

Development and Validation of a Multiple-Choice Test in Algebra and Elementary Trigonometry (MTH 112)

Umar Saidu Bashir*, Muktar Mohammed, Mohammed Usman

Department of General Studies, School of Information and General Studies,
Federal Polytechnic Bali, Taraba State, Nigeria

Abstract

This study developed and validated a Multiple-Choice Test for Algebra and Elementary Trigonometry (MTH 112) For Federal Polytechnic Damaturu, Yobe State Nigeria. The population of the study was 300 MTH 112 Engineering Students. A sample of 81 was selected randomly using stratified random sampling. An evaluative research design was adopted. The instrument initially consisted of 77 multiple-choice items developed in line with the MTH 112 course content. The examination was administered to 81 students, while ten (10) experts in Mathematics Education and Psychometrics validated the instrument. Data were analyzed using the Content Validity Ratio (CVR), Kuder-Richardson Formula 20 (KR-20), Standard Error of Measurement (SEM) formula, item discrimination index, and item difficulty index. The findings revealed that 74 out of the 77 items met the acceptable content validity criterion, while three items were discarded, yielding a Scale Content Validity Index of 0.88. The KR-20 reliability coefficient was 0.984, indicating excellent internal consistency. The Standard Error of Measurement was 3.05, suggesting that the examination provides precise estimates of students' true scores with minimal measurement error. Furthermore, the examination recorded a mean discrimination index of 0.817, indicating excellent ability to discriminate between high- and low-performing students. The mean difficulty index was 0.448, showing that the examination possessed an appropriate moderate level of difficulty. The study concluded that the developed MTH 112 examination is a psychometrically sound instrument with high validity, excellent reliability, appropriate difficulty, and strong discriminatory power. The study recommended among others that the developed test be adopted for assessing students' achievement in Algebra and Elementary Trigonometry in Polytechnics to enhance the quality of educational assessment.

Keywords: Psychometric properties, content validity, Reliability, Discrimination Index, Classical Test Theory

1. Introduction

Assessment of students' achievement is central to education quality; it yields crucial information about what students know, how well they understand concepts, and how much of the curriculum has been internalized (Nitko & Brookhart, 2014). In Mathematics, particularly in foundational courses like Algebra and Elementary Trigonometry, robust assessment helps in diagnosing cognitive gaps, guiding instruction, and preparing students for more advanced studies in science, engineering, and technology. In Nigerian federal polytechnics, courses such as MTH 112 help students build on the mathematics they learned in secondary school and develop the quantitative skills needed for their tertiary education.

Despite its importance, mathematics achievement in Nigeria has remained low, particularly in polytechnics and among students in more remote or underserved regions. The North-East geopolitical zone of Nigeria faces particular challenges: insecurity, displacement, limited infrastructure, frequent interruptions in schooling, and scarcity of well-trained staff in some areas (see e.g., Kwami, Kumazeghe & Umar, 2025). These conditions exacerbate difficulties in both teaching and assessment and may result in inconsistent testing practices and weak measurement of learning in mathematics.

One major issue is that the assessment of courses like Algebra and Elementary Trigonometry often relies on locally developed tests that are not standardized or validated. Such tests may lack proper content coverage, may not align

*Corresponding Author: Umar Saidu Bashir, Ph.D

Email Address: jikanbunu@gmail.com | ORCID: <https://orcid.org/0009-0001-2236-0766>

with intended learning outcomes, and may have items with poor psychometric properties (too easy, too hard, ambiguous, or biased). Without systematic validation (including item difficulty, discrimination, internal consistency, etc.), the reliability and fairness of such assessments are questionable.

