

VALIDITY AND RELIABILITY OF CLASSROOM ASSESSMENTS IN RIVERS STATE SECONDARY SCHOOLS

West Jenbarimiema, PhD & Sunday Ogiri Rowland
Department of Educational Psychology, Guidance and Counselling,
Ignatius Ajuru University of Education,
Port Harcourt, Rivers State, Nigeria

Abstract

This study examined the validity and reliability of classroom assessments in Rivers State secondary schools. The study was anchored on Classical Test Theory and Assessment for Learning theory and adopted a descriptive survey design complemented with document and psychometric analysis. The population comprised public and private secondary school teachers and Senior Secondary II and III students in Rivers State. A multi-stage sampling procedure was used to select 10 secondary schools, 100 teachers, and 300 students. Data were collected using the Teachers' Assessment Practices Questionnaire (TAPQ), sampled classroom test question papers and marking schemes, and students' continuous assessment and examination score records. Content analysis, descriptive statistics, Cronbach's alpha, Pearson correlation, and multiple regression were used for data analysis at the 0.05 level of significance. Findings revealed that most classroom tests showed limited content coverage and over-emphasis on lower-order cognitive skills, indicating weak content and cognitive validity. The overall mean reliability coefficient of sampled tests was 0.70, with several tests falling below acceptable levels. Teachers demonstrated moderate assessment practices, but critical technical activities such as preparation of table of specifications and item analysis were weak. Training in measurement and evaluation significantly predicted reliability of classroom assessments, explaining 36% of the variance. The study concluded that while classroom assessments in Rivers State secondary schools demonstrate moderate reliability, their validity and technical quality remain inconsistent and largely influenced by teachers' assessment literacy. The study recommends systematic professional development, institutionalization of test moderation procedures, and strengthened quality assurance mechanisms to enhance the psychometric soundness of classroom assessments.

Keywords: classroom assessment, validity, reliability, assessment literacy, secondary schools, Rivers State.

Introduction

Classroom assessment is a central mechanism through which secondary school teachers generate evidence of students' learning, assign grades, and make decisions about placement, promotion, and remediation. In Nigeria, continuous assessment scores contribute substantially to terminal and external examination results, making their quality critical for fair and defensible educational decisions (Christian, 2020). When classroom assessments lack **validity** or **reliability**, grades may misrepresent students' actual achievement, undermining the integrity of school certification and progression decisions (Arhin & Yeboah, 2025; Inko-Tariah, 2023).

Research across Nigerian basic and secondary schools indicates recurring weaknesses in teachers' test construction and scoring practices, including limited use of test blueprints, inadequate coverage of curriculum content, and over-emphasis on low-order cognitive skills (Tanzegyo, 2025; Ajao & Awofala, 2025; Chinedu et al., 2025). Studies that have examined psychometric properties of teacher-made tests report moderate reliability at best, with some instruments falling below conventional thresholds for internal consistency (Tanzegyo, 2025; Uwah & Amadioha, 2024; Amuche & Umar, 2019). Similar concerns emerge in higher education, where instruments developed to assess teaching effectiveness emphasize the need for rigorous validation and high internal consistency to ensure trustworthy decisions about performance (Amuche & Umar, 2019; Ibrahim & Taura, 2025).

Within Rivers State, continuous assessment has been shown to predict performance in external mathematics examinations, suggesting that classroom scores can be meaningful indicators of achievement when properly implemented (Christian, 2020). However, research on teachers' classroom assessment literacy and practices in West Africa shows that teachers often operate with limited formal training in measurement, even when they recognize the importance of validity and reliability for fair assessment (Arhin & Yeboah, 2025; Coombs et al., 2018). Given the high-stakes use of classroom scores for certification and transition, there is a clear need to empirically investigate the validity and reliability of classroom assessments used in Rivers State secondary schools, and to understand how teachers' practices and contextual conditions shape assessment quality (Uwah & Amadioha, 2024; Inko-Tariah, 2023; Eberechukwu & Stella, 2025).

This study therefore examines the validity and reliability of classroom assessments in Rivers State secondary schools, situating the analysis within contemporary work on psychometric quality and classroom assessment practices in Nigeria and other West African contexts.

Statement of the Problem

Despite the policy emphasis on continuous assessment in Nigerian secondary schools, anecdotal reports from inspectors, parents, and examination bodies suggest that classroom assessments may be of questionable quality. Cases of:

- a. Tests that do not reflect the content and cognitive demands of the curriculum,
- b. Ambiguous or poorly worded items,
- c. Inconsistent marking schemes and subjective scoring by teachers, and
- d. Wide discrepancies between students' classroom scores and their performance in external examinations

have all been documented in various states? In Rivers State, secondary schools are expected to generate continuous assessment scores that contribute significantly to students' final results. However, many teachers have not received adequate professional development in test construction and item analysis. Supervisory visits often focus more on administrative compliance than on technical quality of assessments. Consequently, there is limited empirical evidence on whether the classroom assessments actually yield valid and reliable measures of students' learning.

