
MATHEMATICS ASSESSMENT IN THE BASIC EDUCATION CERTIFICATE EXAMINATION; AN EVALUATION OF CONTENT VALIDITY AND RELIABILITY OF TESTS IN IMO STATE

By

Ike, Iheanacho Chukwudi, Ahara Obianjun O. & Achigasim Philomena. O.

Department of Educational Psychology/Guidance and Counselling,

Alvan Ikoku Federal University of Education, Owerri.

ikeachor5@gmail.com Contact: 08036861273

&

Unamba, Eugene Chukwuemeka

Department of Early Childhood Care Education,

Alvan Ikoku Federal University of Education, Owerri.

unambaec@yahoo.com Contact: 08038956166

&

Ajaero, Henry O.

Father Cloonan Memorial Secondary School Emekuku, Owerri.

henryajaero@gmail.com Contact: 08037853181

&

Eze Anthonia Nwaneze

Federal Government Girls College Owerri, Imo state.

anthonianwaneze1@gmail.com Contact: 08061357035

Abstract

This study investigated the extent to which the mathematics test items used in the Basic Education Certificate Examination (BECE) from 2019 to 2022 in Imo State, Nigeria, aligned with the content and cognitive expectations of the Basic Education Mathematics Curriculum. The study was guided by two research questions and two null hypotheses, using an evaluation research design guided by Stake's Congruence-Contingency Model, all 240 multiple-choice questions from the four years were analyzed for content distribution and cognitive complexity based on Bloom's revised taxonomy. Reliability was assessed using the Kuder-Richardson Formula 20 (K-R20), while content validity was established through expert judgment. Descriptive statistics and correspondence analysis were employed for data analysis. Results showed no significant deviation between observed and expected content distributions ($p > 0.05$), but a significant misalignment in cognitive level distributions ($p < 0.05$). Findings highlight the need to enhance the cognitive rigor of assessments to better reflect curriculum intentions.

Keywords: BECE, Mathematics Assessment, Content Validity, Cognitive Levels, and Reliability Curriculum Alignment.

Introduction

Assessment plays a critical role in evaluating students' mastery of curriculum content and cognitive skills. The Basic Education Certificate Examination (BECE), administered annually in Nigeria, is a key tool for certifying student learning at the junior secondary level. However, there are growing concerns about the extent to which BECE mathematics tests align with the curricular standards and learning objectives outlined by the Nigerian Educational Research and Development Council (NERDC). Prior research has indicated disparities between test content and curricular demands, leading to questions about the validity and cognitive balance of these assessments (Iweka, 2008; Ogugua et al., 2020). Misaligned assessments can distort teaching focus and hinder students' development of higher-order thinking skills. This study aimed to evaluate the content validity, reliability, and cognitive alignment of BECE mathematics tests administered in Imo State between 2019 and 2022.

Mathematics is a core discipline that underpins human reasoning, scientific inquiry, and technological innovation. It is not merely a school subject but a tool for understanding the world and solving practical problems. Carl, as cited in Yadav (2017), referred to mathematics as the "queen of science," highlighting its foundational role in all scientific domains. Despite this, many students harbor fear and anxiety toward the subject, which impedes their academic progress and limits their future opportunities. Mathematics enhances mental acuity and fosters logical thinking, creativity, and problem-solving skills. According to Ayebale, Habaasa, and Tweheyo (2020), mathematics is central to national development and is often viewed as the bedrock of scientific and technological growth. In recognition of its importance, mathematics is taught from early childhood and is made compulsory at the secondary level (Mollah, 2017). Das and Das (2013) argue that the growing reliance on data and technology in society necessitates strong numeracy skills, further reinforcing the need for effective mathematics instruction. The acquisition of skills such as arithmetic operations, measurement, and financial calculations forms the basis for real-life applications and is best achieved through structured school curricula (Pedagogy of Mathematics, 2019). However, despite concerted efforts, students in Nigeria often exhibit poor performance in mathematics, especially in standardized assessments such as the BECE.

Assessment plays a central role in education by measuring student progress, guiding instruction, and informing policy. Donna, cited in Ogugua et al. (2020), defined examinations as deliberate mechanisms for evaluating knowledge, skills, and performance in specific subject areas. For assessments to be meaningful, they must align with learning objectives (Getinet, Solomon, & Yilfashewa, 2021). Among the characteristics of quality assessments, validity holds primacy. Nwana (2009) asserted that a test lacking validity is essentially meaningless. Validity ensures that a test accurately measures what it claims to measure, while reliability ensures consistency in measurement (Nwana, 2007). Obe, cited in Ogomaka (2016), described validity as the accuracy of a test in evaluating student achievement in a defined content area. Kimberling and Winterstein, as cited in Haradham (2017), stressed that both validity and reliability are critical to the effectiveness of any assessment instrument.

