**International Journal of Advanced Education and Research**
**www.alleducationjournal.com**
**ISSN: 2455-5746**
**Received: 11-11-2021, Accepted: 26-11-2021, Published: 13-12-2021**
**Volume 6, Issue 6, 2021, Page No. 15-19**

# Analysis of content validity of teacher-made tests for physics in Sapele local government area, Delta state Nigeria

**Romy O Okoye, Benedicta I Fejokwu, Benedict C Ikeanumba**
Department of Educational Foundations, Faculty of Education, Nnamdi Azikiwe University, Awka, Anambra, Nigeria

## Abstract
The new secondary school curriculum in Nigeria has indicated that final assessment and examination at national level should incorporate 40% continuous assessment (CA) from Teacher Made Tests (TMTs). However, little has been researched on the validity of teacher made tests. This study, therefore, assessed comparability of content validity of Teacher Made Tests for Physics in secondary schools in terms of item-syllabus alignment, content representativeness and cognitive functioning levels. To generate data, the study used 30 TMTs for physics for Terms 1, 2 and 3, for the academic year 2019/2020 for SS 2 students from Sapele secondary schools in Sapele LGA of Delta state of Nigeria. The study adopted descriptive survey research design. The study used both qualitative and quantitative methodologies. Subject Matter Experts (SMEs) were used to generate the qualitative data by analysing each test item to a content area and cognitive level, while the quantitative data were analysed using descriptive statistics (Mean). The study found among others that the test items from all the schools were developed from the syllabus. However, content coverage and item cognitive functioning levels varied from one school to another. The study, therefore, recommends that secondary school teachers in Sapele should be trained on the development of standard tests to ensure that the secondary school TMTs are comparable in terms of content representativeness and cognitive functioning levels.

**Keywords:** teacher made tests, content validity, content representation, item cognitive functioning level

## Introduction
The Nigerian secondary curriculum underwent a revision in 2011. The policy for the assessment of contents of the revised curriculum indicates that national examinations results should incorporate 40% continuous assessment (CA). This has prompted the National Examinations Council (NECO) as an examining body in Nigeria to change its mode of assessment of secondary education to include CA. Previously, NECO's mode of assessment of students' work was mainly based on final examination.
CA refers to an on-going process which takes place throughout the whole learning process (Le Grange & Reddy, 2018). Continuous assessment is a form of educational examination that evaluates a student's progress throughout a prescribed course. It is often used as an alternative to the final examination system. The purpose of CA is to assess trichotomy of learning of a learner's development: cognitive, affective and psychomotor domains. It can be formal or informal. Formal assessment deals primarily with learner's cognitive domain. Written test is the best method that can be used for assessing cognitive domain (Nxumalo, 2017) [21]. In the case of CA, it is the class teacher who prepares the test items (Marso & Pigge, 2018) [15].
As part of curriculum reform, Nigerian government has introduced a national policy on assessment of secondary education. The policy requires candidates that are assessed at SSCE level to have two components of assessment score: 60% for final examination and 40% for CA. The actual assessment instruments of CA will mainly include TMTs. TMTs need to be valid to gauge students' mastery of learning at the end of a course, term or semester (Gronlund, 2018; Nitko, 2016) [19]. In order to achieve this, TMTs must meet the key attributes of test quality such as validity and reliability (Kinyua & Okunya, 2014) [13]. Validity tells you how accurately a method measures something. If a method measures what it claims to measure, and the results closely correspond to real-world values, then it can be considered valid. There are four main types of validity: construct validity, content validity, face validity and criterion validity. Content validity assesses whether a test is representative of all aspects of the construct (Fiona, 2020). Content validity is defined as the degree to which the items in the test adequately sample the areas of subject and abilities that a course of instruction has aimed to teach (Ebel, 2019). Content validity of a test addresses the match between test items and the content area they are intended to assess. This concept of match is sometimes referred to as alignment, while the content area of the test may be referred to as performance domain.
Subject Matter Experts (SMEs) in a given performance domain generally judge the content validity (Sireci & Faulkner-Bond, 2014) [26]. Thus, content validity is how well the behaviours demonstrated in a testing situation constituted a representative sample of behaviours to be exhibited in the domain (content). According to Bachman (2010) [2], content validity involves two major concepts: content relevance and content representation. However, Sireci (2018) extended the definition of content validity to include aspects of test development. He argues that content validity pertains to four elements of test quality: domain definition, domain representation, domain relevance and appropriateness of test development. When these elements are applied to teacher-made tests, domain definition simply involves the areas where test items are going to be drawn from. Test items should be drawn from the syllabus (Gronlund, 2003). Domain representation implies that

teachers must ensure that test items represent the content areas of the syllabus in proportion to the demands of each content area.

