

# Moderation Process For Valid Mathematics Teacher-Made Test Results: Evidence of Final National Performance in Secondary Schools in Zanzibar

<sup>1</sup>Hassan A. Haji, <sup>2</sup>Dr. Victorini Salema, <sup>3</sup>Dr. Peter Siamoo

<sup>1</sup>Student-Mwange Catholic University

<sup>1,2</sup>Lecturer-Mwange Catholic University

## Abstract

This study assessed the moderation process for valid Mathematics teacher-made test (TMT) results and their implications for national performance in secondary schools in Zanzibar. Grounded in Classical Test Theory (Novick, 1968), the study employed a convergent design under a mixed-methods approach. The target population included 3,768 individuals from 30 public schools, including students, heads of schools, Mathematics teachers, and ZEC subject coordinators. A sample of 436 respondents was selected using purposive and proportionate stratified random sampling techniques. Data collection instruments included interview guides, document analysis guides, and questionnaires. Instrument validity was confirmed by three education research experts who supervise postgraduate students from MWECAU. At the same time, reliability was assessed using the Cronbach Alpha method ( $r = 0.774$  for teacher questionnaires and  $r = 0.759$  for student questionnaires). Ethical guidelines were strictly followed throughout the research. Quantitative data were analysed using descriptive and inferential statistics, while qualitative data were subjected to thematic analysis. The findings revealed a significant gap between teachers' perceptions (high extent) and students' (moderate extent) about quality of test Moderated to enhance students Mathematics performance. The finding of the hypothesis testing reveals low credibility of Mathematics TMT in enhancing students' academic achievement, indicating a need for improving moderation process, assessment procedures and teacher training to align TMT with National Examinations Standards. However, interviews and document reviews highlighted inadequacies in moderation and potential grading biases. These findings underscore the need for improving TMT to support students in teaching and learning to enhance students' performance in Mathematics in FTNE. The study concludes that enhanced communication and greater teacher involvement in moderation are essential for fair and reliable assessments. It recommends the implementation of structured moderation processes and the promotion of open communication to improve the effectiveness of Mathematics TMTs and support student success in FTNA.

**Keywords:** Moderation, Mathematics Teacher-Made Tests, Secondary School Performance, Form Two National Assessment (FTNA).

## 1. Introduction

Moderation of mathematics assessments in a global context enhances assessments' validity, reliability, and fairness, improves students' performance, and reduces disparities across diverse educational settings. It ensures that assessment tools are valid and reliable, leading to better measures of student understanding (Lee & Kim, 2023). Additionally, moderation reduces performance disparities, especially in diverse contexts (Johnson, 2023). Collaborative strategies, such as workshops and capacity-building seminars, help educators develop a shared understanding of standards (Pang et al., 2021), vital for supporting all students, including those facing socio-economic challenges (Schmidt et al., 2023). Inconsistent moderation practices across regions, however can undermine the effectiveness of assessments, highlighting the need for standardised approaches to ensure equitable and high-quality evaluations for all students. The variability in moderation practices across different regions raises concerns about the quality of TMT and their effectiveness in

enhancing national academic performance. Without consistent moderation processes, the potential benefits of the assessments may not be fully realised, ultimately affecting students' learning outcomes on a national level. Despite all these, it is not yet clear how the moderation of TMT influences national mathematics examinations.

In the United States, Johnson (2022) found that well-structured TMT, when properly moderated, significantly enhances student engagement and performance. Similarly, Müller (2023) in Germany noted that students with lower test anxiety tend to perform better in mathematics. The information might be because mathematics deals with facts students can prove before submission. Addressing psychological factors like test anxiety is crucial for effective assessment practices. In Ireland, Cronin et al. (2021) emphasised that moderated mathematics support improves students' grades, confidence, retention, progression, and employability over twenty years. In Japan, Shimizu (2025) highlighted that learning engagement influences how self-efficacy and mathematics anxiety affect students' performance in complex plane problem-solving. While these studies underscore the importance of quality assessments, they raise critical questions about ensuring consistent moderation practices across different educational contexts. How can schools implement uniform moderation processes effectively?

In Asia, research by Tan and Saligumba (2019) and Setiawan (2023) emphasises the importance of developing valid TMT through rigorous content validation and item analysis, which enhances alignment with learning objectives and improves student performance. In Africa, Yılmaz (2022) explored the role of group moderation in enhancing assessment reliability, while Owusu et al. (2021) investigated how mathematics interest and teaching quality impact student performance in Ghana. Despite these valuable insights, a pressing need remains to improve teacher collaboration and address sampling limitations in these studies, as their findings may not be universally applicable across diverse educational contexts. Akendita et al. (2025) further examined the moderating effect of teacher efficacy on the relationship between students' perceptions of mathematics and their achievement. The Internal Moderator report in South Africa also analyses National Senior Certificate examination question papers for clarity, difficulty, and curriculum alignment, enhancing assessment quality and providing essential insights for educators and policymakers (WCED, 2024). However, the inconsistency in moderation practices across regions raises concerns about the overall effectiveness of these assessments and their implications for student performance on a national scale. Addressing these issues is critical for ensuring that assessments are valid, reliable, equitable, and reflect all students' abilities.

In East Africa, particularly Tanzania, the moderation of teacher-made tests is essential for ensuring fair and valid assessments. Challenges such as limited resources, varying teacher qualifications, and insufficient training hinder effective moderation processes (Mhando et al., 2020; NECTA, 2021). These inadequacies often result in assessments that do not accurately reflect students' mathematical abilities, misaligning educational strategies and support systems for struggling learners (Mhando & Mgaiwa, 2021). The National Examination Council of Tanzania (NECTA) emphasises the need for standardised assessment tools to ensure comparability across schools and regions (NECTA, 2021). However, the effectiveness of these standardised tools remains questionable, particularly in how well they accommodate the diverse needs of students and educators in different contexts. While assessment in the classroom might appear fair, the case might be different with the national examinations.