Recent studies in Nigeria reinforce the urgent need for attention to test development and validation. For instance, Green & Baddom (2024) developed and validated a multiple-choice test for Measurement and Evaluation for NTI NCE Students in the South-South states, establishing item difficulty and discrimination indices and demonstrating acceptable reliability. Blessing (2023) employed Item Response Theory (IRT) in the development and validation of a multiple-choice Economics achievement test designed for Senior Secondary School One (SSS1) students. Also, Olotola, Ihechu & Nuraddeen (2023) examined differential item functioning in BECE mathematics multiple-choice items, revealing gender-based item bias. These works show that while test validation research is active, there is little that directly addresses polytechnic mathematics courses, especially MTH 112, and even less focused on the North-East region.

Furthermore, infrastructural challenges and human resources constraints in the North-East exacerbate assessment issues. According to recent media reports and institutional communiqués, many Federal Polytechnics in North-East Nigeria struggle with poor electricity, inadequate facilities, and limited access to modern assessment and measurement tools, which negatively affect both teaching and assessment quality.

Thus, there is a pressing need to develop and validate a standardized multiple-choice test for Algebra and Elementary Trigonometry (MTH 112) specifically for Federal Polytechnics in the North-East. Such an instrument would ensure that assessment is fair, reliable, and valid; align with learning outcomes; and provide data that can be compared across polytechnics. It would also help in identifying areas where students struggle, guide remedial instruction, and support policy and curricular improvements.

Objectives of the Study

The specific objectives of this study were to:

- i. determine the Content Validity Ratio (CVR) and Scale Content Validity Index of the developed MTH 112 examination;
- ii. determine the internal consistency reliability coefficient (KR-20) of the examination;
- iii. determine the Standard Error of Measurement (SEM) of the examination;
- iv. determine the mean discrimination index of the examination items; and
- v. determine the mean difficulty index of the examination items.

2. Literature Review

2.1 Introduction

This section reviews literature relevant to the development and validation of the multiple-choice test (MCT) for Algebra and Elementary Trigonometry (MTH 112). It is organised around four strands: (i) the conceptual and theoretical framework underpinning test development and validation, (ii) empirical studies on multiple-choice item construction and quality, (iii) empirical studies on the psychometric evaluation of mathematics achievement tests using Classical Test Theory (CTT) and Item Response Theory (IRT), and (iv) studies on learners' difficulties in algebra and trigonometry that motivate the content coverage of the instrument. The review closes with a summary that identifies the gap the present study addresses.

2.2 Conceptual Framework

2.2.1 The Multiple-Choice Item as a Measurement Tool

A multiple-choice item is generally described as consisting of a stem, which presents the problem or question, and a set of response options comprising one correct or best answer (the key) and one or more incorrect but plausible alternatives (distractors). This format remains one of the most widely used and enduring modes of educational assessment because it permits efficient, objective scoring across large numbers of examinees while still allowing the measurement of varied cognitive skills, from recall to application and analysis (Butler, 2018).

Guidance on constructing sound multiple-choice items converges on a common set of principles (Haladyna, Downing, & Rodriguez, 2002; Brame, 2013): stems should be stated clearly and, where possible, positively, avoiding negatively phrased leads such as "except" or "not" (Chiavaroli, 2017); options should be homogeneous in content, grammatically consistent with the stem, and similar in length so that no option is cued by its form rather than its content; and combination options such as "all of the above" or "both A and B" are generally discouraged because they can reward test-wiseness rather than genuine content mastery (Haladyna et al., 2002). Distractors, in particular, are treated as carrying instructional value in their own right: well-constructed distractors are built around plausible misconceptions or common computational errors, so that the pattern of options a cohort selects can itself provide diagnostic information about specific gaps in understanding, rather than functioning merely as "wrong answers" (Brame, 2013). Evidence on the optimal number of options is mixed, but Rodriguez's (2005) meta-analysis of 80 years of item-writing research found that three-option items (one key and two functioning distractors) can perform comparably to the traditional four- or five-option format once non-functioning distractors are identified and removed, since non-functioning distractors add length without adding discrimination.