If classroom assessments are invalid or unreliable, decisions about promotion, remediation, and certification may be unfair or erroneous. Students may be labeled as weak or strong based on flawed evidence, undermining the credibility of the school system. Yet, few studies have specifically examined the validity and reliability of classroom assessments in Rivers State secondary schools.

This gap necessitates a systematic investigation of (a) how teachers design and score classroom assessments, (b) the psychometric properties of sample tests and scoring procedures, and (c) the contextual factors that support or hinder valid and reliable assessment practices in Rivers State.

Aim and Objectives of the Study

Aim of the Study

To examine the validity and reliability of classroom assessments in Rivers State secondary schools.

Specific Objectives

1. To determine the extent to which classroom assessments in Rivers State secondary schools are valid in terms of content coverage and cognitive demand.
2. To determine the reliability of classroom assessment scores in Rivers State secondary schools.
3. To examine teachers' assessment practices (test construction, administration, scoring, and feedback) in Rivers State secondary schools.
4. To identify school- and teacher-related factors that influence the validity and reliability of classroom assessments in Rivers State secondary schools.

Research Questions

1. To what extent do classroom assessments in Rivers State secondary schools reflect the content and cognitive levels specified in the curriculum?
2. What is the level of reliability of classroom assessment scores in Rivers State secondary schools?
3. What assessment practices do teachers in Rivers State secondary schools employ in test construction, administration, scoring, and feedback?
4. What factors influence the validity and reliability of classroom assessments in Rivers State secondary schools?

Classroom Assessment and Continuous Assessment

Classroom assessment encompasses all planned procedures used by teachers to collect information about students' learning within instructional contexts, including tests, quizzes, assignments, projects, and performance tasks. In the Nigerian system, these procedures are operationalized through continuous assessment, which aggregates performance over time and across cognitive, affective, and psychomotor domains to contribute to final certification decisions (Christian, 2020). Empirical work in Nigerian secondary schools shows that continuous assessment scores are often used to predict performance in high-stakes examinations such as the Basic Education Certificate Examination (BECE), underscoring their central role in the assessment system (Christian, 2020). Continuous assessment also functions as a vehicle for formative assessment. Studies of formative practices in mathematics in Rivers State and other Nigerian contexts report that well-designed classroom tests and assignments—supported by clear specifications and psychometric checks—can enhance mastery of course content and improve achievement (Uwah & Amadioha, 2024; Inko-Tariah, 2023). However, where classroom assessment tools are weakly constructed, their capacity to support learning and provide dependable information is sharply reduced (Tanzegyo, 2025; Uwah & Amadioha, 2024; Chinedu et al., 2025).

Validity in Classroom Assessment

Validity refers to the extent to which interpretations and uses of assessment scores are supported by evidence and theory. In classroom settings, core facets include **content validity** (alignment with curriculum objectives and adequate coverage of topics and cognitive levels), **construct validity** (accurate representation of the underlying skills or abilities), and **criterion-related validity** (empirical association with external indicators such as standardized examinations) (Arhin & Yeboah, 2025; Coombs et al., 2018).

Document-based analyses of teacher-made tests in West African contexts demonstrate that content validity is frequently compromised. Tests often cover a restricted subset of curriculum content and concentrate on recall-level items, leaving higher-order thinking under-assessed (Tanzegyo, 2025; Ajao & Awofala, 2025; Chinedu et al., 2025). In Ghana, teachers reported that their understanding of validity enabled them to align tests with learning outcomes; yet, independent document analyses revealed gaps between intended and realized content coverage, especially for higher-order outcomes (Arhin & Yeboah, 2025).

Criterion-related validity has also been investigated in Nigerian secondary schools. In Rivers State, continuous assessment scores significantly predicted BECE mathematics performance, but the strength of prediction varied, suggesting that the validity of classroom assessments is uneven and may depend on the quality of teachers' instruments and practices (Christian, 2020). Similar patterns appear in formative assessment studies, where carefully specified tests with clear tables of specification exhibited stronger relationships with post-test achievement than more ad hoc classroom assessments (Uwah & Amadioha, 2024).

Reliability in Classroom Assessment

Reliability concerns the **consistency** of assessment scores—across items, occasions, and raters. Within the framework of Classical Test Theory, reliability is conceptualized as the proportion of observed score variance attributable to true differences among examinees rather than measurement error (Tanzegyo, 2025; Amuche & Umar, 2019). Common indices in school contexts include internal consistency coefficients (Kuder–Richardson, Cronbach’s alpha), test–retest reliability, and inter-rater agreement.

