World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Test construction competency components and pupils' mathematics outcomes in Sierra Leone

Issa John Gbla *, Tamba Kebbie and Joseph J. Mbavai

School of Education, Njala University, Sierra Leone.

World Journal of Advanced Research and Reviews, 2026, 31(01), 1461–1467

Publication history: Received on 15 June 2026; revised on 21 July 2026; accepted on 24 July 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.31.1.1956

Abstract

This study examined how specific components of mathematics teachers' test construction competency predict pupils' learning outcomes in senior secondary schools in Western Area Rural District, Sierra Leone. The article is derived from Objective 4 of the author's doctoral thesis on test construction competency and pupils' mathematics achievement. A convergent mixed-methods design was used, with the quantitative strand reported here through component-level multilevel modelling. The sample comprised 40 senior secondary schools, 80 mathematics teachers and 640 pupils. Teachers completed a Test Construction Competency Questionnaire covering planning/blueprinting, item writing, item review/moderation and post-test item analysis/evidence use. Teacher-made mathematics assessment artefacts were audited, and pupils completed a curriculum-aligned mathematics achievement test. Because pupils were nested within teachers and schools, three-level random-intercept models were estimated. Adding the four competency components significantly improved model fit over the controls-only model (Delta -2LL = 16.00, df = 4, p = .003). Planning/blueprinting significantly predicted pupil achievement (B = 4.10, SE = 1.40, 95% CI [1.36, 6.84], p = .004), as did item writing (B = 2.50, SE = 1.10, 95% CI [0.34, 4.66], p = .024). Item review/moderation and post-test item analysis/evidence use were positive but not statistically significant. The findings suggest that front-end assessment practices, especially planning and item writing, offer practical leverage points for strengthening mathematics assessment quality and improving pupils' achievement in similar resource-constrained settings.

Keywords: Test Construction Competency; Mathematics Achievement; Assessment Literacy; Multilevel Modelling; Senior Secondary Schools; Sierra Leone.

1. Introduction

Classroom assessment is one of the most direct ways through which teachers make pupils' learning visible. In mathematics, this is especially important because concepts are cumulative and early misunderstandings can quickly become barriers to later learning. Well-designed tests can help teachers identify misconceptions, provide feedback and adjust instruction, while weak tests can misrepresent learning and delay the support pupils need. This makes test construction not a clerical task, but a core professional skill within classroom assessment practice (Black and Wiliam, 2018; Brookhart and Nitko, 2019).

Teacher assessment literacy includes the knowledge, judgement and practical skills required to design, interpret and use assessments appropriately. Within this broader construct, test construction competency refers to teachers' ability to plan tests, write technically sound items, review and moderate items, and analyse test results after administration. These competencies are shaped by teachers' training and school context, but they also depend on whether teachers are given the time, tools and support to translate assessment knowledge into classroom practice (Pastore and Andrade, 2019; Xu and Brown, 2016).

* Corresponding author: Issa John Gbla

In Sierra Leone, improving mathematics learning outcomes remains a national education priority. The Education Sector Plan 2022-2026 places strong emphasis on improving the quality of learning, while West African Senior School Certificate Examination (WASSCE) Mathematics performance continues to raise concern because credit-level achievement remains essential for access to many tertiary and professional pathways (MBSSE, 2022a, 2022b). Recent national assessment evidence reinforces this concern. The National Early Grade Reading and Mathematics Assessment reported that many pupils in Primary 2 and Primary 4 struggled with foundational numeracy tasks, including number identification, simple operations and word problems (MBSSE and UNICEF, 2021). Similarly, the Learning and School Safety Study found continuing concerns about pupil learning outcomes in junior and senior secondary schools two years after school closures in Sierra Leone (MBSSE and Leh Wi Lan, 2023).