2.2.2 Classical Test Theory and Item Response Theory

Two complementary psychometric frameworks dominate the validation of achievement tests: Classical Test Theory (CTT) and Item Response Theory (IRT). CTT is built on the notion that an examinee's observed score is the sum of a true score and measurement error, and it yields familiar item-level statistics — difficulty (the proportion of examinees answering an item correctly), discrimination (typically a point-biserial or biserial correlation between item performance and total score), and distractor efficiency — together with test-level reliability estimates such as the Kuder-Richardson Formula 20 (KR-20) for dichotomously scored items (Kuder & Richardson, 1937) or Cronbach's alpha for polytomous or mixed-format instruments (Cronbach, 1951). CTT's appeal lies in its computational simplicity, its modest sample-size requirements, and the fact that it is the only framework of the two that directly evaluates individual distractors.

IRT, in contrast, models the probability of a correct response as a function of both item parameters (difficulty, discrimination, and, in three-parameter models, a pseudo-guessing parameter) and an examinee's underlying ability (theta) on a common latent scale. Because IRT parameters are, in principle, invariant across different examinee samples and different subsets of items, IRT is regarded as offering a more robust basis for tasks such as equating, adaptive testing, and cross-form comparability than CTT, whose item and person statistics are sample-dependent (Ayanwale, 2019; Ayanwale, Chere-Masopha, & Morena, 2022). Comparative studies of mathematics achievement tests have generally found that CTT and IRT converge in classifying items as broadly easy, moderate, or difficult, and as good, fair, or poor discriminators, even though the two frameworks diverge conceptually and can disagree on invariance properties when person ability is estimated from different item subsets (Ayanwale, 2019). This convergence supports the common practice, followed in the present study, of using CTT indices as the primary basis for item retention or revision in test-development studies with moderate sample sizes, while treating IRT as a complementary check where sample size and data structure permit.

2.3 Empirical Studies on the Development and Validation of Mathematics Multiple-Choice Tests

A substantial body of empirical work has applied CTT and IRT to the construction and validation of mathematics achievement tests, much of it situated in Nigerian and comparable West African secondary and tertiary contexts. Such studies typically follow an instrumentation research design: items are drafted against a specified table of specification, trial-tested on a sample of the target population, and then screened using difficulty and discrimination indices, with items falling outside acceptable ranges revised or discarded. Ayanwale's (2019) doctoral study, for instance, developed and validated a dichotomous and polytomous mathematics achievement test on a sample of 1,080 senior secondary students in Osun State, Nigeria, subjecting the data to tests of essential unidimensionality and local dependence before confirming that the constructed items satisfied the assumptions required for IRT calibration.

Ayanwale, Chere-Masopha, and Morena (2022), reviewing the measurement practice of the Examination Council of Lesotho, similarly concluded that while CTT remains attractive for its simplicity, its sample- and test-dependence are limitations that IRT-based approaches can address, and recommended that examining bodies handling multiple-choice, short-response, and constructed-response assessments move toward incorporating modern test theory in item development and scoring. Findings from this stream of research are broadly consistent in showing that IRT-based ability estimates offer greater invariance across item subsets than CTT-based raw or percentage scores, and that IRT

can flag item bias that CTT-based indices alone may not detect. At the same time, the literature notes that IRT's larger sample-size demands and more technical estimation procedures make CTT the more practical first-line tool for test developers working with cohort sizes typical of a single institution or course offering, which is the position adopted for the trial testing phase of the MTH 112 instrument.

2.4 Learners' Difficulties in Algebra and Elementary Trigonometry

Because a test's content validity depends on its items reflecting both the curriculum and the ways learners actually engage with that content, studies on common sources of error in algebra and trigonometry inform the item-writing and content-coverage decisions for the MTH 112 instrument. Usman and Hussaini (2017), analysing the errors of 80 senior secondary students in Zaria Metropolis, Nigeria, using a Mathematics Achievement Test and a Trigonometrical Diagnostic Test interpreted through the Newman Error Hierarchy Model, found that most student errors occurred at the transformation and process-skill stages — that is, students could often set up a trigonometric problem correctly but made errors in manipulating the relevant formula or completing the required algebraic steps. Related work has also linked poor trigonometry performance to specific difficulties with reference angles, angle relationships across quadrants, and conversions between degree and radian measure, and Amedume, Bukari, and Mifetu (2022) found that problem-based instructional approaches could meaningfully improve student achievement in trigonometry, underscoring that these difficulties are amenable to targeted instructional and assessment design.