In Zanzibar, moderated mathematics teacher-made tests (TMT) are critical for accurately assessing student performance, given the unique challenges posed by resource limitations and varying teacher qualifications. The ineffective moderation of mathematics TMT contributes to low performance in national assessment, evidenced by persistently low performance in the Mathematics FTNA results, which have remained below 35% over the past five years in the ZEC Basic Statistical Report (2020–2024). These result in inaccurate assessment procedures that fail to reflect students' abilities in TMT. This situation creates a cycle of poor preparation, as students cannot build the necessary skills and confidence to tackle TMT and national examinations successfully. Students inadequately grasp foundational mathematics concepts, leading to increased anxiety and decreased confidence in their abilities, further spreading their low performance in future assessments (CIRA, 2023). The consequences of low performance are far-reaching, affecting not only the students' immediate educational outcomes but also their long-term opportunities for further education and employment. Teachers are hindered by insufficient assessment procedures, making adapting their teaching strategies effectively to meet students' needs challenging. These contribute to continued low

performance in mathematics and limit opportunities for student success, exacerbating the issue and leading to a lack of student engagement and motivation.

Effective teaching strategies in mathematics, including differentiated instruction, active learning, formative assessments, technology integration, collaborative learning, scaffolding, real-world applications, visual aids, and constructive feedback, aim to enhance student engagement and performance. This study addresses low performance, such as poor test quality, lack of moderation, and insufficient motivation. Improving mathematics education is vital for individual success and the region's socio-economic development, as a mathematically proficient population is better equipped to tackle challenges and contribute positively to society. The moderation process is essential for ensuring the quality of mathematics teacher-made tests, as ineffective moderation leads to unreliable assessments and low student performance. Addressing these issues is crucial for improving students' academic outcomes and future prospects in Zanzibar.

## **2. Statement of the Problem.**

Moderation of Mathematics TMT in Unguja, Zanzibar faces significant challenges, as evidenced by persistently low performance in the FTNA results, which have remained below 35% over the past five years in the ZEC Basic Statistical Report (2020–2024). Despite the collaborative effort taken by the government and other educational stakeholders, the desired achievement has not been achieved. These raise concerns about the Moderation process for valid Mathematics TMT. Various stakeholders express dissatisfaction, including students, parents, and teachers (CIRA, 2023). Students reported struggling with unclear expectations and insufficient preparation for national examinations, contributing to low academic achievement. While previous studies have emphasised the importance of formative assessments in improving students' academic performance (Nziku & Matogwa, 2021; Kyaruzi, 2019; Fosu et al., 2023; Takunyaci & Aydin, 2021), there remains a significant gap in understanding how moderated Mathematics TMT directly influences students' performance in FTNA. The reviewed studies had diverse focuses, and therefore, the influence of moderated TMT on the National Form Two Mathematics examination remains unknown (EDZ, 2019; VSO International. 2019 & World Bank. 2020) based on Mechanism for Enhancing Primary Education Development and effective school leadership and quality education. Instead of assessed the moderation process for valid TMT results and students' mathematical abilities, leading to ineffective educational strategies and insufficient support. If these issues remain unresolved, they will exacerbate performance disparities hinder students' future success due to poor-quality of final mathematics assessments. Therefore, this study assessed how the Moderation process for valid Mathematics TMT enhances Mathematics performance of Form Two students in Unguja, Zanzibar.

## **3. Research Question**

This study was guided by one research question and one hypothesis, which were:-

To what extent does a moderation of Mathematics teacher-made tests enhance Mathematics performance of Form Two students in Unguja, Zanzibar?

### **3.1 Research Hypothesis**

H<sub>1</sub>: There is a significant relationship between the scores of moderated mathematics teacher-made tests and the national assessment mathematics performance of Form Two students in Unguja, Zanzibar.

## **4. Significance of the Study**

The study's findings are significant in informing the Educational stakeholders, the Ministry of Education and Vocational Training in Zanzibar (MoEVTZ), and Policymakers to reduce the inconsistency between the current situation of students' Mathematics test scores and expected results in FTNA. Filling the gaps in Mathematics performance concerning moderated TMT and students' final national Assessments. Particularly among Form Two students who consistently show low achievement in Mathematics in national assessments. By identifying effective assessment practices and providing empirical data, the research aims to inform policymakers and educational authorities about the importance of high-quality moderation practices, guiding the development of policies that emphasise effective assessment methods. The findings significantly contribute to teachers by enhancing their skills in designing meaningful assessment tools, while benefiting Form Two students by improving their academic performance in the Mathematics national assessments. This study enriches the body of knowledge in educational assessment. It offers actionable insights that enhance

educational practices, ultimately supporting future research and advancing Mathematics education in Unguja, Zanzibar.

## **5. Theoretical Framework**

Classical Test Theory (CTT), developed by Lord and Novick (1968), is a foundational framework in psychometrics and educational measurement, particularly in assessing test validity and reliability. CTT clarifies the relationship between observed test scores, accurate scores, and measurement errors. This theory has critical implications for improving test quality, especially in the Test of Mathematics (TMT), where it analyses test items for validity, reliability, actual score variance, and error variance. By focusing on single error sources such as time constraints or item design, CTT guides the validation of teachers' terminal tests, thereby enhancing academic outcomes for form two students. Strategies integrating CTT principles, such as test moderation and alignment with instructional objectives, are essential for reducing score errors and ensuring equitable test administration (Lord & Novick, 1968).

The relevance of CTT is particularly evident in mathematics, where it provides systematic approaches to enhance the quality, validity, and reliability of assessments. By applying CTT principles, mathematics teachers can better align tests with learning outcomes, evaluate item quality, and analyse item difficulty, effectively mitigating errors from various sources, including teacher biases and test environments. Test moderation practices grounded in CTT enable educators to identify and rectify inconsistencies in test items, ensuring fair assessments. Furthermore, CTT's focus on minimising measurement errors through careful test design supports improving students' mathematical performance (Baker & Kim, 2018; Wilson & Allen, 2020). By addressing issues arising from test items and environmental factors, CTT aids in creating accurate assessments of student performance, ultimately contributing to enhanced academic achievement in mathematics and emphasising the importance of robust assessment practices that increase the validity and reliability of the tests designed.