These system-level findings are important, but they do not fully explain what happens inside routine teacher-made classroom tests. In senior secondary mathematics, such tests influence what teachers emphasise, what pupils practise and how schools judge readiness for promotion and public examinations. Studies from West African contexts have reported weaknesses in teacher-made tests, including limited use of tables of specification, uneven content coverage, weak cognitive balance and limited item analysis (Quansah et al., 2019; Kissi et al., 2023). Item-analysis literature also shows that poorly functioning items and weak distractors can distort test difficulty, discrimination and score interpretation (Rezigalla et al., 2024).

However, fewer studies in Sierra Leone have isolated the predictive contribution of specific test construction components to pupils' mathematics outcomes. This article, developed from Objective 4 of the author's PhD thesis, therefore determines the extent to which planning/blueprinting, item writing, item review/moderation and post-test item analysis/evidence use predict pupils' mathematics learning outcomes in senior secondary schools in Western Area Rural District.

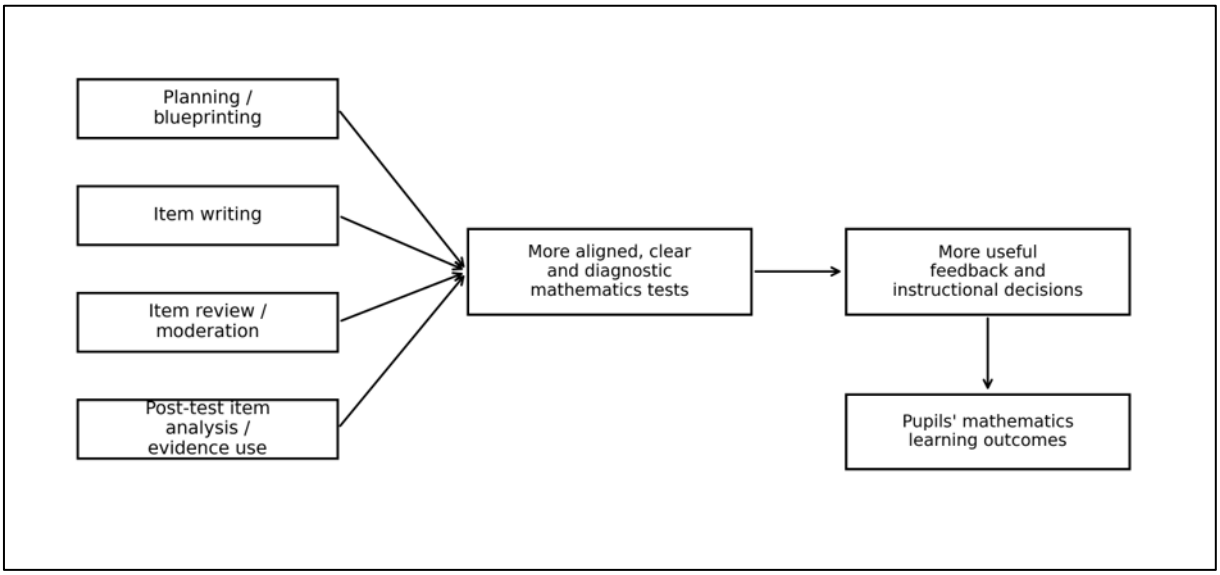

Figure 1 Conceptual pathway linking test construction competency components to pupils' mathematics learning outcomes. Source: Author's conceptualisation based on Black and Wiliam (2018), Brookhart and Nitko (2019), and Xu and Brown (2016)

2. Materials and methods

This manuscript is an article version of Objective 4 of the author's doctoral thesis submitted to Njala University in Sierra Leone. The wider thesis examined mathematics teachers' test construction competency and its predictive influence on pupils' learning outcomes in senior secondary schools in Western Area Rural District, Sierra Leone.

The study used a convergent mixed-methods design, although this article reports the quantitative component-level model most directly linked to Objective 4. Quantitative and qualitative data were collected within the same fieldwork phase and interpreted together in the wider thesis. This design was appropriate because it allowed statistical patterns to be interpreted alongside teachers' practical assessment conditions (Creswell and Creswell, 2018; Fetters et al., 2013).