Content validity specifically refers to the extent to which test items represent the full range of subject matter and objectives intended for measurement. Bush (cited in Getinet et al., 2021) emphasized that content validity evaluates the representativeness of test content. Mehran and Lehmann (cited in Abonyi, 2011) defined it as the degree to which a test adequately samples the domain from which inferences are made. Ugodulunwa and Wakjissa (2016) noted that content validity is achieved when test items align with essential objectives, accurately reflect curriculum weightings, and include various cognitive levels. content validity of a test for

assessing skills in vocational and technical education shows whether the test covers the areas it is supposed to cover. Okoro (2020) observed that establishing the content validity of a test for assessing skills is very essential because it enables the examiner to determine the degree of coverage given by the test to each item of the skills outlined in the curriculum. Anatasi (2019) noted that content validating a test answers two major questions namely; does the test cover a representative sample of the specified skills and knowledge? Is the test performance reasonably free from the influence of irrelevant variables? Answers to these questions imply establishing content validity of a test. Since tests are constructed based on objectives of instruction or course outline there is therefore, a need for thorough assessment of items of the test in order to establish the relevance of the test items to the objectives of the instruction or course. Therefore, content validation of test for assessing skills in vocational and science education is carried out to ensure that the content of the test is relevant to the objective of the assessment sufficiently complete, uncontaminated with other content and at the appropriate level of difficulty.

There are several methods of determining the content validity of a test. Mehrens and Lehman (1984) suggested a method that involves asking experts in the field of measurement and evaluation to thoroughly inspect and judge test for proper wording, consistency and reviewing the test based on the suggestions of the experts. While Butler (1976) suggested that after a test has been constructed and initially content validated by experts during construction, it should be given to subject matter specialists for further review, corrections, and assessment of appropriateness of the test items. Furthermore, Gachino and Galington (1976) suggested that the larger the number of subject specialists used for assessing the appropriateness of test items the better the content validity of that test.

However, the method of content validation by employing experts has been criticized for subjectivity because of lack of quantitateness. Tuckman (1976) suggested that the subjectivity of lack of quantitateness can be eliminated by using rating scale. The rating scale would provide means for the experts to express their views on those items of test to be retained and those to be removed. While Mehrens and Lehman (1984) suggested that two tests should be constructed over the same content areas and be given to same subject matter specialists for analysis. The results should be correlated to produce quantitative value. Another quantitative method of determining content validity of a test was suggested by Ohuns (2020) when he observed that by building a Table of specification into the process of test construction, it is possible to provide quantitative value of content validity. Earlier Green (1976) defined a table of specification as a table that provides general outline of intended emphasis of assessment and the assessment approaches. UNESCO (2002) described the Table of specification as a two chart in which content of a course/ topic are correlated to the outcome/competencies, which describe the skills to be achieved from the course of the study. Oribhabor and Emafor (2016) found that many secondary school assessments in Nigeria are administered without prior validation. Onyekuba and Anyichie (2013) also noted that many teachers lack the training and expertise necessary for constructing high-quality test items, further exacerbating the problem. Persistent underachievement in mathematics among Nigerian students raises concerns about the quality of teaching and assessment. Awofala and Fatade (2023) identified several contributory factors, including ineffective pedagogy, negative student attitudes, and lack of parental involvement, psychological barriers, and administrative shortcomings. Assessment quality, in particular, remains a significant challenge. Okonkwo, cited in Ukozor, Onah, and Ukah (2017), criticized Nigerian assessments for lacking validity and reliability. Many test items, especially multiple-choice questions, are poorly constructed and fail to reflect curriculum objectives, confusing even

capable students. Faleye (cited in Oribhabor & Emafor, 2016) emphasized that test format and item quality greatly affects student performance. Similarly, Osadebe, also cited in Ukozor et al. (2017), noted that students' success depends on expert test construction, comprehensive curriculum coverage, and the implementation of reliable and valid evaluation tools.

This study aims to assess the content validity and reliability of mathematics test items used in the BECE in Imo State. Specifically, the researchers sought to:

Purpose of the study

- i. ascertain how closely the content of mathematics tests constructed by the Education Development Centre (EDC), Imo State, aligns with the Basic Education Mathematics Curriculum.
- ii. determine the extent to which the cognitive objectives assessed by EDC test items reflect those specified in the Basic Education Mathematics Curriculum.