## **6. Literature Review**

### **6.1 The Extent to Which Moderation of Teacher-Made Test of Mathematics Subject Enhances Academic Performance of Students**

The empirical reviews focus on how the moderation of TMT items in Mathematics aims to enhance students' academic performance. It involves reviewing the questions' clarity, accuracy, and alignment with the curriculum. Moderators ensure the process promotes fair grading in assessment tools and academic performance.

Wedel (2021) conducted a study in Germany about the instruction time and student achievement: The moderating role of teacher qualifications. The previous research employed regression analysis and multilevel modelling, alongside instruments such as standardised test scores, teacher qualification metrics, classroom observation protocols, and instructional time logs. The findings indicated that each additional hour of instruction increases test scores by an average of 0.03 standard deviations, with a more substantial effect of 0.04 to 0.05 standard deviations for students taught by well-qualified teachers. The previous study informed the current study by highlighting the importance of examining various moderating factors. However, while the study focused on teacher qualifications, it primarily overlooked other moderating factors, such as student engagement, classroom environment, instructional strategies, and parental involvement, which could influence academic achievement and limit the reliability of its findings. Therefore, the current study assessed the moderation of test items in Mathematics subjects to improve students' performance in national assessments.

Takunyaci and Aydin (2021) conducted a study in Turkey about the implementation of a Group Moderation Assessment Model in an open-ended mathematics examination, using a non-random appropriate sampling method to assess with a team of five mathematics teachers, finding that the raters were positively influenced by each other and formed a reliable evaluation system through their collaborative judgments. The previous study informed this study by demonstrating the value of collaborative judgment in enhancing assessment reliability; however, the previous study employed document analysis, interviews, non-random interviews, and a Non-Random sampling technique. The current study aims to assess the mathematics teachers in moderating test items to enhance students' academic achievements. To extend on it, the current study aimed to assess test items for improving student academic performance. The current study adopted probability and non-probability sampling methods, and the role of the tests moderated the process for enhancing students'

performance in Mathematics. The national employed likelihood and non-probability sampling methods, with the role of the tests' moderating process that improved students' performance in the national Mathematics assessment.

Harahap et al. (2023) conducted a study in Indonesia examining the efficacy of mathematics as a moderator on the relationship between teacher support and middle school students' academic engagement. The study employed a quantitative design with a target population of middle school students, consisting of a sample size 430. A questionnaire was used to assess teacher support and mathematics self-efficacy. Findings revealed that teacher support significantly aids students in understanding mathematics. Previous research utilised Multiple Regression Analysis (MRA) and Structural Equation Modelling (SEM) to analyse perceptions from both teachers and students, with results showing  $\beta = 0.18$  for teacher support and  $\beta = 0.19$  for the interaction between teacher support and mathematics self-efficacy. The current study builds on these findings by incorporating the moderating role of mathematics teacher-made tests (TMT) in enhancing student performance. It employed a convergent design within a mixed-methods approach, using probability and non-probability sampling techniques. It included questionnaires, interviews, and document analysis to provide a more comprehensive understanding of how the moderation of mathematics TMT enhances students' mathematics performance.

Tan and Saligumba (2019) conducted a study in the Philippines about developing Valid and Reliable Teacher-made Tests for Mathematics. The previous research employed surveys, questionnaires, expert assessments by educational specialists, interviews, and document analysis guides, while this study used a convergent design under a mixed approach. However, the tests achieved strong reliability, with Kuder-Richardson (KR-20) coefficients of 0.82 and 0.85, indicating effective alignment with learning objectives. Also, the Findings showed that the tests accurately measured students' understanding and skills in mathematics. The previous study informed the current study by highlighting effective assessment methods, but lacked a broader range of data sources. The current study aims to build on these findings by assessing how the moderation of mathematics TMT enhances students' academic achievements and improves performance in national assessments to bridge the gap.

Fosu et al. (2023) conducted a study in Ghana on the mediation and moderation effects of mathematics interest and teaching quality between self-concept and mathematics achievement, finding that both factors significantly mediated this relationship. The previous study informed this study by highlighting the significance of interest and teaching quality in enhancing mathematics achievement and the importance of procedural accuracy in sampling. However, the previous study employed a sample of 300 students from a population of 1,200, which raised concerns about selection bias due to insufficient details on the randomisation process, affecting generalizability. To address these limitations, the current study employed probability and non-probability sampling techniques, such as stratified random and purposive sampling, to assess the Moderation of Mathematics tests to enhance students' Mathematics performance in national assessments.

Bright (2025) conducted a study in Ghana examining the moderating effect of teacher-student relationships and the mediating role of self-efficacy on the relationship between peer-assisted learning and mathematics performance. Using a descriptive-correlation design, the study sampled 351 students through stratified simple random sampling and employed structured questionnaires for data collection. Findings indicated that teacher-student relationships, self-efficacy, and peer-assisted learning positively influenced mathematics performance, mediating the relationship between peer-assisted learning and performance. To extend this, the current study adopted a convergent design using a mixed-methods approach, and probability and non-probability sampling techniques were used. It employs questionnaires, interviews, and document analysis guides to assess the moderation of mathematics TMT in enhancing students' mathematics performance. Therefore, this study assessed the moderation of mathematics TMT to enhance students' mathematics performance.

## **7. Demonstration of Research Gap**

The reviewed literature on the Moderation of Mathematics TMT reveals significant gaps that necessitate further exploration, particularly in Unguja, Zanzibar. A critical knowledge gap exists regarding the impact of teacher qualifications on factors such as student engagement, as noted by Wedel (2021), who utilised regression analysis, classroom observations, and questionnaires. Methodological concerns are evident in studies like Takunyaci and Aydin (2021), which employed non-random sampling methods, limiting

generalizability, while Fosu et al. (2023) highlighted sampling bias in their Ghanaian study with a sample size of 300 drawn from a target population of 1200. Contextual limitations are present in various studies, such as Tan and Saligumba (2019), which lacked diverse data sources on test reliability, and Puspita (2019), whose small sample size constrained her examination of English test validity in Indonesia. Additionally, Takunyaci's (2021) research, involving interviews with only five teachers, raises concerns about insight breadth. An empirical gap exists in linking structured moderation practices to student performance in national assessments. To address these issues, the current study employs classical test theory and a convergent design under a mixed-methods approach, integrating interviews, document analysis, and questionnaires to enhance academic achievement and contribute to effective mathematical practices.