Recent studies on the validity and reliability of teacher-made tests show that a student's grade point average is usually not consistent with the same student's scores on standardized tests (Kinyua & Okunya, 2014) [13]. Although many researchers claim that teacher-made tests are flawed (Oescher & Kirby, 2010) and inadequate (Wiggins, 2019) due to poor training, there is limited empirical evidence on how content validity of TMTs compare in terms of item-syllabus alignment, content representativeness and cognitive functioning level.

**Item cognitive functioning level**

Benjamin Bloom introduced a taxonomy of educational objectives which consist of cognitive, affective and psychomotor domains. He further divided the cognitive domain into six levels viz: knowledge, comprehension, application, analysis, synthesis and evaluation, with knowledge being the lowest level and evaluation the highest. According to Okoye (2015) [23], some individuals expressed worry about some aspects of Bloom's levels of the cognitive domain, and for that reason, one of his students Loren Anderson, revised Bloom's levels of the cognitive domain to read as follows: Remembering, Understanding, Applying, Analyzing, Evaluating and Creating. Irrespective of which of the two models a teacher wishes to use, they should, while setting questions, ensure that such questions cover the various levels of the cognitive domain. Oescher and Kirby (2010) [22] conducted a research study in secondary Physics in which the findings revealed that 78% of the items were at knowledge level and 13% were at comprehension level. Test items that predominantly assess learners' ability to remember encourage rote learning. According to Baroudi (2017), good test items require more than remembering a fact or reproducing a skill. Since there is a thin gap between higher adjacent levels, the cognitive skills in this study have been put into two categories: lower order and higher order. Lower order cognitive skills consist of remembering and understanding while higher order skills consist of applying, analyzing, evaluating and creating.

**Guidelines for Test Development**

In order to enhance validity and reliability of teacher-made tests, teachers must stick to standard guidelines for test development. Teachers must understand that developing a test is an art. Before engaging in the actual test development, teachers must firstly, consider the purpose of the test. This is because an assessment may be appropriate for one purpose but inappropriate for another purpose (Young, So & Ockey, 2013) [28]. If the test is not appropriate for the purpose for which it was designed, then it is not a valid test. The teacher must develop a test plan that describe the instructional objectives, the content to be measured and the relative emphasis given to each learning outcome (Gronlund, 2011) [9]. This is what Sireci (2018) [25] described as the domain definition, domain representation and domain (content) relevance. In other words, the teacher must consider the content from where the test items are going to be developed, how the items are going to represent the content areas or core elements of the syllabus as well as the levels of cognitive skills the items should operate on. Normally, the syllabus or curriculum dictates the scope and sequence of the content area from which to develop test items. Teachers must be able to develop a table of test specifications (TOS). A TOS is defined as a test blueprint which helps teachers to align objectives, instruction, activity and assessment (Fives & DiDonato-Barnes, 2013) [8]. It is a two-way grid that captures content domain as well as cognitive levels of the Bloom's taxonomy.

Table of test specifications is important in test development. Notar (2014) [20] and Parr and Bauer (2016) [24], in separate research studies found that developing a table of specifications ensures the content validity of a test. Table of specifications helps the test developer to ensure appropriate level of difficulty of items, varied types of items and wide content coverage of the syllabus (Parr & Bauer, 2016) [24]. In addition, a study by Fives and DiDonato-Barnes (2013) [8] revealed that when teachers do not use TOS, there is often a mismatch between the content that is examined in class and the material assessed at the end of unit test. A research study conducted by Ing, et al (2015) found that TMTs were valid in terms of content validity although most of the teachers did not even refer to TOS when developing the test items. It is assumed that if teachers apply most of the guidelines above, they are likely to come up with valid and comparable tests. According to Chakwera (2014) [5], teachers who were knowledgeable in acceptable assessment skills prepared their candidates for national examinations better than the less knowledgeable ones. However, little has been researched on the validity of teacher made tests. This study therefore, assessed content validity of TMTs for physics in secondary schools in terms of item-syllabus alignment, content representativeness and cognitive functioning level of the items.

