JOURNAL OF INNOVATIONS IN SCIENCE EDUCATION (JISE)
Vol. 2(2); 2025

ASSESSING THE VALIDITY AND QUALITY OF MATHEMATICS TEST CONTENT IN THE BASIC EDUCATION CERTIFICATE EXAMINATIONS IN IMO STATE, NIGERIA

[1]Iheanacho C. Ike, [2]Chinemeze J. Ogoke, [3]Chioma L. Ahanotu, [4]John-Donald Ikegwuruka, [5]Henry O. Ajaero

[1]ikeachor5@gmail.com, [2]james4justics@gmail.com,
[3]ahanotuchioma2018@gmail.com, [4]johndonald.ikegwuruka@alvanikoku.edu.ng,
[5]henryajaero@gmail.com

[1, 2,3]Department of Mathematics, Alvan Ikoku Federal University of Education, Owerri. Imo State, Nigeria.

[4]Department of Psychology/ G & C, Alvan Ikoku Federal University of Education, Owerri, Imo State, Nigeria.

[5]Father Cloonan Memorial Secondary School Emekuku, Owerri, Imo State, Nigeria.

Abstract

This study assessed the content validity and quality of Mathematics tests used in the Basic Education Certificate Examination (BECE) in Imo State, Nigeria. Guided by three research questions and three null hypotheses, the study adopted an analytic survey design. The population comprised 240 test items, which also served as the sample, drawn from BECE Mathematics examinations administered between 2019 and 2022. Frequency counts and proportions were used to address the research questions, while chi-square and one-sample proportion tests were employed to test the hypotheses. Findings revealed that the extent of dependence of test items on BECE Mathematics content areas was significantly independent of the years. The proportional distribution of test items across content areas was not significantly greater than 0.75, and there was no significant association between the years and the distribution of item facility indices within the acceptable range. Overall, the quality of the tests, as measured by item facility indices, remained stable, though the proportions of items within the acceptable range varied across the years. It was concluded that although the Mathematics tests demonstrated reliability, they exhibited a low degree of content validity. The study recommends that subject experts in Mathematics, as well as specialists in measurement and evaluation, be actively involved in test construction to ensure alignment with curriculum standards and to enhance the validity of examination items.

Keywords: Mathematics test, Content validity, Item facility indices.

Introduction

Assessment is a fundamental component of education, serving as a means to measure learning outcomes and competencies, thereby guiding informed decisions about students and educational programs. It involves the systematic process of collecting, recording, scoring, and interpreting data related to students' learning progress. The primary goal of assessment is to facilitate improvement in both teaching and learning (Ighodaro & Orheruata, 2024). In educational settings, assessment is integral to evaluating students' academic performance throughout a course (Talebi, Ghaffari, Eskandarzadeh, & Oskouei, 2013). It functions as a diagnostic tool that not only measures student competence but also identifies discrepancies between intended learning objectives and actual learning outcomes, providing valuable feedback to educators, students, and policymakers. Such insights highlight the central role of assessment in shaping both classroom practices and broader educational policies.

Building on this understanding, a teacher's knowledge of assessment and evaluation is both complex and dynamic, requiring continuous engagement with assessment principles and methodologies (Xu & Liu, 2009). In many classrooms, assessments are often designed and administered by teachers themselves, making the validity and reliability of these tests a critical concern (Carroll & Moody; Boothroyd et al., cited in Anna Siri, 2011). Since evaluation serves multiple purposes, ensuring its validity is essential to achieving its intended objectives. Given the significance of standardized assessments in educational decision-making, it is therefore necessary to examine the quality and validity of mathematics test content in the Basic Education Certificate Examination (BECE) in Imo State. This study aims to assess the alignment of BECE mathematics test items with the expected curriculum standards, ensuring that the examination effectively measures students' mathematical competencies.

Mathematics is a fundamental part of human thought and logic, essential for understanding the world and ourselves. It is seen as the foundation of scientific and technological knowledge vital for a nation's socio-economic development (Ayebale, Habaasa & Tweheyo, 2020). It is a way of thinking and organizing logical proofs, primarily concerned with finding answers to questions and problems (Agah, 2020). Despite its importance, many variables have been identified as responsible for poor academic performance in Mathematics in Nigeria, including teachers, students, parents, psychological factors, the Ministry of Education (government) and the role of examination bodies (Awofala & Fatade, 2023). Consequently, educators, researchers, and policymakers are continually seeking solutions to reduce poor performance to the

barest minimum. According to Faleye, cited in Oribhabor & Emafor, (2016) one of the critical area of concern is the nature of test formats used during regular classroom evaluation, since the quality of items has a direct impact on the students' learning outcomes and achievement. Sharma (2000), also noted that the quality of a test depends on each item.