## 8. Research Design and Methodology

The study employed a convergent mixed-methods design, collecting qualitative and quantitative data simultaneously, which were analysed separately before being merged for discussion (Creswell & Creswell, 2023). The target population included 3,768 individuals from 197 public secondary schools, comprising 197 Heads of Schools, 3,060 Form Two students, 288 Mathematics teachers, and 16 ZEC Subject Coordinators across three regions and seven districts. Sampling techniques included total sampling, criterion purposive sampling, stratified random sampling, and censoring. Quantitative data were gathered through questionnaires, while qualitative data were collected via interviews and document analysis guides (Creswell & Creswell, 2018; Ogula et al., 2018; Cohen et al., 2018), with a sample size of 436 respondents. The process ensured proportional representation by considering gender and location in the selection process. To ensure balanced representation across the Urban-West, South, and North regions, the study included ZEC subject coordinators for their expertise in exam preparation. At the same time, school heads were purposively selected based on relevant criteria.

Descriptive statistics were applied to quantitative data, and thematic analysis was utilised for qualitative data, with the validity of research instruments established through expert reviews and reliability checks demonstrating coefficients exceeding the acceptable threshold of 0.7. In the pilot study, 39 participants from rural and urban schools, comprising 30 students, six mathematics teachers, and three school heads from selected public secondary schools, were involved to refine instruments and assess reliability. Rigour and ethical integrity were emphasised through peer review and triangulation measures, enhancing the credibility of results by evaluating findings across multiple sources (Yin, 2018; Okendo et al., 2020). The questionnaire's Likert-type items achieved a Cronbach's Alpha coefficient of 0.72 for teachers. Qualitative data were analysed thematically and presented with direct quotations. In contrast, quantitative data were analysed using inferential statistics through linear regression at a significance level of 0.05 ( $\alpha$ ), and descriptive statistics were presented in tables, frequencies, mean, and percentage using SPSS version 22. Ethical considerations, including informed consent and confidentiality, were rigorously followed throughout the research process to foster trust.

## 9. Data Presentation and Discussion of Findings

This section consists of data collected from Mathematics teachers, Heads of Schools (HoSs, and ZEC Subject coordinators through questionnaires, interview guides, and document review guides. To obtain the required data, ZEC subject coordinators and HoSs were interviewed, and questionnaires for Students and Mathematics Teachers are summarised in Table 1.

**Table 1:** Teachers' Responses on the Extent to Which Moderation of Mathematics Teacher-Made Test Enhances Mathematics Performance of Form Two Students (n = 41)

S/N	Statements	VLE		LE		ME		HE		VHE		Mean
		F	%	F	%	F	%	f	%	f	%	
i.	The moderation process ensured consistency and fairness in grading the tests	0	0.0	2	4.9	23	56.1	10	24.4	6	14.6	3.49
ii.	Test Moderated provides a fair and accurate assessment of students' understanding of mathematics concepts in the national performance	0	0.0	0	0.0	13	31.7	20	48.8	8	19.5	3.88
iii.	The moderation process ensured consistency in grading and scoring among different markers on the test to improve students' academic performance.	0	0.0	0	0.0	15	36.6	16	39.0	10	24.4	3.88
iv.	Moderation of the math test ensures the reliability and validity of a national assessment of students' mathematics skills	1	2.4	2	4.9	10	24.4	16	39.0	12	29.3	3.88

v.	The moderation process ensured the teacher's commitment to maintaining high standards and fair tests for all students to improve academic performance.	0	0.0	3	7.3	12	29.3	17	41.5	9	22.0	3.78
vi.	The moderation process of math tests ensured that the test scores reflected the students' abilities in the national exam.	0	0.0	3	7.3	15	36.6	12	29.3	11	26.8	3.76
vii.	The moderation process improved the accuracy and reliability of the test results and the national assessment	0	0.0	2	4.9	8	19.5	23	56.1	8	19.5	3.90
viii.	The Moderation of the Test ensures the assessment criteria are clearly defined and transparent for all students	0	0.0	3	7.3	11	26.8	19	46.3	8	19.5	3.78
ix.	The moderation process of tests promotes a comprehensive coverage of math content and skills for improving students' academic performance	0	0.0	3	7.3	13	31.7	14	34.1	11	26.8	3.80
x.	The moderation of the test ensured that the result feedback provided valuable insights for future improvement of students' academic achievement	0	0.0	1	2.4	11	26.8	12	29.3%	17	41.5	4.10
<b>Grand Mean</b>												<b>3.82</b>

**Source:** Field data (2024). Key: 1=Very low Extent (VLE), 2=Low Extent (LE), 3=Moderate Extent (ME), and 4=High Extent (HE), 5=Very high Extent (VHE). Mean = (Very Lower extent = 1.00-1.79; Low extent = 1.80- 2.59; moderately extent = 2.60-3.39; High extent =3.40-4.19 and Very high extent = 4.20-5.00.) 0-20%= Very Low Minority, 21%-40%= Minority, 41%-60%= moderate majority, 61%-80%= High Majority and 81%-100%= Very High Majority.

*Test Moderated provides a fair and accurate assessment of students' understanding of mathematics concepts in national performance.*

Data in Table 1 indicate responses of Mathematics teachers regarding the extent to which the Moderation of Mathematics TMT Enhances Mathematics Performance of Form Two Students. Test moderation effectively offers a fair and accurate assessment of students' understanding of mathematics concepts within the context of national performance. A moderate majority (68.3%) of Mathematics teachers expressed a high extent, indicating their belief that moderated tests are reliable reflections of students' understanding of mathematical concepts. Meanwhile, a minority (31.7%) of teachers rated it as a moderate extent. These are supported by a mean score of 3.88, aligning with a high extent of agreement. This mean score suggests a strong consensus among Mathematics teachers regarding the fairness and accuracy of moderated assessments. Thus, this finding confirms that Test moderation enhances the assessment process meaningfully, providing a trustworthy measure of student understanding.