**Research Questions**

The following research questions guided the study

1. To what extent do TMTs for Physics align with the syllabus?
2. How well do TMT items represent the various topics in the course content?
3. What percentage of TMT items are set at different levels of cognitive domain

**Method**

The study used descriptive survey research design. It used both quantitative and qualitative approaches to assess content validity of TMTs for Physics. The study was carried out in secondary schools of Sapele Local Government Area of Delta State. Sapele LGA has 20 public secondary schools. Using simple random sampling technique, 10 secondary schools were selected from which TMTs for Physics for the end of Terms 1, 2 and 3 for SS 2 students of 2019/2020 academic year were used to generate data. Subject Matter Experts (SMEs) were used to collect the data by allocating each item to a particular content area and cognitive level. To avoid bias, the three SMEs came from schools which did not participate in the study. The experts had at least 10 years' experience in teaching Physics in secondary schools. The experts were asked to check the face validity of the items. Specifically, they were asked to find out the suitability and clarity of the language used in the construction of the items. For content validation, the experts were given the questions to guide them. Some of the corrections made by the experts include: inclusion of some omitted items that were supposed to be part of the questions, deletion of items that were not testing any of the concepts taught and restructuring of some questions. These corrections given by the experts were properly effected as regards alignment with the syllabus, course content and levels of cognitive domain. The data were analysed using percentage.

## Results and Discussion

The SMEs were provided with a sample of core elements and topics from Form 3 syllabus for Physics to guide them on allocation of test items to different content areas.

**Table 1:** Core Elements and Topics from Form 3 Syllabus for Physics

| Core Element | Symbol | Topics |
|---|---|---|
| Time | S | Concept of time and ways of measuring time |
| Equilibrium | ⇌ | Equilibrium under parallel and non-parallel forces and moment of a force |
| Electromagnetic Waves | Λ | Types of radiation and Electromagnetic spectrum |
| Simple Harmonic Motion (S.H.M) | F =-kx | Energy of S.H.M, forced vibration and resonance |
| Magnetic Fields | v t e | Concept of magnetic field, magnetic field ground and bar magnet |

### Research Question 1

*To what extent do TMTs for Physics align with the syllabus?*

TMTs for the schools had different number of test items as shown in Table 2.

**Table 2:** Total Number of Test Items per School per Term

| School | Number of items | | |
|---|---|---|---|
| | *Term 1* | *Term 2* | *Term 3* |
| School 1 | 21 | 17 | 18 |
| School 2 | 16 | 20 | 17 |
| School 3 | 20 | 20 | 24 |
| School 4 | 20 | 21 | 20 |
| School 5 | 20 | 22 | 19 |
| School 6 | 21 | 18 | 18 |
| School 7 | 18 | 22 | 22 |
| School 8 | 19 | 21 | 21 |
| School 9 | 18 | 20 | 19 |
| School 10 | 21 | 21 | 20 |

SMEs assigned each item to its appropriate content area. An item was accepted as belonging to a particular content area and cognitive level when 75% of the SMEs reached a consensus. Tables 3, 4 and 5 display the results.