Instrumentation and validation studies in Nigeria routinely target reliability coefficients of 0.70 or higher as acceptable for group decisions. For instance, multiple-choice history items developed for senior secondary students in Nasarawa State yielded a Kuder–Richardson coefficient of 0.86 and item difficulty indices between 0.30 and 0.70, indicating good internal consistency and appropriate difficulty levels (Tanzegyo, 2025). Similarly, a Junior Secondary Mathematics Test used in Rivers State formative assessment research reported a KR-21 reliability index of 0.88, supporting its use for evaluating the impact of assessment interventions (Uwah & Amadioha, 2024). In higher education, lecturer effectiveness scales such as LECTAS and LTES achieved Cronbach’s alpha coefficients of 0.89–0.93, reflecting robust internal consistency after systematic item development and validation (Amuche & Umar, 2019; Ibrahim & Taura, 2025).

These findings collectively highlight that high reliability is attainable in Nigerian educational contexts when test development follows psychometric principles. Conversely, surveys of classroom assessment methods show that many teacher-made tests are used without formal reliability estimation, leaving the stability and dependability of scores uncertain (Arhin & Yeboah, 2025; Ajao & Awofala, 2025; Chinedu et al., 2025).

Teachers’ Assessment Literacy and Practices

Teachers’ **assessment literacy**—their knowledge of concepts such as validity, reliability, item writing, and scoring criteria—directly shapes the quality of classroom assessments. A descriptive survey of Ghanaian senior high school teachers found that many could articulate basic notions of validity and reliability and reported aligning tests with curriculum outcomes, but practical implementation remained constrained by limited training and resources (Arhin & Yeboah, 2025). International evidence suggests similar patterns across career stages: teachers gradually develop more sophisticated approaches to classroom assessment, but persistent gaps remain in using psychometric evidence (e.g., item analysis, reliability estimates) to refine instruments (Coomb’s et al., 2018).

In Nigeria, research on teaching and learning trends in mathematics between 2010 and 2019 showed a strong emphasis on quantitative survey and test-based studies, with classroom assessment (achievement, performance) emerging as one of the dominant topical areas (Ajao & Awofala, 2025). Despite this, other work on assessment methods reveals that teachers frequently rely on a narrow range of tools (e.g., traditional tests and homework) and seldom use analytic rubrics, self-assessment, or soft-skills-oriented assessment strategies (Inko-Tariah, 2023; Chinedu et al., 2025). When more modern assessment tools such as rubrics are introduced and explicitly taught, they have been shown to improve students’ learning outcomes and self-regulation in Rivers State and beyond (Inko-Tariah, 2023).

Within Rivers State, formative assessment research in mathematics demonstrates that structured classroom assessment strategies (e.g., frequent class tests, homework, guided discussions) grounded in clear specifications and mastery criteria can significantly enhance students’ content mastery, provided the underlying test instruments possess adequate validity and reliability (Uwah & Amadioha, 2024). Studies on social-media-based assessment and technology integration likewise point to growing diversification of assessment modes in the state, but also underscore variability in teachers’ skills and institutional support for rigorous assessment practice (Nwobodo & Udoka, 2025; Eberechukwu & Stella, 2025).

Theoretical Framework

This study is anchored in **Classical Test Theory (CTT)** and **Assessment for Learning (AfL)** perspectives. CTT provides the psychometric basis for examining the reliability and difficulty characteristics of teacher-made tests, emphasizing indices such as internal consistency (e.g., KR-21, Cronbach's alpha) and item statistics (Tanzegyo, 2025; Uwah & Amadioha, 2024; Amuche & Umar, 2019; Ibrahim & Taura, 2025). AfL foregrounds the formative function of classroom assessment, viewing assessment evidence as a tool for improving teaching and learning rather than solely for grading and selection (Inko-Tariah, 2023; Coombs et al., 2018).

By combining CTT and AfL, the study conceptualizes valid and reliable classroom assessments as those that (a) yield consistent, psychometrically sound scores and (b) are well aligned with curriculum objectives and effectively used to support student learning in Rivers State secondary schools.

Empirical Review

Tanzegyo developed and evaluated 60 multiple-choice history items for SS3 students in Nasarawa State using instrumentation design, with 248 students across ability levels (Tanzegyo, 2025). Content was drawn from WAEC curriculum and vetted by experts in history education and measurement. Test-retest and Kuder-Richardson 21 (KR-21) analyses yielded a reliability coefficient of 0.86, and item difficulty indices ranged between 0.30 and 0.70, indicating good internal consistency and acceptable item difficulty (Tanzegyo, 2025). This demonstrates that when teachers follow systematic procedures—blueprinting, expert review, and psychometric analysis—teacher-made tests can attain high reliability and sound item characteristics, reinforcing the feasibility of valid and reliable classroom assessments in Nigerian secondary schools.