The researcher further triangulated data by face-to-face interviews with the heads of schools (HoS) and ZEC subject coordinators and document analysis.

Furthermore, during the Head of School interview, he said, *"The school-based test is not moderated; it is just prepared according to what has been taught, not moderated, weekly test, monthly test, and other tests. But the mock exam was moderated by the regional educational committee"* (HoS-S, personal communication, 12 September 2024). Again, during the interview with one of the ZEC subject coordinators, they had this to say, *"Moderation of TMT in mathematics for Form Two students is minimal and inconsistent, while mock examinations receive regional-level moderation"* (ZECSC-2, personal communication, 29 August 2024). During the document analysis guide conducted at School "C," the researcher found no minutes of the Moderation panel. The researcher observed no moderation records available, making it hard to establish whether moderation was done (DAoS-C, 08 September 2024).

Information collected from interviews with Heads of Schools (HoS), ZEC subject coordinators, and document analyses conducted at School "C" reveals disparities between Mathematics teachers from self-report and interviews in moderation assessment practices. Through interviews and document analysis guides, it was found that the school-based test is not moderated like a moderation panel, just submitted academic master and observed in a general view before the testing, indicating insufficient inspection for routine assessments. In contrast, the regional educational committee formally moderates mock examinations. However, a high majority (68.3%) of the mathematics teachers rated it as high. The test provides a fair and accurate assessment of students' understanding of mathematics concepts in national performance. In contrast, a minority (31.7%) of teachers rated it moderate in extent. These findings highlight the urgent need for improved moderation practices to ensure fair and accurate measurement of student performance, aligned

with the study by Puspita (2019), which emphasised that while tests may be valid and dependable, assessment standards have considerable room for improvement.

*Moderation of mathematics tests ensures the reliability and validity of a national assessment of students' mathematics skills.*

The data in Table 1 reveal that the Moderation of Mathematics tests ensures the reliability and validity of a national assessment of students' mathematics skills. A substantial majority (68.3%) of mathematics teachers expressed a high to very high extent of agreement regarding their confidence in the moderation process, which is essential for maintaining consistent assessment standards. In contrast, a minority (24.4%) of teachers rated their agreement as moderate, while a tiny minority (7.3%) of teachers indicated a very low extent of agreement. The mean score of 3.88 further reinforces the firm belief among teachers in the usefulness of moderated assessments in keeping the integrity of assessments. This mean score suggests that the Moderation of Mathematics tests ensures the reliability and validity of a national assessment of students' mathematics skills. These findings suggest that the moderation of mathematics tests plays a critical role in providing reliable and valid assessments for assessing students' mathematical competencies, highlighting the necessity for continued support and enhancement of these moderation processes.

Furthermore, during the interview of the Head of School, "C" had this to say:

The absence of moderation for regular mathematics Teacher-Made Tests (TMT) reduces quality and may negatively impact students' academic performance and engagement. The academic master observes the tests in general before testing, highlighting the need for improved assessment practices in schools (HoS-C, personal communication, 27 September 2024).

Moreover, during the interview with the ZEC subject coordinator-10 remarked and had this to say:

The moderation process every teacher does themselves, not by panel-wise in school-based tests, such as all weekly, monthly, and other tests, usually supervised by the subject teacher, even to moderate the test, the test is submitted to the academic committee for general observation. Most of the time, school-based tests are submitted late to the academic committee. In contrast, mock tests are prepared by the Regional Educational Committee and moderated by the Regional Educational Centre instead of an individual teacher (ZECSC-10, personal communication, 29 August 2024).

Information collected from interviews with the Head of School (HoS) "C", ZEC subject coordinators, and document analyses conducted at School "B" reveals significant concerns about the moderation of mathematics assessment tools. Through interviews and document analysis guides, it was found that the academic master observes the tests in general before testing, moderation typically relies on individual teachers rather than a panel, leading to inconsistencies and delays in submitting school-based tests to the academic committee. However, a high majority (68.3%) of the mathematics teachers rated it as high. The Moderated mathematics tests ensure the reliability and validity of a national assessment of students' mathematics skills. While a minority (24.4%) of the teachers rated it moderate, 7.3% rated it as low. These findings highlight the urgent need for improved moderation practices to enhance assessment quality and support better student performance. These findings align with Fosu et al. (2023), who emphasised that understanding moderating and mediating factors can provide a more comprehensive view of the dynamics influencing mathematics achievement in educational contexts.

*The moderation process ensured the teacher's commitment to maintaining high standards and fair tests for all students to improve academic performance.*

The data in Table 1 reveal that the moderation process ensured the teacher's commitment to maintaining high standards and fair tests for all students to improve academic performance. A high majority (63.5%) of Mathematics teachers indicated a high extent of agreement. The minority (29.3%) of mathematics teachers were rated as moderate extent (ME), while a very low minority (7.3%) of mathematics teachers were rated as having a low extent of agreement. The mean score of 3.78 further underscores the widespread recognition among teachers of the value of moderation in promoting fairness and quality in assessments. This mean score suggests that the moderation process ensured the teacher's commitment to maintaining high standards and fair tests for all students to improve academic performance. This finding indicates that the moderation process is crucial in fostering equitable assessment practices, thereby supporting enhancing students' academic success in the mathematics national assessment.

Also, during the interview of the Head of School, "A" had this to say:

The school-based test is not moderated; it is just prepared according to what has been taught and submitted to the academic committee for general observation before setting the date for testing. However, the mock examinations are moderated by the regional educational committee (HoS-A, personal communication, 18 September 2024).

Moreover, from this perspective, a ZEC subject coordinator-4 remarked and had this to say:

Teachers face challenges in moderating tests due to an insufficient teachers relative to the large number of classes and students, leaving little time for additional activities. Equalising the teacher-to-student ratio could improve the Moderation of Mathematics TMT, ensuring fair and reliable assessment tools aligned with learning objectives. Standardised tests would minimise grading biases and foster collaboration among students and teachers. These, in turn, would provide accurate feedback, helping students identify strengths and weaknesses, enhance engagement, and ultimately improve their performance in mathematics (ZECSC-4, personal communication, 27 August 2024).