**Table 3:** Item Syllabus Alignment for Term 1 Tests

| Core element | Number of Items | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Sch.1 | Sch.2 | Sch.3 | Sch.4 | Sch.5 | Sch.6 | Sch.7 | Sch.8 | Sch.9 | Sch.10 |
| S | 11 | 10 | 13 | 9 | 12 | 13 | 9 | 19 | 11 | 21 |
| ⇌ | 3 | 4 | 1 | 1 | 0 | 3 | 1 | 0 | 1 | 0 |
| Λ | 5 | 2 | 2 | 4 | 5 | 4 | 3 | 0 | 3 | 0 |
| F = -Kx | 2 | 0 | 4 | 6 | 3 | 0 | 5 | 0 | 3 | 0 |
| vte | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Total | 21 | 16 | 20 | 20 | 20 | 21 | 18 | 19 | 18 | 21 |

**Table 4:** Item Syllabus Alignment for Term 2 Tests

| Core element | Number of Items | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Sch.1 | Sch.2 | Sch.3 | Sch.4 | Sch.5 | Sch.6 | Sch.7 | Sch.8 | Sch.9 | Sch.10 |
| S | 5 | 7 | 8 | 11 | 16 | 8 | 12 | 14 | 12 | 14 |
| ⇌ | 1 | 2 | 1 | 0 | 0 | 0 | 1 | 2 | 1 | 2 |
| Λ | 2 | 5 | 5 | 3 | 3 | 2 | 5 | 4 | 3 | 3 |
| F = -Kx | 3 | 5 | 5 | 6 | 1 | 6 | 3 | 1 | 2 | 1 |
| vte | 5 | 1 | 1 | 1 | 2 | 2 | 1 | 0 | 2 | 1 |
| TOTAL | 17 | 20 | 20 | 21 | 22 | 18 | 22 | 21 | 20 | 21 |

**Table 5:** Item Syllabus Alignment for Term 3 Tests

| Core element | Number of Items | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Sch.1 | Sch.2 | Sch.3 | Sch.4 | Sch.5 | Sch.6 | Sch.7 | Sch.8 | Sch.9 | Sch.10 |
| S | 9 | 12 | 10 | 9 | 6 | 9 | 12 | 12 | 8 | 10 |
| ⇌ | 2 | 0 | 1 | 6 | 3 | 3 | 4 | 2 | 3 | 1 |
| Λ | 3 | 2 | 2 | 0 | 3 | 2 | 3 | 2 | 2 | 3 |
| F = -Kx | 4 | 2 | 7 | 4 | 5 | 4 | 2 | 3 | 4 | 4 |
| vte | 0 | 1 | 4 | 1 | 2 | 0 | 1 | 2 | 2 | 2 |
| Total | 18 | 17 | 24 | 20 | 19 | 18 | 22 | 21 | 19 | 20 |

Table 3 shows that schools 1, 6 and 10 were allocated 21 items each, schools 2 was allocated 16 items, schools 3, 4 and schools 5 were allocated 20 items each, schools 7 and 9 were allocated 18 items each and schools 8 was allocated 19 items. Also the table shows that only one question was set from magnetic field while most questions came from Time
Table 4 shows that schools 1 was allocated 17 items, schools 2, 3, and 9 were allocated 20 items each, schools 4, 8 and 10 were allocated 21 items each, schools 6 was allocated 18 items and schools 7 was allocated 22 items. In Table 4 also, the least questions came from Equilibrium while most of the questions came from Time.
Table 5 shows that schools 1 was allocated 18 items, schools 2 was allocated 17 items, school 3 was allocated 24 items, schools 4 and 10 were allocated 20 items each, schools 5 and 9 were allocated 19 items each, schools 6 was allocated 18 items, schools 7 was allocated 22 items and schools 8 was allocated 21 items. The number of items in Tables 3, 4 and 5 for all ten schools corresponds to the number of items for the tests for the schools in Table 2. Therefore, no item was developed from outside the syllabus. In Table 5 also, the least questions came from Magnetic field while most of the questions came from Time.

**Research Question 2**
*How well do TMT items represent the various topics in the course content?*

**Mean percentages of the number of test items per core element for the three terms were calculated from the data in Tables 3, 4 and 5 and the results were used to draw Figure 1**.

