



(RESEARCH ARTICLE)



Mathematics teachers' test construction competency and pupils' learning outcomes in senior secondary schools in western area rural district, Sierra Leone

Issa John Gbla * and Tamba Kebbie

Department of Teacher Education, Faculty of Education, Freetown Polytechnic, Freetown, Sierra Leone.

World Journal of Advanced Research and Reviews, 2026, 30(01), 2547-2552

Publication history: Received on 16 March 2026; revised on 26 April 2026; accepted on 28 April 2026

Article DOI: <https://doi.org/10.30574/wjarr.2026.30.1.1132>

Abstract

Objective: This study examined whether mathematics teachers' test construction competency is associated with pupils' learning outcomes in senior secondary schools in Western Area Rural District, Sierra Leone.

Methods: The paper reports Objective Three of a broader doctoral study and used a cross-sectional correlational design within a convergent mixed-methods framework. The sample comprised 40 senior secondary schools (SSS), 80 mathematics teachers and 640 SSS2/SSS3 pupils. Data were collected with the Teacher Test Construction Competency Questionnaire, the Teacher Assessment Artefact Rubric (TAAR) and a curriculum-aligned mathematics achievement test. Teacher/class-level Pearson correlations were computed, and a three-level random-intercept model was estimated with pupils at Level 1, teachers/classes at Level 2 and schools at Level 3.

Results: Overall TCC-Q correlated strongly with TAAR overall ($r = 0.63$, $p < 0.001$) and positively with mean pupil achievement ($r = 0.32$, $p = 0.004$). TAAR overall had a stronger association with mean pupil achievement ($r = 0.38$, $p < 0.001$). In the multilevel model, teacher overall TCC remained a significant predictor of pupil achievement ($b = 5.915$, $SE = 2.017$, 95% CI [1.962, 9.867], $p = 0.003$).

Conclusion: Pupils achieved higher scores in classes where teachers demonstrated stronger test construction competency. Strengthening blueprinting, item writing, moderation and post-test evidence use should be treated as part of mathematics quality improvement.

Keywords: Assessment Literacy; Classroom Assessment; Mathematics Achievement; Multilevel Modelling; Sierra Leone; Test Construction Competency

1. Introduction

Assessment is central to classroom teaching because it shapes what teachers notice, how they judge progress and how they respond to learning gaps. In mathematics, this role is particularly important because learning is cumulative: when early misconceptions are not detected, later topics become more difficult for pupils to master. Well-constructed classroom tests therefore do more than assign marks. They help teachers identify misconceptions, align instruction with curriculum goals and provide pupils with clearer evidence of what they know and what still requires support (Black & Wiliam, 2018; Brookhart & Nitko, 2019).

The usefulness of classroom assessment depends heavily on teachers' assessment literacy and, more specifically, on their ability to construct sound tests. Contemporary literature treats teacher assessment literacy as a multidimensional professional competence that combines knowledge, technical skill, judgement and the ethical use of assessment

* Corresponding author: Issa John Gbla

evidence (DeLuca et al., 2016; Xu & Brown, 2016; Pastore, 2023; Lei & Lei, 2026). From this perspective, test construction competency is not a minor administrative task; it is an important part of teaching quality and instructional decision-making.

Recent African evidence reinforces this concern. In Ghana, studies have reported gaps in mathematics teachers' classroom assessment literacy, heavy reliance on traditional paper-and-pencil tests and uneven attention to test design, scoring and use of assessment results (Akayuure, 2021; Armah, 2025; Asamoah et al., 2023). Other evidence shows that teachers' test construction competence is related to the quality of the tests they produce (Kissi et al., 2023; Quansah et al., 2019). This is particularly important in examination-oriented systems where teacher-made tests influence feedback, promotion decisions and preparation for external examinations.

In Sierra Leone, the Education Sector Plan 2022-2026 emphasizes the improvement of learning quality in core subjects, including mathematics (Government of Sierra Leone, 2022). However, empirical evidence remains limited on how mathematics teachers' test construction competency relates to pupils' actual achievement, particularly at senior secondary level in Western Area Rural District. This gap matters because teachers' daily assessment practices may either reveal learning problems early or conceal them until pupils face high-stakes examinations.

Against this background, the study examined the relationship between mathematics teachers' overall test construction competency and pupils' learning outcomes in senior secondary schools in Western Area Rural District, Sierra Leone. The null hypothesis tested was that there is no statistically significant relationship between teachers' overall test construction competency and pupils' mathematics achievement.