The issue of test quality is closely tied to the process of item analysis. According to Anna-Siri (2011), item analysis involves examining students' responses and test questions to assess both the quality and quantity of items, as well as the overall test structure. This process plays an integral role in course assessment, as it enables educators to identify the strengths and weaknesses of individual test items, thereby improving the reliability and overall examinations quality. Item analysis yields information about the characteristics of items used in a test, providing the basis for evaluating a test's validity and reliability (Abdullahi, 2020). Through item analysis, teachers and researchers can determine whether items are too difficult, too easy, or ineffective in discriminating between high- and low-achieving students. Such information provides a solid basis for evaluating the validity and reliability of tests and enhances the credibility of assessment practices.

In addressing these concerns, this study adopts the Classical Test Theory (CTT) framework, which assumes that each individual has a true score that would be obtained if there were no errors in measurements. Since measuring instruments are imperfect, the observed score for each person differs from their true ability (Ogunka & Amadi, 2019). Within this framework, test items are described in terms of indices such as discrimination, distracter effectiveness, and difficulty. Of particular importance is the Item Facility Index (IF) (also known as the item difficulty index), which measures the proportion of examinees who answer a test item correctly (Anna-Siri, 2011). This index helps to determine whether an item is appropriately pitched for the target learners. Items that are too easy or too difficult reduce the effectiveness of an assessment, while those within the acceptable range (0.30–0.80) maintain an appropriate balance of challenge and fairness. Items with an index below 0.30 are considered too difficult, whereas those above 0.80 are deemed too easy (Nathan, 2020). Such measures ensure that classroom and standardized tests can effectively differentiate among students while still reflecting meaningful learning.

Closely linked to these considerations are the psychometric properties of validity and reliability, which are fundamental to the usefulness of any test (Elliot, Kratochwoli, Cook & Travers; in Nwana, 2009). According to Nwana (2009), among these

properties, validity is regarded as paramount because a test has little value if it does not measure what it is intended to measure. Content validity, in particular, refers to the extent to which a test adequately represents the content domain from which it is drawn (Bush as cited in Getinet, Solomon, &Yilfashewa, 2021). It is strictly concerned with the extent to which items of a test have achieved representativeness of the content from where the measuring instrument was drawn (Mehran's & Lehmann cited in Abonyi, 2011). Content validity is determined by the test's coverage of essential objectives and content, as well as an adequate sampling of essential curriculum content (Ugodulunwa & Wakjissa, 2016). This requires careful coverage of curriculum objectives and sufficient sampling of relevant content areas.

Statement of the Problem

Unfortunately, many classroom evaluation items used in Nigerian secondary schools are not subjected to proper validation before use, largely due to teachers' limited test-construction skills. (Oribhabor & Emafor, 2016; Onyekuba & Anyichie, 2013). This situation raises genuine concerns about the quality of large-scale examinations such as the Basic Education Certificate Examination (BECE). In Imo State, doubts remain about whether the mathematics tests administered by the Ministry of Education truly undergo rigorous validity procedures. Addressing these doubts forms the central motivation for the present study, which seeks to examine both the distribution of content areas and the item facility indices of BECE mathematics tests.

Purpose of the Study

The main purpose of this study is to assess the Validity and Quality of Mathematics Test Content in the Basic Education Certificate Examinations in Imo state, Nigeria. The researchers in this study specifically:

1. Examined the association between the content distribution of the BECE mathematics test items constructed by Imo State's Education Development Centre (EDC) and the expected content distribution from the Basic Education Mathematics Curriculum across the years of examination
2. Determined whether the proportion of BECE mathematics test items with facility indices within the acceptable range prescribed by Classical Test Theory differs significantly from 0.75.
3. Investigated the association between the distribution of BECE mathematics test items' facility indices within the acceptable range and the years of examination.

Research Questions

The following research questions were answered in the study;

1. To what extent does the distribution of the content of the BECE mathematics test items constructed by Imo State's Education Development Centre (EDC) align with the expected content distribution from the Basic Education Mathematics Curriculum across the years of examination?
2. What is the proportion of BECE mathematics test items with facility indices that fall within the acceptable range of Classical Test Theory, and how does this compare with the benchmark value of 0.75?
3. How does the distribution of BECE mathematics test items' facility indices within the acceptable range vary across the years of examination?

Hypotheses

The following null hypotheses were formulated to guide the study and each tested at the 0.05 level of significance.