Research Questions

1. In what content areas do the percentage distributions of test items in the BECE mathematics examination differ from the expected distribution in the Basic Education Mathematics Curriculum (BEMC)?
2. At what cognitive levels do the percentage distributions of test items in the BECE mathematics examination differ from the expected distribution in the BEMC?

Hypotheses

H₀₁: There is no significant difference between the observed distribution of test items across content areas in the BECE mathematics examination and the expected distribution outlined in the Basic Education Mathematics Curriculum.

H₀₂: There is no significant difference between the observed distribution of test items across cognitive levels in the BECE mathematics examination and the expected distribution outlined in the Basic Education Mathematics Curriculum.

Method

This study employed an evaluation research design to investigate the degree of alignment between the content and cognitive levels of mathematics test items used in the Basic Education Certificate Examination (BECE) in Imo State and the expectations outlined in the Basic Education Mathematics Curriculum. The Stake's Congruence-Contingency Model was adopted, focusing on the extent to which observed test items align with intended curricular goals and cognitive objectives. The study was conducted in Imo State, Nigeria, and covered BECE mathematics tests administered between 2019 and 2022. A total of 240 multiple-choice questions (60 per year) constructed by the Ministry of Education were analyzed using a census approach, meaning all items across the four years were included. To determine the reliability of the test items, a trial test was administered to 120 Senior Secondary School 1 (SS1) students from two private schools in Imo State. The Kuder-Richardson Formula 20 (K-R20) was employed to estimate internal consistency. The computed reliability coefficients were 0.9811 (2019), 0.9461 (2020), 0.9515 (2021), and 0.9580 (2022), indicating high reliability. To establish content validity, the test items were reviewed by one expert in Measurement and Evaluation with a specialization in mathematics, and two experienced mathematics teachers with over ten years of teaching experience. These experts mapped the items to the curriculum and classified them based on content areas and Bloom's revised taxonomy of cognitive domains. A test blueprint was used to ensure consistency in

classification. Expected cognitive level distributions were drawn from the national assessment framework: Lower-Level Cognitive Skills (LLCS) at 60% (Remembering - 25%, Understanding - 35%) and Higher-Level Cognitive Skills (HLCS) at 40% (Applying - 25%, Analyzing - 10%, Evaluating - 5%, Creating - 0%). Descriptive statistics (frequencies and percentages) were used to analyze the research questions, while correspondence analysis was applied to test the null hypotheses. A p -value > 0.05 indicated no significant difference between observed and expected distributions, while $p < 0.05$ indicated a significant difference.

Results

Research Question 1. In what content areas do the percentage distributions of test items in the Basic Education Certificate Examination (BECE) differ from the expected percentage distribution in the Basic Education Mathematics Curriculum (BEMC)?

Hypothesis 1: There is no significant difference between the observed distribution of test items across content areas in the Basic Education Certificate Examination Mathematics and the expected distribution of test items as outlined in the Basic Education Mathematics Curriculum.

Table 1. Percentage distribution of content areas of observed and expected; Analysis of Content Area Distributions

Year/ Themes	Expected	2019	2020	2021	2022	Observed
Number and Numerations	40 (12%)	08(13%)	10(17%)	08(13%)	07(12%)	33(13.75 %)
Basic Operations	88(27%)	17(28%)	21(35%)	15(25%)	18(25%)	71(28.25%)
Algebraic Operations	84(26%)	14(24%)	16(27%)	22(37%)	12(20%)	64(27%)
Mensuration and Geometry	76(24%)	16(26%)	08(13%)	10(17%)	16(26%)	50(20.50%)
Statistics	36(11%)	05(08%)	05(08%)	05(08%)	07(12%)	22(09%)

Table 1 presents the percentage distributions of observed and expected content areas for the mathematics tests. The results from the correspondence analysis indicate that the distributions of observed and expected proportions are closely aligned. This similarity suggests that there is no significant deviation in the distribution of test items when compared to the established curriculum. To further evaluate the alignment between the observed distributions of test items and the expected distributions, we refer to the summary table from the correspondence analysis. The resulting p -value is $p = 0.953$ ($p > 0.05$), indicating that there is no significant difference between the observed and expected distributions of test items. Therefore, we accept the null hypothesis, concluding that the test items are appropriately aligned with the curriculum standards.

Research Question 2. In what cognitive level do the percentage distributions of test items in the Basic Education Certificate Examination differ from the expected percentage distribution in the Basic Education Mathematics Curriculum?