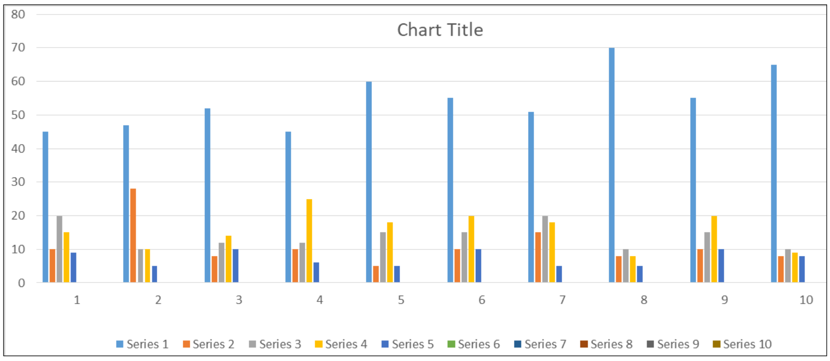


**Fig 1:** Content Representation of Test Items per core Element for Terms, 2 and 3

Figure 1 evidently shows that there was disparity in content representation of the test items of the same core element by the schools. For example, core element NN, content representation ranged from 44% to 74%, core element S, content representation ranged from 3% to 29%, core element SSM, 10% to 20% content representation, core element PRF, content representation ranged from 7% to 25% and core element STAT, content representation ranged from 2% to 8%.

**Research question 3**
*What percentage of TMT items are set at different levels of cognitive domain?*

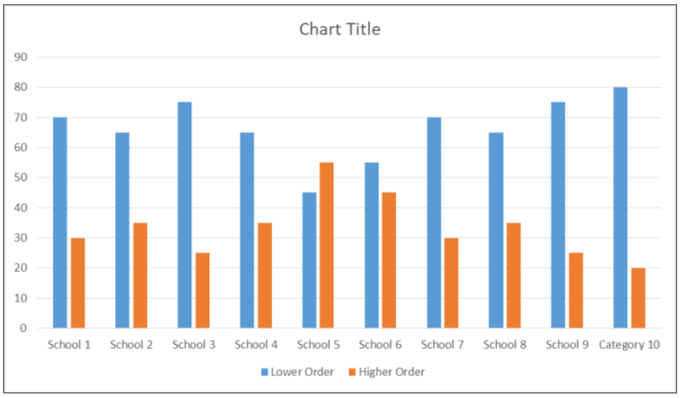


**Fig 2:** Percentages of item cognitive functioning levels for Terms 1, 2and 3

Figure 2 shows that for each of the ten schools, the mean percentages of test items functioning at lower level of cognitive skills were higher than the mean percentages of higher order test items except for School 5, which had higher percentage of higher order items than lower order. Figure 2 also shows that in each school, there was wide disparity between the percentage of lower order items and that of higher order items except for schools 4 and 5. In general therefore, the mean percentages for lower order test items for the 10 schools ranged from 48% to 80% while for the higher order test items, they ranged from 20% to 52%.

**Conclusions**
The findings of the study revealed that teachers develop test items from the syllabus. However, they differ in content representation and item cognitive functioning levels. That is, some core elements were overrepresented at the expense of others and also that most TMTs dwelt more on lower order

cognitive skills than higher order. This is in agreement with Newell (2012) [18] who indicated that TMTs usually measure a limited part of subject area, do not cover a broad range of abilities and rely heavily on memorized facts and procedures.

**Recommendations**

Based on the findings of the study, there is need for secondary school teachers in Sapele LGA to be trained on the development of standard tests for secondary school TMTs to be assessed in terms of content representativeness and cognitive functioning levels for quality assessment. There is also need for NNECO and Ministry of Education for Science and Technology (MoEST) to scale up the study to all districts in Nigeria to assess the validity of TMTs before adopting the assessment system of secondary education of integrating final examination score with CA score.

