 0.87. With an average response of 3.44 and a standard deviation of 1.48 for item 8, it can be seen that respondents largely concur that they are aware of the format of assessments in the field.

Table 3: Teachers' responses to content validity of TMT

S/N	*Statements*	*SA*	*A*	*D*	*SD*	*N*	*Mean*	*Std. Dev*
6	I make use of TOS for the classroom test construction	23	12	70	135	250	2.00	0.97
7	I am knowledgeable about the subject matter I teach.	100	123	12	15	250	3.52	0.87
8	Regarding the evaluation format for the subject I teach, I am clear.	70	67	90	23	250	3.44	1.48
9	I am competent at classifying test items according to how difficult they are.	95	78	37	40	250	2.76	1.46
10	I am capable of categorizing test items according to difficulty level and students' understanding.	26	47	65	112	250	3.03	1.62
11	Table of specification ensures all topics are sampled.	91	108	27	25	250	2.62	1.34
12	I am unable to create a table of specifications.	42	21	92	95	250	2.76	1.05

Some of the respondents agreed that they can categorize assessment items according to how difficult they are, that they are aware of the degree to which their pupils comprehend the subject matter they teach, and that TOS guarantees that all topics are covered. Similar results can be seen for item 9, which has a mean of 2.76 and a standard deviation of 1.46; item 10, which has a mean of 3.03 and a standard deviation of 1.62; item 11, which has a mean of 2.62 and a standard deviation of 1.34; and item 12, which has a mean of 2.76 and a standard deviation of

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8405151

1.05. Findings show that the teachers were unfamiliar with how TOS for test construction were constructed. Teachers may not be using TOS because they do not know how to create them, which is one possible explanation. The teachers agreed that TOS is capable of achieving this, so TOS will make sure that every topic is represented and sampled in TMT. The results are consistent with research by Alvares (2013) and Idris & Alfa (2012), who found that content validity mainly focuses on how subject topics and subtopics are represented or included in a test instrument for the purposes of formative and summative assessment.

Findings also reveal a significant majority of respondents never use or refer to the TOS when creating assessment tools. According to Notar et al. (2004) and Sugianto, (2020), classroom teachers who do not likely follow the rules for creating standard tests are likely to under- or overestimate student achievement and produce assessments with weak or extremely low content validity. Given a specific level of respondents, the TMT at the schools chosen for this study only represented the lessons taught in the classroom by the teachers; the content was never in line with the rules guiding the construction of TOS. The TMT cannot be said to have examined the objectives that it would have been expected to evaluate in the absence of the TOS.

Part IV: Teachers' Utilization of TOS in Teacher-Made Assessment

Item 13's mean value of 3.41 is greater than the acceptable value of 2.5 and its standard deviation is 1.10, as shown in Table 4, the majority of respondents concur that test items should not be constructed using TOS. The results for question 14 show that the respondents unambiguously concur that they cannot create tests using a TOS, with a mean of 3.10 and a standard deviation of 1.00. Similar to item 14, item 15 has a mean of 2.82 and a standard deviation of 1.03. For item 16, the average and standard deviation are 2.99 and 0.85, but for item 17, they are 2.72 and 0.79.

Table 4: The extent of teachers' utilization of TOS in teacher-made assessment

S/N	*Statements*	*SA*	*A*	*D*	*SD*	*N*	*Mean*	*Std. Dev*
13	I don't use a TOS when creating test items	125	45	35	45	250	3.14	1.10
14	I have no idea how to create a test item using TOS.	23	45	67	115	250	3.10	1.00
15	I have never observed any other teachers or colleagues utilizing TOS.	72	92	41	45	250	2.82	1.02
16	TOS ensure content validity	70	90	68	22	250	2.99	1.25
17	The development of TOS for item generation takes time.	50	110	72	18	250	2.72	1.09

The respondents concurred that creating a table of specifications for item generation takes a lot of effort and that they have never seen a fellow teacher use one. The findings demonstrate that neither the teachers at the selected secondary schools nor the exam items they prepare employ tables of specifications. They do not know how to create tests using test blueprints or specification tables. The respondents also agreed that creating a test plan for item production requires a lot of time and effort. The results of this study are in line with those of Notar et al. (2004) and Sugianto (2020), who shared the belief that teachers who do not follow traditional

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8405151

building criteria for test construction are not accurately assessing student achievement. It is alarming because their tests will likely have low content validity because this is the component that each assessment tool's validity depends on the most.