During the document analysis guide conducted at School "F," the researcher found no minutes from moderation panels in the reviewed minute document files. The absence of these minutes was noted during the observation (DAoS-F, 27 August 2024).

Information from interviews with Head of School (HoS) "A", ZEC subject coordinator-4, and document analyses conducted at School "F," reveals significant concerns about the moderation of mathematics assessments. Through interviews and document analysis guides, it was found that there is a lack of proper moderation for school-based tests. At the same time, the Regional Educational Committee oversees mock examinations, and insufficient teacher-to-student ratios hinder effective moderation, suggesting that better distribution of teachers could enhance academic performance through fair assessments. Despite these issues, a high majority (63.5%) of the Mathematics teachers rated it to a high extent about the moderation process, ensuring the teacher's commitment to maintaining high standards and fair tests for all students to improve academic performance. While a minority (29.3%) of the teachers rated it moderate, 7.3% rated it as low. These findings underscore the urgent need for improved moderation practices, particularly regarding teacher-to-student ratios, which align with the study by Wedel (2021) showing that better-qualified teachers lead to increased test quality, instruction, and improved test scores.

*The moderation process of mathematics tests ensured that the test scores reflected the students' abilities in the national exam.*

The data in Table 1 reveal that the moderation process of Mathematics tests ensured the test scores reflected the students' actual abilities toward the national examination. Here, a moderate majority (56.1%) of Mathematics teachers expressed a high to very high extent of agreement, indicating that the test results accurately represent student capabilities. The minority (36.6%) of mathematics teachers were rated as having a moderate extent (ME), with a very low minority (7.3%) indicating a low extent of agreement. The mean score of 3.76 reflects a favourable perception of the accuracy of test scores. This mean score suggests that the moderation process of Mathematics tests ensured the scores reflected the students' abilities toward the national examination. The moderation process is crucial in validating that mathematics test scores genuinely reflect students' abilities as they prepare for national assessments.

Furthermore, in this context, a Head of School T<sub>1</sub> had this to say:

In some ordinary secondary schools, there is no moderation panel; assessments are prepared based on what has been taught, including weekly and monthly tests, which are then submitted to the academic master for arrangement and administration to the students. However, the mock examinations are indeed moderated by the regional educational committee (HoS-T<sub>1</sub>, personal communication, 08 October 2024).

Moreover, from this perspective, a ZEC subject coordinator-7 remarked and had this to say:

Inadequate Moderation for TMT, as it was submitted to the academic teacher just observing the general and then stored to wait for testing, also most of the tests submitted to the academic master on the deadline date to submit to the academic master (ZECSC-7, personal communication, 20 August 2024).

Information from interviews with Head of School (HoS) "T<sub>1</sub>", ZEC subject coordinator-7, and document analyses conducted at School "G" reveals that the Interviews with Head of School T<sub>1</sub> and ZEC Subject Coordinator-7 highlight critical issues in the Moderation of TMT in ordinary secondary schools. The interviews and documents analysis guides it was found that lack of a moderation panel, resulting in assessment tools being submitted directly to the academic master without prior review, tests are often submitted at the last minute due to time constraints faced by mathematics teachers and found no recorded minutes of moderation panel were available for mathematics tests. However, a moderate majority (56.1%) of the Mathematics teachers rated it as a high extent about the moderation process of mathematics tests, ensuring that the test scores reflected the students' abilities in the national exams. While a minority (36.6%) of teachers rated it moderate, 7.3% rated it as low. These findings raise concerns about the credibility and adequacy of mathematics TMT, potentially undermining instructional test strategies and negatively affecting student learning outcomes in national assessments. This finding aligns with the study by Takunyaci and Aydin (2021) in Turkey, who found that the stresses in credible tests indicate the importance of collaborative moderation for ensuring the validity and reliability of assessment tools.

*The moderation process improved the accuracy and reliability of the test results and national assessments.*

The data in Table 1 reveal that the moderation process improved the accuracy and reliability of the test results and national assessment. A high majority (75.6%) of the mathematics teachers rated the high extent of agreement as being between a high and very high extent of agreement about the moderation process to enhance the integrity of assessment results. While a very low minority (19.5%) of the mathematics teachers rated it moderate (ME), 4.9% rated it as low extent, suggesting that some teachers view the process as beneficial. The mean score of 3.90 further reflects a positive perception among educators regarding the impact of moderation on accuracy and reliability in assessments. This mean score suggests that the moderation process improved the accuracy and reliability of the test results and national evaluation. This finding indicates that the moderation process enhances the quality and trustworthiness of test results within the context of national assessments.

Again, during the interview with one of the HoS-T<sub>2</sub> had this to say:

Inadequate standard moderation process in ordinary secondary schools is just prepared according to what has been taught, weekly tests, monthly tests, and other tests, and then submitted to the academic master. The test is arranged for the students. However, the mock exam was moderated by the Regional Educational Committee (HoS-T<sub>2</sub>, personal communication, 27 August 2024).

On the other hand, during interview of ZEC subject coordinator "6" remarked and had this to say, *"The moderation has been done by academic committee observing the test in general format and some of criteria of TMT seen to be not quality compared to FTNE"* (ZECSC-6, personal communication, 29 August 2024).

During the document analysis guide conducted at School "I", the researcher found no minutes available from the minute document file while seated with the Mathematics moderation panels. This lack of minute documents indicates poor means of moderating for Mathematics TMT. The absence of a structured moderation process raises concerns about the credibility of Mathematics TMT, as it compromises the necessary checks and balances for an effective moderation process in assessment tools constructed by teachers (DAoS-1, 17 September 2024).