Algebra is similarly documented as a persistently challenging strand of the Nigerian mathematics curriculum. Anietimfon, Zakariya, Idris, and Falalu (2026), in a quasi-experimental study of 382 Nigerian secondary school students, note that algebra "remains a challenging aspect of the secondary school mathematics curriculum in Nigeria, contributing to persistent learning difficulties and poor achievement," and report that structured, sequenced instruction produced significantly better algebra performance than conventional teaching. Taken together, this literature motivates continued interest in both instructional intervention and improved diagnostic assessment of algebra and trigonometry in Nigerian educational settings.

These findings have direct implications for a multiple-choice test of algebra and elementary trigonometry: distractors modelled on documented misconceptions (for example, common sign errors, incorrect quadrant reasoning, or formula-transformation slips) are more likely to function diagnostically than arbitrary incorrect values, and item coverage should be weighted toward the specific sub-skills — algebraic manipulation, angle and radian conversion, and multi-step trigonometric problem-solving — that the literature identifies as recurrent sources of difficulty.

2.5 Summary and Gap in Literature

The reviewed literature establishes, first, that sound multiple-choice item construction rests on well-documented stem- and distractor-writing principles (Haladyna et al., 2002; Rodriguez, 2005; Chiavaroli, 2017), and second, that CTT and IRT provide complementary, largely convergent tools for evaluating item and test quality once a mathematics MCT has been trial-tested (Ayanwale, 2019; Ayanwale et al., 2022). It further shows that Nigerian and West African studies have applied these frameworks extensively to secondary-school mathematics instruments, and that algebra and trigonometry are consistently documented as areas of persistent learner difficulty that instruments in this domain should be designed to probe (Usman & Hussaini, 2017; Amedume et al., 2022; Anietimfon et al., 2026). However, relatively few studies combine a curriculum-aligned National Diploma (NBTE) course such as MTH 112 with a full psychometric validation reporting both classical item statistics and reliability alongside a discussion of distractor functioning informed by documented learner misconceptions. The present study addresses this gap by developing and validating an MTH 112 multiple-choice test using classical item analysis and reliability estimation, situated within the specific institutional and curricular context of National Diploma engineering mathematics.

3. Methodology

In this section, the researcher discusses the steps that were adopted in the research work, which include research design, population of the study, sample size and sampling, instrument of the study, validation of instrument, reliability of instrument, and method of data analysis techniques.

3.1 Research Design

An evaluative research design was adopted for this study, as it is appropriate for developing and validating a standardised achievement test. The study was carried out among National Diploma (ND I) Engineering students offering

Algebra and Elementary Trigonometry (MTH 112) at Federal Polytechnic Damaturu, Yobe State, North-East Nigeria. The data generated from the examination and the expert ratings were analysed using content validity and internal consistency reliability procedures.

3.2 Population of the Study

The population of this study comprised 300 MTH 112 Engineering students of Federal Polytechnic Damaturu, Yobe State. These students were chosen because they were offering the course, and their responses provided the data required for the psychometric validation carried out in this study.

3.3 Sample and Sampling

The study employed a stratified random sampling technique to select a sample of 81 students from the population of 300 National Diploma (ND I) Engineering students offering Algebra and Elementary Trigonometry (MTH 112) at Federal Polytechnic Damaturu. Stratification was observed to ensure representation across the relevant strata within the class.

A sample size of 81 students was considered adequate for item analysis and psychometric validation using Classical Test Theory (CTT) (DeVellis, 2017; Boone, Staver, & Yale, 2014).