**References**


1. Anderson LW, Krathwohl DR, Airasian PW, Cruikshank KA, Pintrich PR, Raths J, Wittrock MC. A taxonomy of learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives. New York: Pearson, Allyn & Bacon, 2011.
2. Bachman LF. Fundamental considerations in language testing. New York: Oxford University Press, 2010.
3. Baroudi ZM. Formative assessment: Definition, elements and role in instructional practice. Postgraduate Journal of Education Research,2010:8(1):37-48.
4. Boothroyd RA, McMorris RF, Pruzek RM. What do teachers know about measurement and how did they find out? A Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA, 2012.
5. Chakwera EWJ. Content validity of independently constructed curriculum based examinations (Doctoral Thesis). Amherst: University of Massachusetts, 2014.
6. Chikoko GW. Effects of two teaching strategies: Guided discovery as opposed to expository approach (M.Ed. Thesis). Zomba Publishing Company: University of Nigeria, 2018.
7. Ebel RL. Essentials of educational measurement. (3rd ed.). Eaglewood Cliffs, New Jersey: Prentice Hall, 2019.
8. Fives H, DiDonato-Barnes N. Classroom test construction: The Power of a Table of Specifications. Practical assessment, research & evaluation,2013:18(3):1-7. Retrieved April 27, 2018, from http://pareonline.net/pdf/v18n3.pdf
9. Gronlund NE. Measurement and evaluation in teaching. New York: Macmillan Publishing Co. Inc:, 2011.
10. Gronlund NE. Assessment of student achievement (7th ed.). Boston, MA: Allyn and Bacon, 1998.
11. Gronlund NE. Assessment of student achievement (7th ed.). 2nd Customs Edition for the University of Aberta, 2013.
12. Ing ME, Musah MB, Al-Hudawi SHV, Tahir LM, Kamir NM. Validity of teacher made assessment: A table of specification approach. Asian Journal of Social Science,2015:11(5):193-200.
13. Kinyua A, Okunya C. Validity and reliability of teacher made tests: Case study of year 2 Physics in the Nyahururu District of Kenya. African Educational Research Journal,2014:2(2):61-71.
14. Le Grange L, Reddy C. Continuous assessment: An introduction and guidelines to implementation. Cape Town:
15. Juta Marso RN, Pigge FL. An analysis of teacher made test: Testing practices, cognitive demands and item construction error. Paper presented at the Annual Meeting of the National Council on Measurement in Education. New Orleans, 2018.
16. Messick S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist,2014:50(9):741-749.
17. Ministry of Education, Science and Technology, National Education Sector Plan: A statement. Lilongwe: Ministry of Education, Science and Technology, 2018.
18. Newell RJ. A different look at accountability: The Edvisions approach. Phi Delta Kappan,2012:84 (3):208-211.
19. Nitko AJ. Educational assessment of students. (2nd ed.). New Jersey: Prentice-Hall, 2016.
20. Notar CE, Zuelke DC, Wilson JD, Yunker BD. Table of Specifications: Insuring accountability in teacher made tests. Journal of Psychology,2014:31(2):115-129. Retrieved, 2014. from from http://www.free patentsonline.comarticle/JournalInstructionalPsycholog y/119611686.html.
21. Nxumalo ZF. The role of continuous assessment in primary school. (M.Ed. Dissertation). Kwazulu-Natal: University of Zululand, 2017.
22. Oescher J, Kirby PC. Assessing teacher-made tests in secondary math and science classrooms. Paper presented at the Annual Meeting of the National Council on Measurement in Education. Boston, 2010.
23. Okoye RO. Educational and psychological measurement and evaluation. Awka: Erudition Publishers, 2015.
24. Parr AM, Bauer W. Teacher made test reliability: A comparison of test scores and student study habits from Friday to Monday in a high school biology class in Monroe County Ohio (Master's Thesis). Graduate School of Marietta College. Retrieved 24th January, 2016-2017 from http://etd.ohiolink.edu/sendpdf.cgi/ Parr%20 Anita.pdf?marietta1142864088
25. Sireci SG. Gathering and analyzing content validity Data. Educational Assessment,2018:5(4):299-321.
26. Sireci SG, Faulkner-Bond M. Validity evidence based on test content. Psicothema,2014:26(1):100-107.
27. Wiggins G. A true test: Toward more authentic and equitable assessment. Phi Delta Kappan,2019:70:703-710.
28. Young JW, So Y, Ockey GJ. Guidelines for best test development practices to ensure validity and fairness for international English language proficiency assessments. Retrieved, 2013-2017. from:https:// www.ets.org/s/about/pdf/best_practices_ensure_validity _fairness_english_langua ge_assessments.pdf