5. CONCLUSION

The results of the study led to the conclusion that teachers in the majority of the sampled schools have little or no understanding of the TOS as a result of not participating in workshops, seminars and training sessions connected to the TOS (Hartell & Strimel, 2019). They were unable to develop test items utilizing TOS for the courses they taught, mostly because of this. On the other hand, it was shown that TMT had average content validity. A table of specifications is undoubtedly a crucial step in ensuring a reliable instrument. It aids in making sure that what is taught and what is tested match. The majority of teachers did not use the TOS while creating test questions for evaluation. When teachers develop assessment instruments, they do not take into consideration the TOS but rather student skills and responses, as seen by the average content validity attained while they did not do so. The low or poor content validity obtained in TMT in the selected schools was due to teachers' lack of knowledge of the standard TOS's design and their insufficient awareness of its importance in the sample schools. TMT is more than just a tool for evaluation. TMT can help instructors obtain perspective on their lessons, help students decide what to study, and establish the objectives of education.

A few suggestions were offered to increase the TMT's validity. The majority of instructors in the sampled schools lack the necessary skills in TOS formulation and design, which leads to poor-quality TMT and compromises the content validity of the TMT (Michael & Burlingame, 1996). Therefore, it is important to educate and familiarize teachers, administrators, Ministry of Education representatives, and other educational stakeholders with the theory and concepts of TOS. They should also learn how to create tables of specifications and use them when creating assessment test items. This can be accomplished through specialized training initiatives, including seminars, and workshops. In order to gather high-quality data, maintain fairness in the scoring, and eliminate scoring errors so that decisions about student performance can be made with confidence, classroom teachers should promote the use of a variety of scoring procedures and assessment methodologies (McMillan, 2000). Although creating a TOS may take a lot of time and work, it is essential for giving the most accurate and useful information about the TOS and for effectively facilitating readiness testing.

References

1) Ajogbeje O. J. (2012a). *Path–analytic model and the effect of some teaching strategies on variables affecting achievement in junior secondary school mathematics in ondo state.* Unpublished Ph. D. Thesis, Ekiti State University, Ado–Ekiti, Ekiti State

2) Ajogbeje, O. J. (2012b). Effect of formative testing on students' achievement in junior secondary school mathematics. *European Scientific Journal. 8(8),* 94-105.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8405151

3) Ajogbeje, O. J. & Alonge, M. F. (2012). Effect of feedback and remediation on students' achievement in junior secondary school mathematics. *International Education Studies. 5(5)*, 153-162.

4) Akinboboye, J. T. & Ayanwale, M. A. (2021). Bloom Taxonomy Usage and Psychometric Analysis of Classroom Teacher Made Test, African Multidisciplinary Journal of Development (AMJD), 10(1), 2021

5) Aliah, H. (2020). The Analysis of Junior High School Teacher-Made Tests for the Students in Enrekang, FOSTER: Journal of English Language Teaching and Learning, 1(2), 122–138. https://doi:10.24256/foster-jelt.v1i2.14

6) Allen, M. J. (2004). Assessing academic programs in higher education. San Francisco, CA, Jossey-Bass Inc.

7) Alonge, M. F. (1989). Measurement and Evaluation in Education and Psychology. Ado– Ekiti, Nigeria: Adedayo Printing (Nigeria) Limited.

8) Alvares, K. K. (2013). Table of Specification and Test Construction. https://www.katekimberlyalvares/table of specification

9) Amelia, R. N., Sari, A. R. P., & Astuti, S. R. D. (2021). Assessment of Chemistry Learning: How is The Quality of The Tests Made by The Teacher? Journal of Educational Chemistry (JEC), 3(1), 11–22. https://doi:10.21580/jec.2021.3.1.6582

10) Anthony, A. (ed.) (1986). Measurement and Evaluation. Institute of Education University of Nigeria, Nsukka. Associateship certificate in Education series 30-31.