3.4 Research Instrumentation

The main instrument for data collection was a Multiple-Choice Test (MCT) developed by the researcher based on the course outline for MTH 112 (Algebra and Elementary Trigonometry).

The test consisted of four-option multiple-choice items, covering key topics such as:

- i. Basic algebraic operations and equations
- ii. Functions and graphs
- iii. Inequalities and logarithms
- iv. Trigonometric ratios and identities
- v. Trigonometric equations and applications

A Table of Specifications (Test Blueprint) guided item development to ensure balance across content areas and cognitive levels (knowledge, comprehension, application, and analysis).

3.5 Validation of Instruments

Draft test items were reviewed by a panel of ten experts consisting of mathematics lecturers, test developers, and psychometricians from universities and polytechnics.

Their feedback was used to refine ambiguous or irrelevant items and to ensure alignment with course objectives.

3.6 Reliability of the Instruments

Internal consistency reliability was determined using the Kuder–Richardson Formula 20 (KR-20). A coefficient of ≥ 0.70 was taken to indicate that the instrument was reliable for measuring student achievement (Tavakol & Dennick, 2011).

3.7 Method of Data Analysis

Lawshe's (1975) formula for Content Validity Ratio (CVR) was adopted in determining the content validity of the developed MTH 112 examination:

$$CVR = \frac{n_e - N/2}{N/2}$$

Where: CVR = Content Validity Ratio

n_e = number of panelists indicating “essential”

N = total number of panelists.

Ten (10) Mathematics Lecturers from Polytechnics participated in evaluating the relevance of the items in the MTH 112 test. In determining the internal consistency reliability coefficient of MTH 112, the Kuder-Richardson (KR-20) formula was adopted. The formula used is:

$$r_{xx} = \frac{k}{k-1} \left(1 - \frac{\sum pq}{\delta^2} \right)$$

Where:

r_{xx} = reliability coefficient

k = proportion of correct responses to a particular item P = proportion of correct responses to a particular item

q = proportion of incorrect responses to that item (or $1 - P$)

δ^2 = variance of the scores on the test

The standard error of measurement of MTH 112 was calculated as follows:

$$SEM = S_x \sqrt{1 - r_{xx}}$$

SEM=standard error of measurement. S_x = standard deviation of test scores. r_{xx} = reliability coefficient.

Item analysis was conducted using Itean software, a program for item analysis and detailed psychometric reporting based on classical test theory.

4. Results and Discussions

4.1 Objective One: Content Validity Ratio (CVR)

What is the Content Validity Ratio of the developed MTH 112 examination?

The instrument was subjected to expert validation by ten Mathematics educators and psychometric experts. Out of the 77 items initially developed, 74 satisfied the acceptable content validity criterion and were retained, while three items (Items 7, 16 and 40) were discarded. The overall Scale Content Validity Index reported from the expert ratings was **0.88**, indicating high content validity.

Table 4.1 Content Validity Summary

Statistic	Value
Number of Experts	10
Number of Draft Items	77
Items Retained	74
Items Discarded	3
Scale Content Validity Index	0.88

4.2 Objective Two: Coefficient of Internal Consistency

Using the KR-20 formula specified in the proposal, the reliability coefficient is

Table 4.2 Reliability coefficient

Statistic	Value
Number of candidates	81
Number of items	74
KR-20 Reliability	0.9844

Interpretation

A KR-20 coefficient of

$$K R20 = 0.984$$

indicates **excellent internal consistency**.

Using the commonly accepted guideline

- i. ≥ 0.90 = Excellent

- ii. 0.80–0.89 = Very good
- iii. 0.70–0.79 = Acceptable

The developed MTH 112 examination possesses excellent reliability, indicating that the items consistently measure students' achievement in Algebra and Elementary Trigonometry.

4.3 Objective Three: Standard Error of Measurement (SEM)

The proposal specifies

$$SEM = S_x \sqrt{1 - r_{xx}}$$

Where:

S_x = standard deviation of total scores

r_{xx} = KR-20 reliability coefficient.