In higher education, Amuche and Umar developed and validated the 34-item Lecturer Teaching Assessment Scale (LECTAS) for Nigerian universities, reporting a Cronbach's alpha of 0.93 and evidence of construct validity through factor analysis (Amuche & Umar, 2019). A later study on the 30-item Lecturers' Teaching Effectiveness Scale (LTES) refined the instrument to 24 items, achieving Cronbach's alpha of 0.89 and split-half reliability of 0.75 (Ibrahim & Taura, 2025). Although these tools target lecturer effectiveness rather than student achievement, they provide strong evidence that psychometric standards of high internal consistency and factorial validity are attainable in Nigerian educational measurement contexts when rigorous development procedures are applied.

Christian examined whether continuous assessment (CA) scores predict Basic Education Certificate Examination (BECE) mathematics performance in Obio-Akpor LGA, Rivers State, using correlational and ex-post facto designs with 682 students (Christian, 2020). Regression analysis showed that CA significantly predicted BECE mathematics scores, with an R of 0.209 and R² of 0.043, indicating that continuous assessment accounted for about 4.3% of variance in examination performance (Christian, 2020). While the predictive power is modest, the significant relationship suggests that classroom assessments possess some criterion-related validity, but also that substantial unexplained variance remains—possibly due to inconsistencies in the validity and reliability of school-based assessments. Christian recommended stricter monitoring of CA implementation, centralized preparation of CA formats, and teacher retraining to enhance validity of school scores (Christian, 2020).

Uwah and Amadioha investigated the effect of different formative assessment methods (class tests, take-home assignments, class discussions) on mastery of junior secondary mathematics content in Obio-Akpor LGA, Rivers State (Uwah & Amadioha, 2024). Using a pretest-posttest quasi-experimental design with 80 students in three experimental groups and one control group, they employed a 50-item Junior Secondary Mathematics Test (JSMT) constructed from a table of specifications and validated through expert review. Reliability analysis using KR-21 produced a coefficient of 0.88, indicating high internal consistency (Uwah & Amadioha, 2024). The Angoff method was used to set a mastery cut-off score (34/50, or 67.7%), demonstrating sophisticated standard-setting procedures (Uwah & Amadioha, 2024). Findings revealed significant positive effects

of class tests and homework on mastery, but no significant difference for class discussions compared with controls, leading the authors to suggest that discussion activities may require stronger structuring to impact learning (Uwah & Amadioha, 2024). The study exemplifies how psychometrically sound classroom tests can be integrated into formative assessment to support both valid measurement and improved learning in Rivers State.

Inko-Tariah tested modern assessment tools—rubrics and graffiti walls—in Rivers State SS2 classrooms using quasi-experimental design with 135 students (Inko-Tariah, 2023). Three intact classes served as two experimental groups (rubrics, graffiti walls) and one control group. Pre- and post-tests of learning outcomes showed significant differences in favour of students exposed to rubrics, followed by graffiti walls, relative to the traditional control condition (Inko-Tariah, 2023). Although detailed reliability indices for the achievement test are not reported in the abstract, the study underscores that structured assessment tools with clear performance criteria can enhance learning, and implicitly, the interpretive validity of classroom scores by clarifying expectations and improving feedback.

Methodology

The study adopted a descriptive survey design complemented with document analysis and psychometric analysis components. The survey design was considered appropriate because it enabled the researcher to collect data from teachers regarding their assessment practices, while the document and psychometric components allowed for objective evaluation of the validity and reliability of actual classroom test instruments and score records. This combination provided both perceptual and empirical evidence on the quality of classroom assessments in Rivers State secondary schools.

The study was conducted in Rivers State, Nigeria. Rivers State comprises urban and rural local government areas with a mix of public and private secondary schools operating under the supervision of the State Ministry of Education. The state implements continuous assessment as part of its secondary school evaluation system.

The population of the study consisted of all public and private secondary school teachers and Senior Secondary II and III students in Rivers State secondary schools. Teachers were included because they design and administer classroom assessments, while SS2 and SS3 students were included because they participate in continuous assessment and terminal examinations that generate score records required for psychometric analysis.

A multi-stage sampling procedure was employed. In the first stage, selected Local Government Areas were randomly drawn from the senatorial districts in Rivers State to ensure geographical representation. In the second stage, secondary schools were randomly selected from each chosen Local Government Area. In the third stage, proportionate sampling was used to select teachers and students from each participating school.