Information collected from interviews with HoS "T<sub>2</sub>", ZEC subject coordinator-6, and document analyses conducted at School "I," reveals significant inadequacies in the Moderation of TMT in secondary schools. Through interviews and a document analysis guide, the tests are prepared based on the material taught and submitted directly to the academic master. Mock exams receive an overview from the Regional Educational Committee. However, a high majority (75.6%) of the mathematics teachers rated the moderation process as high, indicating that it improved the accuracy and reliability of the test results and national assessments. A very low minority (19.5%) of the mathematics teachers rated it as moderate extent, and 4.9% were low extent. These findings raise concerns about the credibility of the mathematics TMT moderation process. These findings diverge from classical test theory by Rasch and Lord, which emphasises validity and reliability through reviews of test items constructed to ensure assessment tools accurately measure student performance.

Generally, the study reveals that the Moderation of Mathematics teacher-made tests (TMT) significantly enhances student performance. A majority (68.3%) of the Mathematics teachers believe that moderated assessments accurately reflect student understanding, as indicated by an overall mean score of 3.88. These

suggest a strong confidence in the effectiveness of moderated assessments, contributing to their credibility and reliability. However, interviews and document analyses highlight inconsistencies in moderation practices. Heads of Schools and ZEC subject coordinators pointed out that regular school-based tests often lack formal moderation, depending instead on individual teacher preparation without a structured framework. The absence of recorded moderation panel minutes further emphasises the urgent need for improved practices to ensure fair and accurate measurement of student capabilities. These findings highlight the critical importance of effective moderation in enhancing the quality and trustworthiness of mathematics assessments. These findings align with the need for assessment standards that accurately measure student learning outcomes and include valid, reliable, and consistent practices.

### **The Extent to which Quality of Moderated Teacher-Made Test of Mathematics Subject Enhances the Academic Achievement of Students**

The data were collected from Students Heads of Schools (HoSs) and ZEC Subject coordinators through questionnaires, interview guides, and a document review guide. To obtain the required data, ZEC subject coordinators and HoSs were interviewed, and questionnaires for Students are summarised in Table 2

**Table 2: Students' Responses in the extent to which Quality of Moderated Teacher-Made Test of Mathematics Subject Enhances Academic Achievement of Students (n = 303)**

N/S	Statements	VLE		LE		ME		HE		VHE.		Mean
		f	%	f	%	f	%	F	%	f	%	
	The quality of Mathematics tests links learning objectives to the national level.	99	32.7	73	24.1	59	19.5	34	11.2	38	12.5	2.47
	The test quality ensures the questions are fair and unbiased.	103	34.0	36	11.9	66	21.8	33	10.9	65	21.5	2.74
	Quality tests its results, which helps me identify areas to improve my performance	64	21.1	42	13.9	49	16.2	54	17.8	94	31.0	3.24
	Well-designed math tests correct the difficulty level of my current skills.	67	22.1	41	13.5	53	17.5	66	21.8	76	25.1	3.14
	A well-standardised test improves students' ability to solve math problems in the national exam	54	17.8	43	14.2	51	16.8	58	19.1	97	32.0	3.33
	The grading criteria used in tests are fair and consistently applied in the final exam.	78	25.7	57	18.8	59	19.5	48	15.8	61	20.1	2.86
	Well-designed tests improve students' critical thinking & apply mathematical concepts in the national exam	55	18.2	41	13.5	68	22.4	56	18.5	83	27.4	3.23
	Well-designed math tests help students identify their strengths and weaknesses before sitting for the national exam	66	21.8	35	11.6	46	15.2	47	15.5	109	36.0	3.32
	The time allotted in the test is sufficient and helps to attempt all the required problems in the final exam.	67	22.1	60	19.8	63	20.8	32	10.6	81	26.7	3.00
	Quality Math tests help motivate students to prepare attentively for the national exams.	72	23.8	35	11.6	34	11.2	38	12.5	124	40.9	3.35
<b>Grand Mean</b>												<b>3.07</b>

**Source:** Field data (2024). Key: agreement 1=Very low Extent (VLE), 2=Low Extent (LE), 3=Moderate Extent (ME), and 4= High Extent (HE), 5=Very high Extent (VHE). Mean = (Very Lower extent = 1.00-1.79; Low extent = 1.80- 2.49; Moderate extent = 2.50-3.39; High extent =3.40-4.19 and Very high extent = 4.20-5.00.) 0-20%= Very Low Minority, 21%-40%= Minority, 41%-60%= moderate majority, 61%-80%= High Majority and 81%-100%= Very High Majority.

*The quality of Mathematics tests links learning objectives to the national level.*

Data in Table 2 shows that a moderate majority (56.8%) of Form Two students perceive a low alignment between the quality of Mathematics teacher-made tests (TMT) and national curriculum objectives, with only

19.5% rating the tests as moderately aligned and 23.7% expressing confidence in their standard quality. The mean score of 2.47 indicates a low agreement among students that the tests effectively link learning objectives to the national level, emphasising a significant perception gap. The findings suggest a considerable gap in students' perceptions of assessment to align with the national curriculum standard, highlighting the need for further action to improve students' academic achievement.

The researcher further triangulated data by conducting a face-to-face interview and document analysis with the heads of schools (HoSs) and ZEC subject coordinators.

Furthermore, during the interview of the Head of School "E", he had this to say, *"The absence of quality in regular mathematics teacher-made tests adversely affects students' academic performance and engagement, emphasising the necessity for enhanced assessment practices in schools"* (HoS-E, personal communication, 20 September 2024). During the interview with the head of school-N, they said, *"Teacher-made mathematics tests for Form two students vary in quality and consistency, while mock examinations consistently meet regional quality standards"* (HoS-N, personal communication, 17 September 2024). During the document analysis conducted at School "D," the researcher found no samples of tests moderated into standard quality; there were no meeting note records from the moderation panels (DAoS-D, 20 September 2024).