From the dataset

Standard deviation = 24.415

KR-20 = 0.9844

Therefore,

$$SEM = 24.415 \sqrt{1 - 0.984}$$

$$SEM = 3.08$$

Interpretation

The SEM of 3.05 marks indicates that an examinee's observed score is expected to differ from their true score by approximately ± 3 marks due to measurement error. This relatively small SEM suggests that the developed MTH 112 examination yields precise measurements of students' achievement.

4.4 Objective Four: the mean discrimination index of the developed examination

Using the upper 27% and lower 27% groups (Classical Test Theory), the discrimination indices were computed.

Summary

Table 4.3 Mean Discrimination Index

Statistic	Value
Number of Items	74
Mean Discrimination Index	0.817

According to Ebel's classification:

Table 4.4 Mean Discrimination Index

Discrimination Index	Interpretation
≥ 0.40	Excellent
0.30–0.39	Good
0.20–0.29	Fair
< 0.20	Poor

The obtained mean discrimination index (0.817) indicates that the examination items possess excellent discriminating power. Overall, the items effectively distinguish between high-performing and low-performing students.

4.5 Objective Five: the mean difficulty index

The difficulty index of each item was computed as the proportion of candidates answering the item correctly.

Summary

Statistic	Value
-----------	-------

Number of Items	74
Mean Difficulty Index	0.448

Using the usual interpretation:

Table 4.5 interpretation of Difficulty Index

Difficulty Index	Interpretation
<0.30	Difficult
0.30–0.70	Moderate
>0.70	Easy

The mean difficulty index of 0.448 indicates that, on average, the developed examination has moderate difficulty, which is considered desirable for achievement tests.

4. Discussion of Findings

The findings of this study demonstrate that the developed Multiple-Choice Achievement Test (MTH 112) possesses satisfactory psychometric properties, making it suitable for assessing students' achievement in Algebra and Elementary Trigonometry in Federal Polytechnics. The results are discussed in relation to previous empirical studies.

The content validation exercise yielded a Scale Content Validity Index (S-CVI) of 0.88, indicating a high level of agreement among the expert validators regarding the relevance, clarity, and representativeness of the test items. This finding supports the recommendation of Lamm, K. W. and colleagues (2020), who argued that expert review is fundamental to establishing content validity during instrument development and that high content validity provides evidence that an instrument adequately measures the intended construct. The present study is consistent with their methodological framework, as both studies emphasize systematic expert judgment as an essential stage in instrument validation. However, while Lamm et al. (2020) proposed a comprehensive framework applicable to various educational and social science instruments, the present study extends these recommendations to the development of a standardized mathematics achievement test specifically designed for Federal Polytechnics in North-East Nigeria.

Similarly, the present finding corroborates the work of Green, B. M. (2024), who developed and validated a multiple-choice achievement test for Measurement and Evaluation among NTI-NCE students. Green reported acceptable content validity, appropriate item characteristics, and satisfactory internal consistency reliability, concluding that systematic test development enhances the quality of educational assessment. Likewise, the current study achieved excellent psychometric indices, including a KR-20 reliability coefficient of 0.984, suggesting that the instrument consistently measures students' achievement in Algebra and Elementary Trigonometry. While Green's study focused on educational measurement for teacher education programmes, the present study extends the application of rigorous psychometric procedures to mathematics education within the Nigerian Polytechnic system, thereby filling an important gap in tertiary mathematics assessment.

The exceptionally high KR-20 reliability coefficient (0.984) obtained in this study is also consistent with the observations of Tavakol, Mohsen and Dennick, Reg (2011), who emphasized that internal consistency reliability is a critical indicator of the quality of objective tests. According to these authors, reliability coefficients exceeding 0.90 indicate excellent internal consistency, although excessively high values may also suggest that some items are highly similar or measure nearly identical aspects of the construct. Consequently, while the reliability estimate obtained in the present study demonstrates that the instrument produces dependable scores, future studies may further examine item redundancy through additional psychometric procedures such as Rasch analysis or Item Response Theory to ensure optimal item diversity.