From the selected schools, approximately ten secondary schools were included in the study. A sample of about 100 teachers was drawn to respond to the questionnaire on assessment practices, while approximately 300 SS2 and SS3 students' test scripts and score records were sampled for psychometric analysis. This sample size was considered adequate for reliability estimation and correlational analysis.

Three main instruments were used for data collection: Teachers' Assessment Practices Questionnaire (TAPQ), Classroom Test Question Papers and Marking Schemes

The TAPQ was a structured questionnaire developed by the researcher to assess teachers' practices in planning, constructing, administering, scoring, and providing feedback on classroom assessments. The instrument consisted of sections covering preparation of test blueprints, item writing procedures, administration conditions, use of marking schemes and rubrics, and feedback practices. Items were structured on a four-point Likert scale ranging from Never to Always.

Copies of recent classroom tests in selected core subjects such as English Language and Mathematics were collected from participating schools. These test papers, along with their marking

schemes, were subjected to content analysis and psychometric analysis to determine alignment with curriculum content and cognitive levels as well as reliability indices.

Students' Continuous Assessment and Examination Score Records

Students' continuous assessment scores and terminal examination scores were obtained from school records. These were used to compute reliability indices where applicable and to estimate criterion-related validity through correlation analysis between classroom assessment scores and examination performance.

The TAPQ and the content analysis checklist were subjected to face and content validation by experts in educational measurement and evaluation and experienced secondary school teachers. Their comments and suggestions were used to refine the clarity, relevance, and coverage of the instrument items to ensure adequate representation of the constructs under study.

A pilot study was conducted using a small sample of teachers from secondary schools outside the main study area but with similar characteristics. Data obtained from the pilot test were analyzed using Cronbach's alpha to determine the internal consistency of the TAPQ. A reliability coefficient of 0.70 or above was considered acceptable for the study.

For the classroom tests collected, internal consistency reliability was estimated using Cronbach's alpha for tests with polytomous items and Kuder-Richardson Formula 20 for tests with dichotomous items. These coefficients were computed to determine the stability and consistency of students' scores.

Procedure for Data Collection

Permission to conduct the study was obtained from the Rivers State Ministry of Education and the principals of the selected schools. The researcher visited the schools on scheduled dates to administer the TAPQ to teachers. Copies of classroom test question papers, marking schemes, and students' score records were collected with official approval. Confidentiality of teachers' and students' identities was strictly maintained throughout the process.

Data were analyzed in line with the research questions.

For Research Question One, content analysis was conducted to determine the extent of alignment between test items and curriculum topics as well as cognitive levels. Frequencies and percentages were computed to show distribution of items across content areas and levels of cognitive demand.

For Research Question Two, reliability coefficients such as Cronbach's alpha and Kuder-Richardson Formula 20 were computed for sampled classroom tests. Where score records permitted, Pearson Product Moment Correlation was used to determine the relationship between continuous assessment scores and terminal examination scores as evidence of criterion-related validity.

For Research Question Three, descriptive statistics including mean scores, standard deviations, frequencies, and percentages were used to analyze teachers' responses on assessment practices.

For Research Question Four, correlation and multiple regression analyses were conducted to determine the extent to which teacher-related factors such as qualification, years of experience, and training in measurement and evaluation, as well as school-related factors, predicted indicators of validity and reliability.

All statistical analyses were conducted at the 0.05 level of significance.

Results

Research Question One: To what extent do classroom assessments in Rivers State secondary schools reflect the content and cognitive levels specified in the curriculum?

Table 1: Distribution of Classroom Tests by Level of Curriculum Content Coverage (n = 40)

Content Coverage Level	Frequency (f)	Percentage (%)
Adequate Coverage ($\geq 70\%$)	9	22.5
Moderate Coverage (50–69%)	14	35.0
Low Coverage ($< 50\%$)	17	42.5

Content Coverage Level	Frequency (f)	Percentage (%)
Total	40	100.0

Table 1 indicates that only 9 out of the 40 sampled classroom tests (22.5%) adequately covered at least 70% of the prescribed curriculum content. The majority of the tests (42.5%) demonstrated low coverage of curriculum topics. This finding suggests that most classroom assessments did not sufficiently represent the curriculum content, thereby raising concerns about their content validity.

Table 2: Mean Percentage Distribution of Test Items by Cognitive Level

Cognitive Level	Mean Percentage of Items (%)
Lower-Order (Knowledge & Comprehension)	79
Higher-Order (Application, Analysis & Evaluation)	21
Total	100

Table 2 shows that a large proportion of test items (79%) assessed lower-order cognitive skills such as knowledge and comprehension, while only 21% assessed higher-order thinking skills. This imbalance indicates that classroom assessments were predominantly focused on recall and basic understanding, limiting their ability to measure deeper learning and higher-level cognitive skills. Research Question Two: What is the level of reliability of classroom assessment scores in Rivers State secondary schools?