Hypothesis 2: There is no significant difference between the observed distribution of test items across cognitive levels in the Basic Education Certificate Examination Mathematics

and the expected distribution of test items as outlined in the Basic Education Mathematics Curriculum.

Table 2. Percentage distribution of cognitive levels of observed and expected; Cognitive Level Distribution Analysis

Table 2 illustrates the percentage distributions of cognitive levels observed over four years compared to the expected percentage distribution. The analysis reveals significant disparities between the observed and expected distributions across cognitive levels. Specifically, Level 1 (Remembering) and Level 3 (Applying) demonstrated the most substantial deviations from the anticipated distribution. Level 5 (Evaluating) also showed a notable deviation, albeit less

Year/ Cognitive	Expected	2019	2020	2021	2022	Observed
Remembering	25%	08 (14%)	11 (18%)	09 (15%)	13 (21%)	41(17%)
Understanding	35%	20 (33%)	20 (33%)	21 (35%)	25 (42%)	86(35.75%)
Applying	25%	23 (38%)	21 (35%)	24 (40%)	15 (25%)	83(34.5%)
Analysing	10%	07 (12%)	06 (10%)	06 (10%)	07 (12%)	26(11%)
Evaluating	5%	02 (03%)	02 (03%)	00 (00%)	00 (00%)	04(1.5%)
Creating	0%	00 (0%)	00 (00%)	00 (00%)	00 (00%)	00(00%)

pronounced than Level 3, indicating its significant contribution to the overall differences between observed and expected distributions. Conversely, Level 6 (Creating) aligned more closely with expectations. The correspondence analysis summary table indicates a p-value of $p = 0.040$ ($p < 0.05$), suggesting that the observed distribution of test items across cognitive levels in the constructed BECE significantly differs from the expected distribution as outlined in the mathematics curriculum. This finding provides strong evidence to reject the null hypothesis, indicating that the observed distribution of items across cognitive levels in the BECE does not align with the expected distribution in the mathematics curriculum.

Discussion of Findings

The basic education curriculum encompasses five overarching themes, each subdivided into specific content areas. The results indicate that, on average, over the four-year period, the percentage distributions of content in the mathematics tests constructed by the state's Ministry of Education closely corresponded to the expected percentages in the basic education mathematics curriculum. This alignment suggests a significant correlation between the expected and observed content percentages, highlighting the importance of maintaining consistency in content allocation. These findings contrast with previous studies by Iweka (2008), Ogugua et al. (2020), and Buba and Kojigili (2020), which reported discrepancies between observed and expected content percentages.

Another significant finding of this study is the notable differences at Level 1 (Remembering) and Level 3 (Applying), where actual distributions deviated from expected benchmarks. This implies an overemphasis on procedural and conceptual recall with inadequate assessment of higher-order thinking skills, such as evaluation and creativity—critical components in mathematics learning (Bloom, 1956; Abonyi, 2011). The results underscore the need to review the distribution of test items across cognitive levels to ensure that assessments accurately reflect the intended learning outcomes as prescribed by the curriculum. This conclusion is supported by the works of Oribhabor et al. (2016), Ogugua et al. (2020), Buba et al. (2020), and Gittinet et al. (2021). It is evident that the weighting of cognitive objectives did not adhere to a predetermined plan or specification that considers students' academic levels.

Furthermore, although high reliability coefficients were recorded across all years, validity remains a concern, particularly as some test items lacked correct options. This reiterates the assertion by Iwuji (in Ogomaka, 2016) that a test may be reliable without being valid, underscoring the need for comprehensive test development strategies that balance both metrics.

Conclusion

The study concludes that while the BECE mathematics tests in Imo State from 2019 to 2022 demonstrate satisfactory content validity and reliability, they fall short in aligning with the cognitive demands of the curriculum. The disproportionate emphasis on lower-level thinking skills undermines the curriculum's intent to foster analytical and evaluative reasoning among learners. These findings call for systemic improvements in the assessment construction process, particularly the use of test blueprints that incorporate both content balance and cognitive complexity. Consequently, the BECE mathematics test items from 2019 to 2022 exhibit a low degree of content validity.

Recommendations

Based on the findings, the following recommendations are proposed:

1. When developing test items, it is essential to prioritize both validity and reliability. The responsible office should prepare a well-structured test plan that accurately represents the contents and cognitive levels outlined in the syllabus.
2. Engage experts in mathematics and measurement and evaluation to guide the construction of tests, ensuring that the cognitive skills of students are appropriately considered.
3. Teachers and experts responsible for preparing examinations should receive ongoing training on how to construct questions using a table of specifications. This training will help ensure a congruent distribution of test items across cognitive levels.