The study was conducted in Western Area Rural District, Sierra Leone, a rapidly growing peri-urban district with public and approved private senior secondary schools. The analytic sample for Objective 4 comprised 640 pupils nested within 80 mathematics teachers/classes and 40 schools. The school sample was selected through proportionate stratification by management type and geographic area. Two focal mathematics teachers were selected in each school, and pupils were systematically sampled from the focal SSS2 and SSS3 classes.

Teachers completed a 40-item Test Construction Competency Questionnaire (TCC-Q) measuring four domains: planning/blueprinting (10 items), item writing (12 items), item review/moderation (9 items) and post-test item analysis/evidence use (9 items). Domain scores were computed as mean scores, with higher scores indicating stronger competency. Teacher-made mathematics assessment artefacts were also appraised using a Teacher Assessment Artefact Rubric (TAAR). Pupils' learning outcomes were measured using a curriculum-aligned mathematics achievement test comprising 30 multiple-choice items and 5 essay items.

The analysis used three-level random-intercept multilevel models because pupils were nested within teachers/classes and teachers were nested within schools. Teacher-level predictors were grand-mean centred. Model 0 was an unconditional model. Model 1 added control variables, including class level, teacher qualification, teaching experience and school type. Model 2 added the four TCC component scores to test their predictive contribution beyond the control variables. This approach is consistent with standard practice for clustered educational data (Hox et al., 2018).

Table 1 Study sample, variables and measurement sources

Element	Description
Study design	Convergent mixed-methods design; this article reports the Objective 4 quantitative component-level model.
Setting	Senior secondary schools in Western Area Rural District, Sierra Leone.
Schools	40 senior secondary schools selected from 138 eligible schools.
Teachers/classes	80 mathematics teachers/classes, two focal teachers per sampled school.
Pupils	640 pupils, with 16 pupils per school sampled from SSS2 and SSS3 focal classes.
Main predictors	Planning/blueprinting, item writing, item review/moderation and post-test item analysis/evidence use.
Outcome variable	Pupils' mathematics achievement test score.
Analytic approach	Three-level random-intercept multilevel models with pupils at Level 1, teachers/classes at Level 2 and schools at Level 3.

Measurement checks supported the use of the data for the Objective 4 model. TAAR inter-rater agreement was good (ICC(2,k) = 0.82, 95% CI [0.73, 0.88], $p < .001$), the pupil achievement test showed acceptable to good reliability (KR-20 = 0.75 for the multiple-choice section; alpha = 0.77 for the full test), and MCQ item difficulty was appropriate overall (mean p = 0.53).

3. Results and discussion

3.1. Descriptive context for the model

The pupil achievement test produced an overall mean score of 52.64 (SD = 12.43), indicating moderate achievement in the sampled classes. Teachers' overall TCC-Q score was also moderate (M = 2.96, SD = 0.41). Among the four domains, planning/blueprinting (M = 3.19), item writing (M = 3.12) and item review/moderation (M = 3.07) were moderate, whereas post-test item analysis/evidence use was low (M = 2.39). This pattern is important because it suggests that teachers were relatively stronger in front-end assessment activities than in the back-end use of test evidence.

3.2. Model fit and variance components

Table 2 presents the three nested models used to test Objective 4. The controls-only model did not improve fit over the unconditional model, but adding the four TCC component scores in Model 2 significantly improved fit over Model 1. Teacher-level variance also declined from 13.99 in Model 1 to 9.73 in Model 2, while the school-level variance reduced

to a boundary estimate. This indicates that the component block explained meaningful between-teacher/class differences in pupils' achievement.