11) Bandele, S. O. (2004). Educational Research in Perspectives (1st ed.). Ibadan, Nigeria: Niyi Comm. & Printing Ventures

12) Bloom, B. S. (ed.) (1956). Taxonomy of educational objectives, Handbook 1: The cognitive domain. New York: Mckay.

13) Brualdi, A. (1998). Implementing Performance Assessment in the Classroom. Journal of Practical Assessment, Research & Evaluation, 6(2).

14) Chan, Y. F. & Gurnam, K. S. (2006). Investigating learning challenges faced by students of higher education. Procedia-Social and Behavorial Sciences.186 (2015), 604-612.

15) Denga, D.I. (2003). Educational measurement continuous assessment and psychological testing. (3rd ed.). Calabar: Rapid Educational Publishers Ltd.

16) Erwin, T. D. (1991). Assessing student learning and development. San Francisco, CA, Jossey-Bass Inc.

17) Eweda, G., Bukhary, Z. A. & Hamed, O. (2020). Quality assurance of test Blueprinting Enhanced Reader, Journal of Professional Nursing, 36, 166–170. https://doi:10.1016/j.profnurs.2019.09.001

18) Federal Republic of Nigeria (2014). National Policy on Education, Abuja, NERDC Feron, E., Schils, T. & Ter Weel, B. (2016). Does the Teacher Beat the Test? The Value of the Teacher's Assessment in Predicting Student Ability, Economist (Netherlands), 164(4), 391–418. https://doi:10.1007/s10645-016-9278-z

19) Fives, H., & DiDonato-Barnes, N. (2013). Classroom Test Construction: The Power of a Table of Specifications. Practical Assessment, Research & Evaluation, 18(3), 1-7. https://pareonline.net/pdf/v18n3.pdf

20) Grimm, L. G., & Yarnold, P. R. (2006). Reading and Understanding Multivariate Statistic. Washington, DC. American Psychological Association.

21) Griswold, P. A. (1990). Assessing Relevance and Reliability to improve the Quality of Teacher-Made Tests. Journal of NASSP Bulletin, 74, 18-24. https://dx.doi.org/10.1177/019263659007452305

22) Gronlund, N. E. (2006). Assessment of students' achievement (8th ed.). Boston, MA: Pearson.

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8405151

23) Gullickson, A.R. (1986). Teacher education and teacher perceived needs in educational measurement. Journal of Educational Measurement, 247-354.

24) Hakim, L. & Irhamsyah, (2020). The analysis of the teacher-made test for senior high school at state senior high school 1 Kutacane, Aceh Tongogara, Journal Ilmiah DIDAKTIKA, 21(1), 10–20.

25) Hartell, E. & Strimel, G. J. (2019). What is it called and how does it work: examining content validity and item design of teacher-made tests, International Journal of Technology and Design Education, 29(4), 781–802. https://doi:10.1007/s10798-018- 9463-2

26) Hopkins, K.D. (1990). Educational and psychological measurement (7th ed.). Englewood cliffs, NJ: Prentice Hall.

27) Huba, M. E. & Freed, J. E. (2000). *Learner-centered assessment on college Campuses Shifting focus from teaching to learning.* Boston: Allyn & Bacon.

28) Idris, U.S.B; Ugochukwu, E.M; Olalere, A.M and Shehu, H. (2016). Procedures for Improving Content Validity of Teacher-Made- Test: Way forward to Ensuring Quality Research and Evaluation. Paper Presented at the 18th Annual National Conference by ASSEREN held in Owerri, (11 – 15 July, 2016).

29) Ing, L. M.; Musah, M. B.; Al-Hudawi, S. H. V.; Tahir, L. M. & Kamil, N. M. (2015). Validity of teacher-made assessment: A table of specification approach Asian Social Science, 11(5), 193-200.

30) Joshua, M. T. (2005). Fundamental test and measurement in education. Calabar: The University of Calabar Press.

31) Karim, S. A., Sudiro, S. & Sakinah, S. (2021). Utilizing test items analysis to examine the level of difficulty and discriminating power in a teacher-made test. Journal of English Education, Literature, and Culture. 6(2), 256–269.

32) Kolawole, E. B. (2001). Test & Measurement (Revised Edition). Lagos, Nigeria: Bolabay Publications.