The findings regarding item difficulty and discrimination are equally encouraging. The mean difficulty index of 0.448 indicates that the developed test possesses a moderate level of difficulty, while the mean discrimination index of

0.817 demonstrates that the items effectively distinguish between high-performing and low-performing students. These results agree with the recommendations of Tavakol and Dennick (2011), who maintained that effective educational tests should contain items with moderate difficulty and high discrimination to maximize measurement precision. Compared with their general recommendations for objective test construction, the present instrument demonstrates even stronger discrimination power, indicating that the retained items possess excellent diagnostic capability.

The present findings also support those reported by Blessing, N. (2023), who applied Item Response Theory (IRT) in developing and validating a multiple-choice Economics achievement test for Senior Secondary School students. Blessing reported that systematic psychometric evaluation improved the quality of test items and enhanced the validity of the resulting instrument. Although the present study employed Classical Test Theory (CTT) using KR-20 reliability, item difficulty, and discrimination indices rather than IRT, both studies reached a common conclusion that rigorous psychometric analysis substantially improves the quality of achievement tests. The major difference lies in the analytical framework adopted: Blessing utilized IRT to estimate item parameters independent of sample characteristics, whereas the present study employed Classical Test Theory because of its practicality, relatively modest sample size, and widespread application in educational assessment within Nigerian tertiary institutions.

Furthermore, the relatively small Standard Error of Measurement ($SEM = 3.05$) indicates that the observed test scores closely approximate students' true achievement levels. This finding provides additional evidence of score precision and complements the high reliability coefficient obtained. The result aligns with the assertion of Tavakol and Dennick (2011) that reliable educational instruments should also demonstrate acceptable measurement precision, thereby increasing confidence in decisions based on students' scores.

Overall, the findings indicate that the developed MTH 112 achievement test satisfies essential psychometric criteria, including content validity, internal consistency reliability, measurement precision, item difficulty, and item discrimination. The study therefore contributes to the growing body of literature on educational measurement in Nigeria by providing a standardized mathematics achievement instrument specifically designed for Federal Polytechnics. Unlike previous studies that focused on teacher education, business studies, or secondary school subjects, the present study addresses a relatively under-researched area of mathematics assessment in Nigerian Polytechnic education, thereby providing a valuable resource for lecturers, curriculum developers, examination bodies, and educational researchers.

5. Conclusion

Based on the findings of this study, it is concluded that the developed MTH 112 (Algebra and Elementary Trigonometry) test possesses satisfactory psychometric properties and is suitable for measuring Engineering students' achievement in the course. The content validation process established that the examination adequately represents the instructional objectives and course content, as evidenced by the high Scale Content Validity Index (0.88), with only three items failing to meet the required content validity standard. The examination also demonstrated excellent internal consistency. The test also exhibited excellent discriminating power as well as possession of an appropriate level of difficulty. The study concludes that the developed MTH 112 examination satisfies the essential psychometric requirements of a quality achievement test for Algebra and Elementary Trigonometry in Polytechnics.

6. Recommendations

Based on the findings and conclusions of the study, the following recommendations are made:

1. The developed MTH 112 test should be adopted by Polytechnics for assessing students' achievement in Algebra and Elementary Trigonometry.
2. Polytechnics should establish item banks containing psychometrically sound examination questions that have been validated and tested for reliability, difficulty, and discrimination indices before use in semester examinations.

Conflicts of Interest:

The authors declare that there is no conflict of interest associated with this study. Although the research was financially supported by the Tertiary Education Trust Fund (TETFund), the funder had no influence on the study design, data collection, analysis, interpretation of the findings, or the decision to publish the results.