2. Materials and methods

2.1. Research design and setting

This paper reports the quantitative evidence for Objective Three of a broader doctoral study. The design for the present analysis was cross-sectional and correlational, situated within a wider convergent mixed-methods framework. The study was conducted in Western Area Rural District, Sierra Leone, a fast-growing peri-urban district with a mixture of public and approved private senior secondary schools.

2.2. Population, sample and sampling procedure

The school-level population comprised 138 senior secondary schools in the district. Using proportionate stratified sampling, 40 schools were selected, representing approximately 29.0% of the population. Two mathematics teachers were selected from each school, one linked to SSS2 and one linked to SSS3, giving a teacher/class sample of 80. To preserve teacher-pupil linkage for multilevel analysis, 16 pupils were selected from each school: 8 from the SSS2 focal class and 8 from the SSS3 focal class. The final analytic sample therefore comprised 640 pupils, 80 teachers/classes and 40 schools.

2.3. Instruments

Three instruments were used. First, the Teacher Test Construction Competency Questionnaire (TCC-Q) measured teachers' self-reported competence in planning and blueprinting, item writing, item review and moderation, and post-test analysis and evidence use. The overall 40-item scale showed strong internal consistency (Cronbach's alpha = 0.90; omega = 0.91), and exploratory factor analysis supported a four-factor structure (KMO = 0.780; Bartlett's chi-square (780) = 1767.366, $p < 0.001$; cumulative variance explained = 58.4%).

Second, the Teacher Assessment Artefact Rubric (TAAR) was used to audit teacher-made assessment artefacts, including tests, marking guides and evidence of moderation or post-test analysis. Third, pupils completed a curriculum-aligned mathematics achievement test consisting of 30 multiple-choice items and 5 essay questions administered within 60 minutes.

2.4. Data collection procedure

Data were collected during scheduled school visits. Teachers completed the TCC-Q, assessment artefacts were collected and rated using the TAAR, and pupils completed the mathematics achievement test under standardised conditions. The linking of teacher, class and pupil records was maintained for analysis, while names and school identifiers were removed from the analytic dataset.

2.5. Data analysis

Analysis proceeded in two stages. First, Pearson correlation coefficients were computed at teacher/class level. At pupil level, teacher-cluster robust tests were used to account for the dependence of pupils within the same teacher/class. Second, a three-level random-intercept model was estimated with pupils at Level 1, teachers/classes at Level 2 and schools at Level 3. Teacher overall TCC was grand-mean centred before entry into the model. Model fit was compared using -2 log likelihood (-2LL), Akaike information criterion (AIC) and Bayesian information criterion (BIC). Statistical significance was evaluated at the 0.05 level.

2.6. Ethical considerations

Administrative permission and ethical clearance were obtained before data collection. Participation was voluntary, and teachers and pupils were informed that the study was for research purposes rather than staff appraisal or pupil grading. Confidentiality was maintained by anonymising schools, teachers and pupils in the dataset and in all reporting.

3. Results and discussion

The analysis examined whether teachers' overall test construction competency was associated with pupils' mathematics achievement. Because the study combined self-report data, artefact-audit evidence and pupil achievement scores, the findings are presented as bivariate associations followed by multilevel estimates that account for the nested structure of the data.

3.1. Bivariate associations

Table 1 Key associations among teacher competency, assessment artefacts and pupil achievement

Association	R	p-value	n
TCC-Q overall × TAAR overall	0.63	< 0.001	80
TCC-Q overall × mean pupil achievement	0.32	0.004	80
TAAR overall × mean pupil achievement	0.38	< 0.001	80
Teacher TCC-Q overall × pupil achievement	0.29	< 0.001	640
TAAR overall × pupil achievement	0.33	< 0.001	640

Note. The first three rows are teacher/class-level correlations ($n = 80$); rows 2 and 3 use teachers' class mean achievement. The final two rows are pupil-level associations ($n = 640$), with p-values estimated using teacher-cluster robust tests. TCC-Q = Teacher Test Construction Competency Questionnaire; TAAR = Teacher Assessment Artefact Rubric.

Overall TCC-Q was strongly related to overall TAAR ($r = 0.63$, $p < 0.001$), showing that teachers who reported stronger test construction competency also tended to produce higher-quality assessment artefacts. The relationship with pupil achievement was positive at both levels of analysis. At teacher/class level, overall TCC-Q was associated with mean pupil achievement ($r = 0.32$, $p = 0.004$), while TAAR overall showed a somewhat stronger relationship ($r = 0.38$, $p < 0.001$). At pupil level, the association between teacher overall TCC-Q and pupil achievement remained positive and statistically significant ($r = 0.29$, $p < 0.001$). This pattern suggests that observable assessment practice may be more closely linked to achievement than self-reported competence alone.