Table 2 Model fit and variance components for the TCC component models

Model	Predictors included	Variance components (teacher / school / pupil)	Fit statistics	Improvement
M0	No predictors	16.50 / 5.20 / 99.30	-2LL = 4588.40; AIC = 4596.40; BIC = 4614.25	Baseline
M1	Class level, qualification, experience and school type	13.99 / 6.19 / 99.25	-2LL = 4588.22; AIC = 4604.22; BIC = 4639.91	Delta -2LL = 0.18, df = 4, p = .996
M2	Controls plus four TCC domains	9.73 / 0.00 / 99.19	-2LL = 4572.22; AIC = 4596.22; BIC = 4649.76	Delta -2LL = 16.00, df = 4, p = .003

3.3. Component effects on pupils' achievement

Table 3 shows the fixed effects for the four TCC components in the final model. Planning/blueprinting and item writing were statistically significant predictors of pupil achievement. By contrast, item review/moderation and post-test item analysis/evidence use had positive coefficients but did not reach statistical significance after the other domains and control variables were included.

Table 3 Final multilevel model: TCC component effects on pupil achievement

Predictor	B	SE	95% CI	p value	Interpretation
Planning/blueprinting	4.10	1.40	[1.36, 6.84]	.004	Significant positive predictor
Item writing	2.50	1.10	[0.34, 4.66]	.024	Significant positive predictor
Item review/moderation	1.10	0.95	[-0.76, 2.96]	.245	Positive, not significant
Post-test item analysis/evidence use	0.60	0.80	[-0.97, 2.17]	.452	Positive, not significant

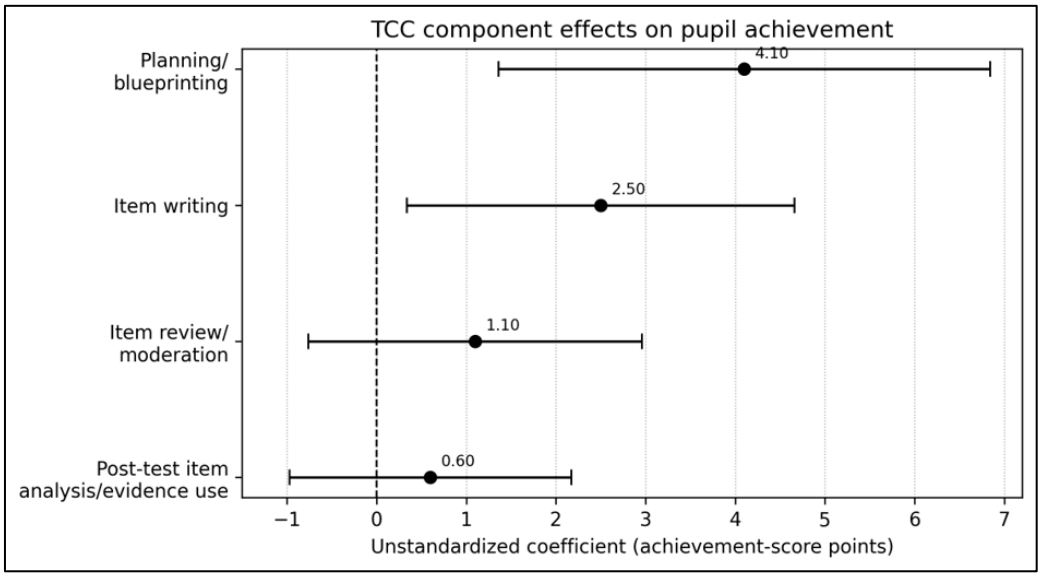

Figure 2 Coefficients and 95% confidence intervals for the four TCC components in the final multilevel model

The coefficient pattern shows that the most immediate assessment levers were the activities that occur before a test is administered. Planning/blueprinting had the strongest estimated association with pupils' achievement. This is educationally plausible: when teachers plan a test before writing questions, they are more likely to cover the intended syllabus areas, balance cognitive demand and reduce construct underrepresentation. In practical terms, a good blueprint helps the teacher ask: What have I taught? What should pupils be able to do? How much weight should each topic and cognitive level receive?