33) Kolawole, E. B. (2010). Principles of Test Construction and Administration (Revised Edition). Lagos, Nigeria: Bolabay Publications.

34) Majid, F. (2011). School-Based Assessment in Malaysian Schools: The Concerns of the English. Journal of US-China Education Review, 8(10), 1-15. https://education.uitm.edu.my/v1/images/stories/publication/faizah/article7.pdf

35) Marso, R. N. & Pigge, F. L. (1987 April). The states of classroom teachers test construction proficiencies: assessment by teachers, principals and supervisors validated by analysis of actual teachers made test. Paper presented at The Annual Meeting of the National Council on Measurement in Education: San Francisco.

36) McMillan, J. H. (2000). Fundamental Assessment Principles for Teachers and School Administrators. Journal of PARE online, 7(8). https://pareonline.net/getvn.asp?v=7&n=8

37) Mehrens, W. A., & Lehmann, I. J. (1987). Using Teacher-Made Measurement Devices. Journal of NASSP Bulletin https://dx.doi.org/10.1177/019263658707149605

38) Micheal, J. M., & Burlingame, G. M. (1996). The Reliability and Validity of the Outcome Questionnaire. Clinical Psychology and Psychotherapy, 3(4), 249-258. https://dx.doi.org/10.1002/(SICI)1099-

39) Mulyani, H., Tanuatmodjo, H. & Iskandar, R. (2020). Quality analysis of teacher- made tests in financial accounting subject at vocational high schools," Journal Pendidikan Vokasi, 10(1), 1–9. https://doi:10.21831/jpv.v10i1.29382

40) Musah, M. B., Tahir, L. M., Al-Hudawi, S. H., Issah, M., Hussein, A. R., & Ibrahim, M. (2022). Testing content validity of teacher-made test: Profiling teacher perceptions and demographic variables. International Journal of Evaluation and Research in Education, 11(2), 878-887. https://doi.org/10.11591/ijere.v11i2.21992

The Seybold Report
ISSN 1533-9211
DOI: 10.5281/zenodo.8405151

41) Notar, C. E., Zuelke, D. C., Wilson, J. D., & Yunker, B. D. (2004). Table of Specifications: Insuring Accountability in Teacher-Made Tests. Journal of Psychology, 31(2). https://www.freepatentsonline.com/article/Journal-Instructional-psychology/119611686.html

42) Odiagbe, S.I. (2016). Determining the Items Difficulty and discrimination index of Mathematics Achievement Tests from Junior Secondary Schools in Kwali Area Council. Paper Presented at the 18th Annual National Conference by ASSEREN held in Owerri, (11–15 July, 2016).

43) Omoruan, B. E. (2018). Valid Teacher-Made Tests: It's Implication on Students' Achievement in Business Education, ATBU, Journal of Science, Technology & Education (JOSTE), 6(2), 65–77.

44) Onunkwo, G.I.N. (2002). Fundamental of Educational Measurement. Owerri: Capes Publishers.

45) Özdemir, G., Özdemir, A. & Gelbal, P. D. S. (2021). Determination of cyber accessibility of teacher made tests/exams, International Journal of Assessment Tools in Education, 8(3), 553–569. https://doi:10.21449/ijate.780556

46) Paloma, C. & Banta, T. W. (1999). Assessment Essentials Planning, Implementing, and Improving Assessment in Higher Education. San Francisco, CA, Jossey- Bass Inc.

47) Pratt, D. (1980). Curriculum: Design and development. New York: Harcourt Brace Jovanovich.

48) Setiabudi, A., Mulyadi, & Puspita, H. (2019). An Analysis of Validity and Reliability of a Teacher-Made Test (Case Study at XI Grade of SMA N 6 Bengkulu), Journal of English Education and Teaching (JEET), 3(4), 522–532.

49) Ugwu, N. G. & Mkpuma, S. O. (2019). Ensuring Quality in Education: Validity of Teacher-made Language Tests in Secondary Schools in Ebonyi State, American Journal of Educational Research, 7(7), 518–523. https://doi:10.12691/education7-7-12

50) Sugianto, A. (2020). Item Analysis of English Summative Tests: EFL Teacher-Made Test, Indonesian EFL Research & Practice, 1(1), 35–54.