$H0_1$: There is no significant association between the distributions of BECE mathematics test items across content areas constructed by Imo state's EDC and the expected content distribution from the Basic Education Mathematics Curriculum across the years of examination.

$H0_2$: The proportion of BECE mathematics test items with facility indices within the acceptable range does not differ significantly from 0.75.

$H0_3$: There is no significant association between the distribution of BECE mathematics test items' facility indices within the acceptable range and the years of examination.

Methodology

This study employed an analytic survey research design to investigate the relationship between the content validity and quality of the Basic Education Certificate Examination (BECE) Mathematics test across different examination years. As noted by Oppenheim (as cited in IvyPanda, 2024), an analytic survey is specifically structured to explore associations between variables.

The research was conducted in Imo State, Nigeria, focusing on a population of two hundred and forty (240) items from past Mathematics examination question papers for the BECE, constructed by the Imo State Ministry of Education from 2019 to 2022. Each question paper comprised 60 multiple-choice questions. A census sampling technique was utilized to include all items from each examination year.

The primary instrument for this study was the BECE Mathematics past question papers. To establish the reliability of the Mathematics tests constructed by the Education Development Center (EDC) for each year, a trial test was conducted. This trial involved administering the instruments to 120 senior secondary level one (SS1) students from three private schools in Imo State. The Kuder-Richardson Formula 20 (K-R20) was employed to calculate the reliability coefficients for each instrument, resulting in coefficients of 0.9811, 0.9461, 0.9515, and 0.9580. To ensure the face validity of the instruments, the researchers consulted two experienced mathematics teachers and one expert in Measurement and Evaluation.

Research questions were addressed using descriptive statistics, including frequency counts and percentages/ proportions. Additionally, null hypotheses were tested using the Chi-square test of independence and one-sample proportion tests at a significance level of 0.05.

Results

The results are presented in order of the research questions and hypotheses raised.

Research Question 1: To what extent does the distribution of the content of the BECE mathematics test items constructed by Imo State's Education Development Centre (EDC) align with the expected content distribution from the Basic Education Mathematics Curriculum across the years of examination?

$H0_1$: There is no significant association between the distribution of BECE mathematics test items across content areas constructed by Imo state's EDC and the expected content distribution from the Basic Education Mathematics Curriculum across the years of examination.

Table 1: Extent of dependency of the distribution of the observed contents of Imo State's EDC Basic Education Certificate Examination and Chi-square test of independence

	CONTENT	YEAR				
		2019	2020	2021	2022	p-value
1	2Whole numbers/ counting	2	1	1	2	
2	Factors/multiples/HCF/LCM	3	1	3	3	0.93
3	Squares/roots/prime numbers	2	7	3	1	
4	Number bases	1	1	1	1	
5	Fractions/ratios/rates	7	9	6	9	
6	Estimation/ approximation	3	6	7	5	
7	Transactions in homes/offices	3	4	2	3	
8	Directed numbers/Order of operations	1	1	0	1	
9	Change of Subject of the formula	0	0	1	1	
10	Algebraic Simplifications	7	9	14	6	
11	Simple equations	3	2	4	1	
12	Simultaneous equations	1	1	1	1	
13	Quadratics equations	1	2	1	3	
14	Inequalities	2	2	1	0	
15	Areas/perimeters/volumes	4	2	0	5	
16	Plane shapes/ similar shapes	0	1	2	2	
17	Angles/ polygons	2	1	6	6	
18	Trigonometry	3	3	2	2	
19	Angle of elevation/depression	0	0	0	1	
20	Bearing	0	0	0	0	
21	Constructions	0	0	0	0	
22	Data collection/presentations	2	1	1	5	
23	Measures of central tendency	3	2	1	1	
24	Probability	1	2	3	1	
		60	60	60	60	

Table 1. Shows the extent of distribution of the curriculum contents as represented in the Imo state's constructed mathematics test from 2019 to 2022. No question was set in bearing and construction across the four years under study. In 2022 one question came out in angle of elevation and depression but none in 2019, 2020 and 2021. The Chi-Square test of independence yielded a p-value of 0.93 ($p > 0.05$), leading us to accept the null hypothesis. This result provides a sufficient evidence to conclude that the extent of dependence of test items drawn from the BECE mathematics content areas is significantly independent of the years.

Furthermore, the summary of the data analysis results to answer RQ2 and Ho2 are presented in table 2.

Research Question 2: What is the proportion of BECE mathematics test items with facility indices that fall within the acceptable range of Classical Test Theory, and how does this compare with the benchmark value of 0.75?

$H0_2$: The proportion of BECE mathematics test items with facility indices within the acceptable range does not differ significantly from 0.75.