Item writing was also a significant predictor. This finding is consistent with the idea that clear, accurate and appropriately challenging items produce stronger assessment evidence. In mathematics, a poorly worded item can test reading confusion rather than mathematical understanding; a weak distractor can make a multiple-choice item too easy; and an ambiguous stem can punish pupils who understand the concept but misread the intended task. Strong item writing therefore supports both fairness and the diagnostic value of test scores (Hambleton and Jones, 1993; Rezigalla et al., 2024; Kubiszyn and Borich, 2024).

The non-significant results for item review/moderation and post-test item analysis should not be interpreted as evidence that these practices are unimportant. Rather, they may reflect the way these practices were implemented in the sampled schools. The wider thesis evidence indicated that moderation and post-test review were weaker and less routinely documented than planning and item writing. Where a practice is rare, rushed or informal, its independent statistical effect may be difficult to detect. This interpretation aligns with assessment literacy literature showing that teachers' assessment knowledge does not always translate into enacted practice without institutional support, time and clear routines (Brookhart, 2024; Shapovalov et al., 2022).

3.4. Decision on the hypothesis

Hypothesis H03 stated that the components of test construction competency do not significantly predict pupils' mathematics learning outcomes. The hypothesis was rejected because the four-component block significantly improved model fit (Delta -2LL = 16.00, df = 4, p = .003), and two components, planning/blueprinting and item writing, were statistically significant independent predictors. The findings therefore show that the components of test construction competency did not contribute equally; the strongest evidence came from the front-end processes of designing the test plan and constructing the items.

3.5. Practical implications

The findings have direct implications for teacher professional development and school-based quality assurance. First, mathematics departments should make the use of simple tables of specification routine for end-of-unit and end-of-term tests. Second, item-writing support should focus on clarity of stems, alignment with learning objectives, balanced cognitive demand, appropriate difficulty and functional distractors. Third, schools should introduce short moderation routines that are realistic for teachers' workload, such as a 15-minute peer review before tests are printed. Fourth, post-test item analysis should begin with manageable practices, including identifying very easy, very difficult or commonly misunderstood items before moving gradually toward more formal statistics.

3.6. Limitations

The study was conducted in one district and focused on senior secondary mathematics, so the findings should be generalised cautiously to other subjects, levels or districts. The design supports predictive inference but not strong causal claims. Although the multilevel model adjusted for selected background variables, other factors such as prior achievement, home support and language proficiency may also influence pupils' performance. In addition, some components, especially post-test item analysis, may have been under-developed in school practice, which could reduce variability and weaken statistical detection of their independent effects.

4. Conclusion

This article shows that specific components of mathematics teachers' test construction competency predicted pupils' learning outcomes in senior secondary schools in Western Area Rural District, Sierra Leone. Planning/blueprinting and item writing were significant independent predictors, while item review/moderation and post-test item analysis/evidence use were positive but not statistically significant. The evidence suggests that improving mathematics achievement requires attention not only to what teachers teach, but also to how teachers design the tests through which learning is measured, interpreted and acted upon. For immediate school improvement, the strongest entry points are practical training in blueprinting and item writing, supported by simple departmental routines for moderation and gradual adoption of post-test item analysis.

Compliance with ethical standards

Acknowledgments

The author acknowledges Njala University, the Ministry of Basic and Senior Secondary Education, participating school leaders, mathematics teachers and pupils in Western Area Rural District for their support during the wider doctoral study from which this article was developed.

Disclosure of conflict of interest

The author declares no conflict of interest.

Statement of ethical approval

Ethical approval and administrative permissions were obtained from the relevant authorities before data collection. Participation was voluntary.