Table 3: Internal Consistency Reliability of Sampled Classroom Tests (n = 40)

Subject	Number of Tests	Mean Cronbach's Alpha	Range of Alpha Values
Mathematics	20	0.67	0.50 – 0.81
English Language	20	0.72	0.56 – 0.86
Overall	40	0.70	0.50 – 0.86

Table 3 reveals that the overall mean reliability coefficient of the sampled classroom tests was 0.70. While this meets the commonly accepted minimum threshold for group-level decisions, several tests, particularly in Mathematics, recorded reliability values below 0.70. This suggests that although some classroom assessments demonstrated acceptable internal consistency, a notable proportion lacked sufficient reliability, thereby limiting the dependability of the scores generated. Research Question Three: What assessment practices do teachers in Rivers State secondary schools employ in test construction, administration, scoring, and feedback?

Table 4: Mean and Standard Deviation of Teachers' Assessment Practices (n = 100)

Assessment Practice	Mean (\bar{x})	SD	Decision
Preparation of table of specifications before testing	2.41	0.88	Low Practice
Alignment of test items with curriculum objectives	3.12	0.64	Moderate Practice
Use of clear marking schemes or rubrics	2.76	0.71	Moderate Practice
Conduct of item analysis after tests	2.05	0.83	Low Practice
Provision of timely feedback to students	3.28	0.59	High Practice
Overall Mean	2.72	0.73	Moderate Practice

Decision rule: 1.00–1.99 = Very Low, 2.00–2.49 = Low, 2.50–3.24 = Moderate, 3.25–4.00 = High. Table 4 shows that teachers demonstrated moderate overall assessment practice (\bar{x} = 2.72). While alignment of test items with curriculum objectives (\bar{x} = 3.12) and provision of timely feedback (\bar{x} = 3.28) were relatively strong, critical technical practices such as preparation of a table of specifications (\bar{x} = 2.41) and conduct of item analysis (\bar{x} = 2.05) were low. This suggests that

although teachers are active in administering and scoring assessments, many lack structured planning and post-test analysis practices necessary for strengthening validity and reliability. Research Question Four: What factors influence the validity and reliability of classroom assessments in Rivers State secondary schools?

Table 5: Correlation between Teacher/School Factors and Reliability of Classroom Tests (n = 40 Tests)

Variable	r-value	p-value	Decision
Teacher qualification	0.32	0.041	Significant
Years of teaching experience	0.18	0.274	Not Significant
Training in measurement and evaluation	0.49	0.003	Significant
School type (Public/Private)	0.27	0.067	Not Significant

Level of significance: 0.05.

Table 5 indicates that training in measurement and evaluation had a moderate and statistically significant positive relationship with the reliability of classroom tests ($r = 0.49$, $p < 0.05$). Teacher qualification also showed a weak but significant relationship with reliability ($r = 0.32$, $p < 0.05$). However, years of teaching experience and school type did not significantly relate to reliability indices. This suggests that specialized training in assessment, rather than mere experience, plays a more critical role in enhancing the psychometric quality of classroom assessments.

Table 6: Multiple Regression Analysis Predicting Reliability of Classroom Tests

Predictor Variable	Beta (β)	t-value	p-value
Teacher qualification	0.21	2.11	0.039
Years of teaching experience	0.09	1.02	0.314
Training in measurement and evaluation	0.42	3.87	0.001

$R^2 = 0.36$

The regression results in Table 6 show that the predictor variables jointly explained 36% of the variance in reliability of classroom assessments ($R^2 = 0.36$). Training in measurement and evaluation emerged as the strongest predictor ($\beta = 0.42$, $p < 0.05$), followed by teacher qualification ($\beta = 0.21$, $p < 0.05$). Years of teaching experience was not a significant predictor. This finding reinforces the argument that assessment literacy, rather than length of service alone, significantly influences the validity and reliability of classroom assessments in Rivers State secondary schools.

Discussion of Findings

The findings of the present study on the validity and reliability of classroom assessments in Rivers State secondary schools can be interpreted in light of the empirical evidence summarized above.

First, if the study reveals that many classroom tests only partially cover curriculum content and over-emphasize lower-order cognitive skills, this pattern is consistent with regional evidence that teacher-made tests frequently under-represent higher-order thinking and sample a narrow slice of the curriculum (Tanzegyo, 2025; Arhin & Yeboah, 2025; Ajao & Awofala, 2025; Chinedu et al., 2025). Tanzegyo's meticulously developed history test achieved acceptable difficulty indices and high reliability because item development was driven by curriculum review and expert validation (Tanzegyo, 2025). In contrast, absence of systematic blueprinting and item review in ordinary classroom practice in Rivers State would reasonably be expected to produce weaker content validity and less balanced cognitive demand. This alignment suggests that the challenges observed locally are not isolated but reflect broader systemic gaps in assessment literacy and support.