Table 2, proportions of items facility indices of Imo state's EDC constructed mathematics test and the one sample proportion test to certify if the content areas are significantly greater than 0.75 at a significance level of 0.05.

YEARS	P	%	Z_{cal}	$Z_{critical}(0.05)$
2019	0.7	70	-0.894	1.645
2020	0.77	77	0.354	
2021	0.7	70	- 0.894	
2022	0.4	40	-6.26	
Average	0.6425	64	-1.923	

Table 2 shows that from each of the sixty (60) items in the Imo state EDC's constructed mathematics tests, considering Classical Test Theory (acceptable range of item facility indices between 0.3 and 0.7) the following proportions were calculated; 0.7 (70%), 0.77 (77%), 0.7 (70%) and 0.4 (40%) across the years under study. The overall average proportion is; 0.64 (64%). The result of the one sample proportion test shows that in

2019, 2020, 2021, and 2022, Zcal. = -1.923 while $Z_{critical}$, = 1.645. $Z_{cal.} < Z_{critical}$, indicating that the null hypothesis is accepted.

Furthermore, the summary of the data analysis results to answer RQ3 and Ho3 are presented in table 3.

Research Question 3: How does the distribution of BECE mathematics test items' facility indices within the acceptable range vary across the years of examination?

$H0_3$: There is no significant association between the distribution of BECE mathematics test items' facility indices within the acceptable range and the years of examination.

Table 3, distribution of the item facility indices within the accepted range (0.3 to 0.7) according to CTT model and Chi-Square test of independence at a 0.05 significance level for distribution of the item facility of the constructed tests.

ITEM FACILITY RANGE / YEARS	**ACCEPTED (0.3-0.7)**	**REJECTED (0.1, 0.2, 0.8, 0.9, 1.0)**
2019	19	41
2020	22	38
2021	21	39
2022	12	48
Df	3	
p_{-value}	0.19	

Table 3, shows the extent of distribution of item facility indices within accepted range (neither too easy nor too difficult). In 2019, 19 out of the 60 items were within acceptable range, 22 items were within range in 2020, 21 items were within range in 2021 and 12 items were within range in 2022. The value is 4.77; degree of freedom (df) is 3, at significant level of 0.05 and p-value of 0.19. The p-value is greater than the significance level of 0.05 ($p > 0.05$), therefore, reject the null hypothesis.

Discussion

In Table 1, the findings revealed that while the overall content distribution was high, some content areas were overemphasized, underemphasized, or omitted entirely. The Chi-Square test of independence analysis showed that the distributions of the test items were consistent across different examination years, indicating no significant dependency on the year of the examination. This agree with Seitz (2017), who found that some topics were over-represented while others were missing across years. Also it agrees with Morrow, Rate, and Evans (2021) they found that lack of standardized content led to inconsistencies in assessment topics over time. But study according to Martone and Sireci (2009) disagrees with the present study. The overemphasis on certain contents and neglect or omission of others may lead to complacency among teachers and students, as they concentrate on certain areas while disregarding others. This situation may arise due to a lack of attention from teachers, curriculum developers, and education experts in formulating the necessary content, objectives, and exercises based on the content of textbooks and syllabi.

Table 2, the proportion distribution of the test items drawn from the Basic Education Mathematics Content Areas is not significantly greater than 0.75. The calculated proportions over the four years were 0.7, 0.77, 0.7, and 0.4, with only one year (2020) having a proportion exceeding the 0.75 threshold. This variability suggests inconsistency in the alignment of test items with the expected difficulty level over the years. Additionally, the proportion of test items drawn from the curriculum content areas did not meet the desired benchmark of 0.75, particularly in 2022, when the proportion dropped to 0.4, indicating a potential misalignment. The null hypothesis was accepted, indicating that the proportion of test items meeting these criteria is not significantly greater than 0.75. The distribution of item facility indices within the accepted range was found to be independent of the examination years. This findings align with the established guidelines in CTT, which suggest that an acceptable range for item facility is between 0.3 and 0.7.

From Table 3, revealed that there is no significant association between the years and the distribution of the extent of dependence of item facility indices within the accepted range. This aligns with Lord's (1952) study on temporal stability of test item difficulties, which found out that the difficulty and discrimination of test items remained stable across different years of administrations.

Conclusion

This study revealed that mathematics examinations developed by the Imo State Education Development Centre lacked adequate content validity, with some curriculum areas overemphasized while others were underrepresented. Although the quality of items, as measured by facility indices, remained generally stable, the proportion of items within the acceptable range varied across the years. These findings suggest that inconsistencies in test construction may contribute to fluctuations in students' performance. To address this, there is a need for a systematic review of curriculum coverage and test design processes to ensure that examination items consistently reflect the intended content and difficulty levels. The results, however, are limited to BECE mathematics tests in Imo State and may not be generalized to other states or national examinations.