Statement of informed consent

Informed consent was obtained from teachers and school leaders, while pupil participation followed school-approved parent/guardian consent procedures where required, and pupils provided assent. Participants' identities were protected through anonymisation and aggregate reporting.

Data availability statement

The data used for this article were generated as part of the author's doctoral research. Access to the full dataset is restricted to protect school, teacher and pupil confidentiality.

References

[1] Black P, Wiliam D. Classroom assessment and pedagogy. Assessment in Education: Principles, Policy & Practice. 2018;25(6):551-575. doi:10.1080/0969594X.2018.1441807

[2] Brookhart SM. Educational assessment knowledge and skills for teachers revisited. Education Sciences. 2024;14(7):751. doi:10.3390/educsci14070751

[3] Brookhart SM, Nitko AJ. Educational assessment of students. 8th ed. Pearson; 2019.

[4] Creswell JW, Creswell JD. Research design: Qualitative, quantitative, and mixed methods approaches. 5th ed. SAGE Publications; 2018.

[5] Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs: Principles and practices. Health Services Research. 2013;48(6 Pt 2):2134-2156. doi:10.1111/1475-6773.12117

[6] Hambleton RK, Jones RW. Comparison of classical test theory and item response theory and their applications to test development. Educational Measurement: Issues and Practice. 1993;12(3):38-47. doi:10.1111/j.1745-3992.1993.tb00543.x

[7] Hox JJ, Moerbeek M, van de Schoot R. Multilevel analysis: Techniques and applications. 3rd ed. Routledge; 2018.

[8] Kissi P, Baidoo-Anu D, Anane E, Annan-Brew RK. Teachers' test construction competencies in an examination-oriented educational system: Exploring teachers' multiple-choice test construction competence. Frontiers in Education. 2023;8:1154592. doi:10.3389/feduc.2023.1154592

[9] Kubiszyn T, Borich G. Educational testing and measurement: Classroom application and practice. 12th ed. Wiley; 2024.

[10] Ministry of Basic and Senior Secondary Education. Sierra Leone Education Sector Plan 2022-2026: Transforming learning for all. Government of Sierra Leone; 2022a.

[11] Ministry of Basic and Senior Secondary Education. WASSCE 2021 performance analysis report: Mathematics and core subjects. Government of Sierra Leone; 2022b.

[12] Ministry of Basic and Senior Secondary Education, UNICEF. Sierra Leone national early grade reading and mathematics assessment (NEGRA/NEGRMA): Baseline study report. Government of Sierra Leone/UNICEF; 2021.

[13] Ministry of Basic and Senior Secondary Education, Leh Wi Lan. Learning and School Safety (LASS) Study 2022: Briefing note 1 - Status of pupil learning outcomes in junior and senior secondary schools two years on from school closures in Sierra Leone. Government of Sierra Leone; 2023.

[14] Pastore S, Andrade HL. Teacher assessment literacy: A three-dimensional model. Teaching and Teacher Education. 2019;84:128-138. doi:10.1016/j.tate.2019.05.003

[15] Quansah F, Amoako I, Ankomah F. Teachers' test construction skills in senior high schools in Ghana: Document analysis. International Journal of Assessment Tools in Education. 2019;6(1):1-8. doi:10.21449/ijate.528390

[16] Rezigalla AA, Eleragi AMESA, Elhussein AB, Alfaifi J, AlGhamdi MA, Al Ameer AY, et al. Item analysis: The impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. BMC Medical Education. 2024;24(1):445. doi:10.1186/s12909-024-05433-y

[17] Shapovalov YA, Dugan AA, Soland J. Developing the educator assessment literacy professional learning screener: Research synthesis (ED627088). ERIC; 2022. Available from: https://files.eric.ed.gov/fulltext/ED627088.pdf

[18] Xu Y, Brown GTL. Teacher assessment literacy in practice: A reconceptualization. Teaching and Teacher Education. 2016;58:149-162. doi:10.1016/j.tate.2016.05.010