References

- Abonyi, O. S. (2011). *Instrumentation in behavioural research: A practical approach*. Enugu: Fulladu publishing company.
- Anderson, L. W. & Krathwohl, E. (2001). *A taxonomy for learning, teaching and assessing. A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

- Awofala, O.A., & Fatade, O.A. (2023). Nigerian students' poor performance in mathematics: who do we blame? *Nigerian Online Journal of Educational Sciences and Technology (NOJEST)*, 5 (1), 80-91
- Ayebale, L., Habaasa, G. & Tweheyo, S. (2020). Factors affecting students' achievement in mathematics in secondary school in developing countries. *Statistical Journal of IAOS*. 36(1), 73-76.
- Bloom, B. S. (1956). *Taxonomy of educational objectives*. Handbook: the cognitive domain. New York: Mc Kay.
- Buba, M. A., & Kojigili, S. T. (2020). Content validation of basic education certificate examination questions in basic science in Adamawa State, Nigeria. *Journal of Education, Society and Behavioural Science*, 33(2), 42-54. <https://doi.org/10.9734/JESBS/2020/v33i23020>
- Das, R. & Das, G. C. (2013). Math anxiety: the poor problem solving factor in school mathematics. *International Journal of Scientific and Research Publications*, 3(4), 1-5.
- Getinet, A. W., Solomon, F., & Yilfashewa, S. (2021). *Evaluating the Content Validity of Grade 10 Mathematics Model Examinations in Oromia National Regional State, Ethiopia*. <https://doi.org/10.1155/2021/5837931>
- Haradham K. M. (2017). Two criteria for good measurement in research: validity and reliability. *Annals of Spiru Horet University*. 17(3): 58-82
- Ihekweba, C. N. & Osuala, R. C. (2017). Principles of Measurement and Evaluation in Education. Owerri: Zionbless publishers Nigeria.
- Iweka, F. E. O. (2008). Content validity of the senior secondary certificate examination in Biology. <https://www.globalacademicgroup.com>
- Mollah, K. (2017). Mathematics anxiety among the school students. *Pramana research journal*. 7(11), 122-129.
- Nwana, O. C. (2009). *Educational Measurement and Evaluation*. Nigerian Academy of Education.
- Ogomaka, P.M.C., Ekwonye, E.C., Ukozor, F. I. & Onah, F.E. (2016). *Measurement and Evaluation: Comprehensive Text book for students and teachers*. Owerri: Flashpoints publishers.
- Ogugua, B. C., Agah, J.J., Acholonu, V.V, Azubuike, R.N., Okeke, M.A., Agbo, P.L. & Ene, C. U. (2020). Content validity of WAEC financial accounting questions. *International Journal of Learning, Teaching and Educational Research*. 19(6), 161-178.
- Onyekuba, C. & Anyichie, A. C. (2013). "Measuring teachers' competencies in constructing classroom-based tests in Nigeria secondary schools need for a test construction skill inventory", *Educational Research and Reviews*. 8(8), 431- 441.
- 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, Science and Technology*. 3(2): 83-90
- Pedagogy of Mathematics (2019). *Aims and Objectives of Teaching Mathematics*.
- Ugodulunwa, C.A., & Wakjissa, S. G. (2016). What teachers know about validity of classroom tests: evidence from university in Nigeria. *Journal of Research and Method in Education*. 6(3). 14-19. Google scholar.
- Ukozor, F. I., Onah, F. E., & Ukah, F. E. (2017). Application of three parametric latent trait model in development of secondary school economics achievement test. *Nigerian Journal of Educational Research and Evaluation*. 16(1), 79-99.

APENDIX

CORRESPONDENCE ANALYSIS SUMMARY TABLE FOR CONTENT AREA

Dimension	Singular Value	Inertia	Chi Square	Sig.	Proportion of Inertia		Confidence Singular Value
					Accounted for	Cumulative	Standard Deviation
1	.059	.003			1.000	1.000	
Total		.003	.684	.953 ^a	1.000	1.000	.071

a. 4 degrees of freedom

CORRESPONDENCE ANALYSIS SUMMARY TABLE FOR COGNITIVE LEVELS

Dimension	Singular Value	Inertia	Chi Square	Sig.	Proportion of Inertia		Confidence Singular Value
					Accounted for	Cumulative	Standard Deviation
1	.242	.058			1.000	1.000	
Total		.058	11.676	.040 ^a	1.000	1.000	.068

a. 5 degrees of freedom