Reference

- Amedume, L. K., Bukari, H. I., & Mifetu, R. K. (2022). Effect of problem-based learning approach on students' achievement in trigonometry. *International Journal of Trends in Mathematics Education Research*, 5(4), 348-357.
- Anietimfon, E. W., Zakariya, Y. F., Idris, I. O., & Falalu, M. K. (2026). Effects of Gagné's instructional model on performance and retention in Algebra among secondary school students in Nigeria. *Frontiers in Education*, 11, Article 1747651. <https://doi.org/10.3389/educ.2026.1747651>
- Ayanwale, M. A. (2019). *Efficacy of item response theory in the validation and score ranking of dichotomous and polytomous response mathematics achievement tests in Osun State, Nigeria* [Doctoral thesis, Institute of Education, University of Ibadan].
- Ayanwale, M. A., Sanusi, I. T., Adelana, O. P., Aruleba, K. D., & Oyelere, S. S. (2022). Teachers' readiness and intention to teach artificial intelligence in schools. *Computers and Education: Artificial Intelligence*, 3, 100099.
- Blessing, N. (2023). *Application of item response theory in the development and validation of multiple-choice test in Economics for Senior Secondary School One students* (Unpublished master's dissertation). Michael Okpara University of Agriculture, Umudike.
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences* (Vol. 10, pp. 978-94). Dordrecht: Springer.
- Brookhart, S. M. (2014). *How to design questions and tasks to assess student thinking*. ASCD.
- Brame, C. J. (2013). *Writing good multiple choice test questions*. Vanderbilt University Center for Teaching. <https://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions/>
- Butler, A. C. (2018). Multiple-choice testing in education: Are the best practices for assessment also good for learning? *Journal of Applied Research in Memory and Cognition*, 7(3), 323-331. <https://doi.org/10.1016/j.jar-mac.2018.07.002>
- Chiavaroli, N. (2017). Negatively-worded multiple choice questions: An avoidable threat to validity. *Practical Assessment, Research, and Evaluation*, 22, Article 3.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Sage Publications.
- Essien, A. A., & Moleko, M. (2025). Research on language and multilingualism in mathematics education in Sub-Saharan Africa. *Africa Education Review*, 22(1), 1-25.
- Green, B. M. (2024). *Development and validation of measurement and evaluation multiple-choice achievement test for NTI-NCE students in South-South States, Nigeria* (Unpublished master's dissertation). Michael Okpara University of Agriculture, Umudike.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-334.
- Kayode, D. I. (2024). Development and validation of multiple-choice test items for Junior Secondary School Business Studies curriculum. *Journal of Educational Studies Trends & Practice*, 4(8), 40-50
- Kwami, Y. A., Kumazeghe, S. Z., & Umar, N. (2023). Effect of problem-based learning on the academic performance and interest in welding and fabrication among Nigeria Certificate in Education (NCE) Mechanical Technology Education students in North-Eastern Nigeria. *ATBU Journal of Science, Technology and Education*, 11(2), 72-80.
- Kuder, G. F., & Richardson, M. W. (1937). The theory of estimation of test reliability. *Psychometrika*, 2(3), 151-160.

- Lamm, K. W., Lamm, A. J., & Edgar, D. (2020). Scale development and validation: Methodology and recommendations. *Journal of International Agricultural and Extension Education*, 27(2), 24–35. <https://doi.org/10.5191/jiaee.2020.27224>
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Nitko, A. J., & Brookhart, S. M. (2014). *Educational assessment of students* (7th ed.). Pearson.
- Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, 24(2), 3–13.
- Tavakol, M., & Dennick, R. (2011). Post-examination analysis of objective tests. *Medical Teacher*, 33(6), 447–458. <https://doi.org/10.3109/0142159X.2011.564682>
- Usman, M. H., & Hussaini, M. M. (2017). Analysis of students' error in learning of trigonometry among senior secondary school students in Zaria Metropolis, Nigeria. *IOSR Journal of Mathematics*, 13(2), 1–4.