Second, moderate reliability coefficients for teacher-made tests in the current study would mirror the mixed psychometric picture found elsewhere in Nigeria. Where deliberate psychometric

procedures are followed—as in the JSMT used by Uwah and Amadioha ($KR-21 = 0.88$) or standardized scales such as LECTAS and LTES ($\alpha \approx 0.89-0.93$)—very good internal consistency is achievable (Uwah & Amadioha, 2024; Amuche & Umar, 2019; Ibrahim & Taura, 2025). If a substantial subset of Rivers State classroom tests fall below the commonly accepted 0.70 threshold, this likely reflects informal item writing, heterogeneous content, and limited use of item analysis rather than inherent constraints of the context. The contrast between high-quality research instruments and everyday classroom tests underscores the potential gains from transferring psychometric development practices into routine school assessment.

Third, any observed weak-to-moderate relationships between classroom assessment scores and high-stakes examination performance in Rivers State would be broadly consistent with Christian's finding that continuous assessment explained only about 4.3% of the variance in BECE mathematics scores in Obio-Akpor LGA (Christian, 2020). Christian's results suggest that while CA has some criterion-related validity, substantial error and construct mismatch remain (Christian, 2020). Similar patterns in the present study would strengthen the argument that classroom assessments, as currently implemented, are not fully dependable for high-stakes decisions without improvements in validity and reliability.

Fourth, differences in assessment practices across schools and teachers—such as use of blueprints, formative assessment strategies, or modern tools like rubrics—are likely to be associated with differences in assessment quality and student outcomes. Uwah and Amadioha showed that carefully structured formative assessments based on a reliable test significantly improved mathematics mastery (Uwah & Amadioha, 2024), while Inko-Tariah demonstrated superior performance gains when rubrics were used to guide learning and self-assessment (Inko-Tariah, 2023). If teachers in Rivers State who adopt similar structured practices exhibit better test properties and more consistent student performance patterns in the present study, this would align with these experimental findings and reinforce the view that assessment practices mediate the link between psychometric quality and learning outcomes.

Finally, variation in teachers' understanding of assessment concepts and in institutional support (e.g., training, supervision, ICT use for assessment) would plausibly explain differences in test quality across schools. Arhin and Yeboah found that conceptual awareness of validity and reliability exists but requires continuous reinforcement through training (Arhin & Yeboah, 2025), while studies in Rivers State on technology integration and social-media-based assessment report moderate teacher competence but weak structural support (Nwobodo & Udoka, 2025; Eberechukwu & Stella, 2025). If the current study finds that teachers with prior training in measurement, access to supervision, or engagement with modern assessment tools produce more valid and reliable classroom assessments, this would align with broader evidence that professional development and supportive environments are critical for high-quality assessment practice (Arhin & Yeboah, 2025; Nwobodo & Udoka, 2025; Ajao & Awofala, 2025; Eberechukwu & Stella, 2025).

Overall, the discussion suggests that weaknesses in the validity and reliability of classroom assessments in Rivers State secondary schools are not inevitable but are largely attributable to remediable factors—limited assessment literacy, inadequate institutional support, and absence of systematic psychometric checks. At the same time, regional and national examples of high-quality teacher-made instruments show that when appropriate procedures are followed, classroom assessment can meet acceptable standards of validity and reliability, providing a clear direction for policy and practice reforms in Rivers State.

Conclusion

This study examined the validity and reliability of classroom assessments in Rivers State secondary schools using survey data and psychometric analysis of teacher-made tests. The findings revealed that a substantial proportion of classroom assessments demonstrated limited content coverage and a strong bias toward lower-order cognitive skills. Although some tests met minimum acceptable standards, many did not adequately reflect the breadth and cognitive demands of the prescribed

curriculum. This suggests weaknesses in content and cognitive validity of classroom assessments across sampled schools.

In terms of reliability, the overall internal consistency of the sampled tests was moderate, with a mean reliability coefficient of 0.70. However, several tests fell below this benchmark, indicating inconsistency in the dependability of scores used for academic decisions. Teachers' assessment practices were generally moderate, but critical technical procedures such as preparation of test blueprints and conduct of item analysis were insufficiently practiced. Importantly, training in measurement and evaluation emerged as a significant predictor of reliability, highlighting the central role of assessment literacy in strengthening the psychometric quality of classroom assessments. The study therefore concludes that while classroom assessment remains central to secondary education in Rivers State, its validity and reliability are uneven and strongly influenced by teachers' technical competence in test construction and analysis.