Recommendation

Based on the findings and conclusions, it is recommended that;

1. Teachers, who are the principal actors, should be involved in the examination board to produce standard test instruments that cover both the required contents and objectives while ensuring validity and reliability.
2. Experts in the field of mathematics and measurement and evaluation should be consulted on how to construct tests, putting item facility indices into consideration.
3. Teachers and experts responsible for preparing examinations should receive essential orientations through training on how to construct questions using a table of specification to ensure congruent distribution of test items.

References

Abdullahi, I. (2020). Analyses of psychometric properties of Mathematics Basic Education Certificate Examination Questions (BECEQ) in Gombe State. Association of educational researchers and evaluators in Nigeria. *Unpublished manuscript.*

Abonyi, O. S. (2011). *Instrumentation in behavioural research:* A practical approach. Enugu: Fulladu publishing company.

Agah, M.P. (2020). The relevance of mathematics education in the Nigerian contemporary society: Implication to second education. *Journal of Education, Society and Behavioural Science*, 33(5); 36-43

Alade, M. & Odunsi, O. N. (2023). Multiple-choice Construction Competencies and Items Quality: Evidence from selected Senior High School Subject Teachers in Kosofe Local Government Area, Lagos State, Nigeria. *International Journal of Educational Research*, 12(1),

Anna-Siri, M. F. (2011). The use of item analysis for the improvement of objective examinations. *Procedia - Social and Behavioral Sciences* 29. 188 – 197. https://doi.org/10.1016/j.sbspro.2011.11.224

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.

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

Ighodaro, A. E., & Orheruata, M. U. (2024). Comparison of methods to detect item parameter drift in NABTEB SSCE for 2012-2015 chemistry multiple choice tests using IRT techniques. *Nigerian Journal of Educational Research and Evaluation*, 23; 203-215.

IvyPanda. (2024). The research surveys: Descriptive and analytical. Retrieved from https://ivypanda.com/essays/the-research-surveys-descriptive-analytical/

Lord, F. M. (1952). The relation of the reliability of multiple-choice tests to the distribution of item difficulties. *Psychometrika* 17,181-194. https://doi.org/10.1007/BF02288781

Marrow, R., Rata, E., & Evans, R. (2021). A study of the New Zealand mathematics curriculum. The impact of inconsistent curriculum design on student achievement. *Educational Studies in Mathematics*, 107(3); 311-329. https://arxiv.org/abs/2106.06674

Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of educational research*, 79(4); 1332-1361 https://doi.org/10.3102/0034654309341375

Mohammed (2018). Use of classical test theory to evaluate the quality of 40 multiple choice research methodology examination items in Maryamu Abacha University, Niger republic. *Unpublished institutional report*

Moyinoluwa, T. D. (2015). Analysing the psychometric properties of Mathematics in public examinations in Nigeria. Research on Humanities and Social Sciences, 5(7), 1-8. Retrieved from http://www.iiste.org/journals/index.php/RHSS/article/view/21512/22269

Nathan, T. (2020). What is classical item facility (p-value)? Retrieved from https://assess.com/classical-item-difficulty-p-value/

Nwana, O. C. (2009). *Educational measurement and evaluation*. Nigerian Academy of Education.

Odukoya, J.A., Adekeye, O., Igbinoba, A.O. &Afolabi, A. (2018). Item analysis of university-wide multiple choice objective examinations: the experience of a Nigerian private university. *Quality & Quantity* 52, 983–997. https://doi.org/10.1007/s11135-017-0499-2

Ogunka, R. I., & Amadi, B. D. (2019). Classical test theory estimation of psychometric properties of 50 items multiple – choice English language test in secondary schools in Ikwerre local government area, Rivers state. *ASSEREN Journal of Education.* 4(1); 1-11.

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

Seitz, T. (2017). Curriculum alignment among the intended, enacted, and assessed curricula for grade 9 mathematics. *Journal of Curriculum and Assessment Studies*, 10(2); 45-62. Retrieved from https://jcacs.journal.yorku.ca

Sharma, S.R. (2000). *Modern teaching strategies*. New Delhi: Omsons Publications.

Talebi, A. G., Ghaffari, R., Eskandarzadeh, E., & Oskouei, A. E. (2013). Item Analysis: An effective tool for assessing exam quality, designing appropriate exam and determining weakness in teaching. *Research and Development in Medical Education*, 2(2); 69-72

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.

Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly*, 43(3); 493–513