3.2. Multilevel model

Table 2 Three-level multilevel models predicting pupil mathematics achievement from teacher overall TCC

Model	Predictor / fit index	Estimate
Model 0	Intercept (grand mean)	52.64
Model 0	Var (school intercept)	5.20
Model 0	Var (teacher/class intercept)	16.50
Model 0	Var (Level-1 residual)	99.30
Model 0	ICC (school)	0.043

Model 0	ICC (teacher/class)	0.136
Model 0	Fit	-2LL = 4588.40; AIC = 4596.40; BIC = 4614.25
Model 1	Intercept (grand mean)	52.64
Model 1	Teacher overall TCC (b, SE, 95% CI, p)	5.915; 2.017; [1.962, 9.867]; 0.003
Model 1	Var (school intercept)	4.80
Model 1	Var (teacher/class intercept)	13.90
Model 1	Var (Level-1 residual)	99.20
Model 1	ICC (school)	0.041
Model 1	ICC (teacher/class)	0.118
Model 1	Fit	-2LL = 4580.10; AIC = 4590.10; BIC = 4612.41

Note. Model 0 is the unconditional model. Model 1 adds teacher overall TCC, grand-mean centred, at Level 2. Pupils are nested in teachers/classes, and teachers/classes are nested in schools. TCC = test construction competency; ICC = intraclass correlation coefficient; -2LL = -2 log likelihood; AIC = Akaike information criterion; BIC = Bayesian information criterion.

4. Discussion

The null model showed that achievement varied across both teachers/classes and schools, although clustering was stronger at teacher/class level (ICC = 0.136) than at school level (ICC = 0.043). When teacher overall TCC was entered in Model 1, it emerged as a statistically significant positive predictor of pupil achievement ($b = 5.915$, $SE = 2.017$, 95% CI [1.962, 9.867], $p = 0.003$). Model fit also improved, with -2LL reducing from 4588.40 to 4580.10 and AIC reducing from 4596.40 to 4590.10. The teacher/class-level variance decreased from 16.50 to 13.90, suggesting that differences in test construction competency explained part of the between-class variation in pupil achievement.

The results provide a consistent message: pupils tended to perform better in classes where mathematics teachers demonstrated stronger test construction competency. This association was visible in simple correlations and remained significant after accounting for the nesting of pupils within teachers/classes and schools. In practical terms, the model indicates that a one-unit increase in overall TCC was associated with an estimated 5.92-point increase in pupil achievement.

The strong association between TCC-Q overall and TAAR overall is also important. It suggests that, in this sample, teachers' self-reported assessment competence was not merely aspirational; it was linked to observable qualities of their assessment artefacts. At the same time, the slightly stronger association between TAAR overall and pupil achievement indicates that enacted assessment practice may be more closely tied to learning outcomes than self-perception alone. This supports the argument that assessment competence should be understood as both knowledge and practice (Xu & Brown, 2016; Pastore, 2023).

These findings are consistent with evidence that teachers' test construction competence is associated with the quality of teacher-made tests (Kissi et al., 2023; Quansah et al., 2019). The present study extends that line of work by showing that competence is also associated with pupils' measured mathematics achievement. The results further align with broader evidence that teacher quality indicators are meaningfully related to students' academic outcomes (Engida et al., 2024). In this study, the reduction in teacher/class-level variance after adding TCC to the model highlights the classroom as an important site where achievement differences are produced.

The findings also help explain why assessment capacity is important in systems where classroom tests are a dominant assessment tool. Previous studies have shown that mathematics teachers may rely heavily on traditional tests while using diagnostic and data-driven assessment practices less frequently (Akayuure, 2021; Armah, 2025; Asamoah et al., 2023). Where classroom tests are frequently used, the technical quality of those tests becomes highly consequential for feedback, remediation and instructional planning.

Nevertheless, the findings should be interpreted with appropriate caution. The cross-sectional design supports evidence of association but does not prove causation. In addition, assessment interventions do not automatically improve achievement unless teachers are supported to use assessment evidence for feedback, re-teaching and follow-up. Evidence from formative assessment research shows that effects can vary depending on implementation depth and

duration (Bostrom & Palm, 2023). For this reason, test construction training should be practical, sustained and linked to teachers' real assessment tasks.