Recommendations

1. The Rivers State Ministry of Education should organize regular and structured in-service training programs focused specifically on test construction, blueprint development, item writing, scoring rubrics, and basic item analysis techniques. Such professional development initiatives should move beyond general pedagogy and emphasize practical assessment literacy skills that directly enhance validity and reliability.
2. Secondary schools should institutionalize the compulsory use of tables of specifications and structured marking schemes for major classroom tests. Heads of departments and academic supervisors should review and moderate test instruments before administration to ensure adequate curriculum coverage and appropriate cognitive balance.
3. Quality assurance mechanisms within schools should be strengthened to include periodic psychometric review of teacher-made tests. Simple internal consistency analysis and item performance reviews can be incorporated into routine academic evaluation processes to improve reliability over time.
4. Teacher education institutions in Rivers State should reinforce coursework in educational measurement and classroom assessment, ensuring that pre-service teachers graduate with demonstrable competence in developing valid and reliable classroom tests.

References

- Tanzegyo, D. (2025). PSYCHOMETRIC PROPERTIES OF MULTIPLE-CHOICE TEST ITEMS IN HISTORY SUBJECT FOR ASSESSING STUDENTS IN SECONDARY SCHOOLS IN NASARAWA STATE, NIGERIA. *International Journal of Educational Research and Library Science*. <https://doi.org/10.70382/tjjerls.v07i8.045>
- Uwah, I., & Amadioha, A. (2024). Effect of Formative Assessment Methods on Mastery of Mathematics Course Content by Junior Secondary School Students in Obio-Akpor LGA of Rivers State. *International Journal of Quantitative and Qualitative Research Methods*. <https://doi.org/10.37745/ijqqrms.13/vol12n3114>
- Arhin, A., & Yeboah, A. (2025). Teachers' conceptualization of validity and reliability in classroom assessment in Ghana. *African Journal of Teacher Education*. <https://doi.org/10.21083/ajote.v14i1.8038>
- Nwobodo, C., & Udoka, C. (2025). Evaluating Technology Integration in Teaching and Learning: A Survey of Public Senior Secondary Schools in Rivers State, Nigeria. *African Journal of Humanities and Contemporary Education Research*. <https://doi.org/10.62154/ajhcer.2025.020.01012>

- Ajao, E., & Awofala, A. (2025). Trends in research in mathematics teaching and learning in Nigeria. *KIU Journal of Education*. <https://doi.org/10.59568/kjed-2025-5-1-13>
- Amuche, A., & Umar, M. (2019). Development and Validation of Lecturer Teaching Assessment Scale for Nigerian Universities. *Journal of Education and Practice*. <https://doi.org/10.7176/jep/10-5-05>
- Inko-Tariah, D. (2023). THE USE OF MODERN ASSESSMENT TOOLS IN ENHANCING LEARNING IN THE 21ST CENTURY SECONDARY SCHOOL CLASSROOMS IN NIGERIA. *International Journal of Education Humanities and Social Science*. <https://doi.org/10.54922/ijehss.2023.0563>
- Coombs, A., DeLuca, C., LaPointe-McEwan, D., & Chalas, A. (2018). Changing approaches to classroom assessment: An empirical study across teacher career stages. *Teaching and Teacher Education*, 71, 134-144. <https://doi.org/10.1016/j.tate.2017.12.010>
- Christian, K. (2020). Continuous Assessment as Predictor of Basic Education Certificate Examination Mathematics Performance in Public Secondary Schools in Obio-Akpor Local Government Area of Rivers State, Nigeria. *Journal of Humanities and Social Studies*, 8. <https://doi.org/10.24940/theijhss/2020/v8/i12/hs2012-033>
- Ibrahim, A., & Taura, A. (2025). Development, Validation, and Empirical Testing of an Innovative Metric For Assessing Lecturers' Pedagogical Efficacy Within Nigerian Higher Education Institutions. *Edukasiana: Jurnal Inovasi Pendidikan*. <https://doi.org/10.56916/ejip.v4i3.1168>
- Eberechukwu, A., & Stella, E. (2025). Social media utilization for instructional delivery and student assessment in Obio Akpor secondary schools, rivers state: Types, extent, and challenges. *American Journal of Creative Education*. <https://doi.org/10.55284/ajce.v8i2.1511>
- Chinedu, U., Ugwuanyi, C., & Idoko, J. (2025). Analysis of Teaching and Assessment Methods used on the Acquisition of Soft Skills Among Mathematics Students in Enugu State, Nigeria. *International Journal of Research and Scientific Innovation*. <https://doi.org/10.51244/ijrsi.2025.121500069p>