The practical implication is clear. School and district improvement efforts should move beyond general assessment workshops and provide hands-on support in blueprinting, writing valid items, reviewing and moderating tests, preparing marking guides and analysing test results. When mathematics teachers construct clearer, fairer and more instructionally useful tests, pupils are more likely to receive accurate feedback and targeted teaching support.

5. Conclusion

This study found a positive and statistically significant relationship between mathematics teachers' overall test construction competency and pupils' learning outcomes in senior secondary schools in Western Area Rural District, Sierra Leone. Teachers who demonstrated stronger test construction competency tended to have pupils with higher mathematics achievement, and this pattern remained significant in a three-level multilevel model.

The evidence suggests that classroom assessment quality should be treated as part of the wider effort to improve mathematics learning. In practical terms, schools, teacher educators and district authorities should invest in sustained, practice-based support for test blueprinting, item writing, moderation, marking guides and the instructional use of assessment evidence. Strengthening these routines can make classroom tests fairer, more informative and more useful for learning. Because the study was cross-sectional and limited to one district, the results should be interpreted as evidence of association rather than definitive proof of causation.

Compliance with ethical standards

Acknowledgments

The authors acknowledge the participating schools, mathematics teachers and pupils for their cooperation during data collection

Disclosure of conflict of interest

The authors declare no conflict of interest.

Statement of ethical approval

Ethical clearance and administrative permission were obtained from the relevant authorities before fieldwork commenced.

Statement of informed consent

Informed consent was obtained from participating teachers, and pupil participation was conducted with appropriate school-level permission and assent procedures.

Data availability statement

The data supporting the findings are available from the author upon reasonable request, subject to confidentiality and ethical restrictions.

References

- [1] Akayuure, P. (2021). Classroom assessment literacy levels of mathematics teachers in Ghanaian senior high schools. *Contemporary Mathematics and Science Education*, 2(2), Article ep21013. <https://doi.org/10.30935/conmaths/11286>
- [2] Armah, P. H. (2025). Assessment practices of Ghanaian senior high school mathematics teachers: A national survey of practices, self-rated skills, and predictors. *European Journal of Mathematics and Science Education*, 6(3), 161-177. <https://doi.org/10.12973/ejmse.6.3.161>

- [3] Asamoah, D., Shahrill, M., & Abdul Latif, S. N. (2023). Towards developing classroom assessment literacy: Exploring teachers' approaches to assessment across cultures. *Cogent Education*, 10(2), Article 2280301. <https://doi.org/10.1080/2331186X.2023.2280301>
- [4] Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551-575. <https://doi.org/10.1080/0969594X.2018.1441807>
- [5] Bostrom, E., & Palm, T. (2023). The effect of a formative assessment practice on student achievement in mathematics. *Frontiers in Education*, 8, Article 1101192. <https://doi.org/10.3389/educ.2023.1101192>
- [6] Brookhart, S. M., & Nitko, A. J. (2019). *Educational assessment of students* (8th ed.). Pearson.
- [7] DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251-272. <https://doi.org/10.1007/s11092-015-9233-6>
- [8] Engida, M. A., Iyasu, A. S., & Fentie, Y. M. (2024). Impact of teaching quality on student achievement: Student evidence. *Frontiers in Education*, 9, Article 1367317. <https://doi.org/10.3389/educ.2024.1367317>
- [9] Government of Sierra Leone. (2022). *Sierra Leone education sector plan 2022-2026: Transforming learning for all*. Government of Sierra Leone.
- [10] Kissi, P. K., Baidoo-Anu, D., Anane, E., & Annan-Brew, R. K. (2023). Teachers' test construction competencies in examination-oriented educational system: Exploring teachers' multiple-choice test construction competence. *Frontiers in Education*, 8, Article 1154592. <https://doi.org/10.3389/educ.2023.1154592>
- [11] Lei, W., & Lei, Z. (2026). Formative assessment literacy: A systematic review. *Language Testing in Asia*, 16, Article 7. <https://doi.org/10.1186/s40468-025-00418-0>
- [12] Pastore, S. (2023). Teacher assessment literacy: A systematic review. *Frontiers in Education*, 8, Article 1217167. <https://doi.org/10.3389/educ.2023.1217167>
- [13] Quansah, F., Amoako, I., & Ankomah, F. (2019). Teachers' test construction skills in senior high schools in Ghana: Document analysis. *International Journal of Assessment Tools in Education*, 6(1), 1-8. <https://doi.org/10.21449/ijate.481164>
- [14] Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162. <https://doi.org/10.1016/j.tate.2016.05.010>