Information collected from interviews with HoS "E", ZEC subject coordinator, and document analysis conducted at School "D" reveals significant issues in the Moderation of Mathematics TMT. Through interviews and document analysis guides, it was found that the lack of moderation for regular tests diminishes quality and could negatively impact students' academic performance. The Moderation for Form Two students is minimal and inconsistent, with only Mock Exams receiving oversight from the Regional Educational Committee. However, a minority (23.7%) of the Students rated it as a high extent about the quality of Mathematics tests, linking learning objectives to the national level. In contrast, a very low minority ( 19.5%) of the students rated it as Moderate Extent, and a moderate majority (56.8%) of the students rated it as low extent. These findings underscore gaps in test quality and the credibility of assessment tools, emphasising the need for structured moderation to ensure that Mathematics TMT is valid, reliable, and effectively measures student learning outcomes. The findings align with Lord and Novick's classical test theory (1968), which emphasises the importance of validity and reliability in assessments and the necessity for systematic moderation to enhance assessment efficiency.

*Test quality ensures the questions are fair and unbiased.*

Data in Table 2 indicates that the test quality ensures the test questions are fair and unbiased. In this case, a very low minority (32.4%) of the students reported a high to very high extent of agreement. These suggest that some students believe the test's standard effectively filters out bias, thereby enhancing the fairness of the assessments. While a minority, 21.8% of the students rated their satisfaction with the objectivity of the test questions, a substantial moderate majority (45.9%) disagreed to a certain extent about fairness and impartiality. The mean score of 2.74 places this perception within the moderate extent category. The mean scores highlight a need for improvement in the moderation process to alleviate students' concerns about test fairness. These findings reveal a gap between students' expectations and experiences, emphasising the importance of enhancing objectivity in assessments to create a fair and positive learning environment.

Again, during the interview with one of the HoS-C, I had this to say:

In our school, test quality has been made for school-based tests like weekly and monthly assessments, but it is not similar to the quality of standardised tests. Once terminal tests are submitted to the academic master, the academic committee reviews them before scheduling the School-based Exam. The Regional Educational Committee moderates Mock Exams for a standard quality (HoS-C, personal communication, 10 September 2024).

In addition to that, during interview with head of school-"I" who had this to say, *"The quality of mathematics teacher-made tests (TMT) is limited and varies, as efforts to ensure their validity through general observation by the academic committee for Form two students are insufficient, while mock examinations receive regional-level quality assurance"* (HoS-I, personal communication, 08 October 2024).

Moreover, from this perspective, a ZEC subject coordinator-4 remarked and had this to say, *"Quality of teacher-made tests in mathematics for form two students is limited and varies, while mock examinations receive regional-level moderation"* (ZECSC-4, personal communication, 29 August 2024).

Information collected from interviews with HoS "C" as well as ZEC subject coordinator and document analysis conducted at School "B" reveals the Head of School "C" and "I" noted the inadequacies in the moderation process for school-based assessments, stating that while the academic committee reviews terminal tests, only mock examinations receive oversight from the regional educational committee. The ZEC subject coordinator highlighted the minimal and inconsistent quality standards for teacher-made mathematics tests for Form Two students. Document analysis indicated a lack of meeting records from the moderation panel, reflecting poorly structured moderation practices. However, a minority (32.4%) of the Students rated it as a high extent about the test quality, ensuring the questions are fair and unbiased. In contrast, a very low minority (21.8%) of the students rated it as Moderate Extent, and a moderate majority (45.9%) of students rated it as low extent. Although students perceive some fairness in the assessments, rated at a moderate extent, there is significant room for improvement in test quality to ensure consistent assessment scores. These findings underscore gaps in the quality of mathematics assessments and the need for systematic approaches to enhance assessment practices across schools. These align with research by Fosu et al. (2023), which emphasises the importance of identifying standard quality assessments and their influencing factors for improving mathematics academic achievement.

*The grading criteria used in tests are fair and consistently applied in the final examination.*

Data in Table 2 indicates that the Grading criteria used in tests are fair and consistently applied in the final examinations. A minority (35.9%) of the students rated as a high to very high extent of agreement, indicating that there is a perception among some students that grading criteria are applied equitably, which is crucial for ensuring a transparent and trustworthy assessment process. While a very low minority (19.5%) of the students rated it as a moderate extent (ME), this suggests that students find the grading somewhat consistent. Conversely, a moderate majority (44.5%) of students indicated a low extent of agreement, reflecting concerns about fairness and consistency in how they perceive their grades being awarded. The mean score of 2.86 places this perception within the moderate extent; this score emphasises the need for improved transparency and perceived fairness in grading practices to enhance student confidence in the testing process. The findings highlighted the necessity for clarity in grading and scoring criteria to effectively address students' concerns about improving the tests' quality and fairness.

Again, during the interview with one of the heads of School (HoS), H had this to say:

There is no standard quality made at the ordinary secondary school level. Tests are prepared based on taught content, class activities, and various assessments, which the academic master then arranges. However, mock examinations are moderated by the regional educational committee (HoS-H, personal communication, 08 October 2024).

During an interview with one of the heads of School (HoS)-M, who said, *"There is low quality moderation of terminal test by the examination committee at the ordinary secondary school level. Tests are prepared based on taught content, class activities, and various assessments, which are then arranged by the academic master"* (HoS-M, personal communication, 12 October 2024). Moreover, from this perspective, a ZEC subject coordinator-5 remarked and had this to say, *"Of course, the standards quality of Mathematics teachers made test provides fair, reliable assessments aligned with learning objectives. Also, it reduces grading biases and promotes collaboration among teachers"* (ZECSC-5, personal communication, 27 August 2024).

Information collected from interviews with the Head of School (HoS) "M," ZEC Subject Coordinator-5, and document analysis at School "J" reveals critical issues in the moderation and quality of the test. Through interviews and document analysis guides, it was found that tests are based on taught content and arranged by the academic master. At the same time, mock examinations receive oversight from the regional educational committee. There is also insufficient quality in Mathematics TMT, raising concerns about reliability and credibility. Form Two students disagreed on the fairness and consistency of their grades. However, a minority (35.9%) of the Students rated it as a high extent that the grading criteria used in tests are fair and consistently applied in the final examination.

In contrast, a very low minority (19.5%) of the students rated it as Moderate Extent, and a moderate majority (44.5%) of the students rated it as low extent. This inconsistency underscores the need for improved moderation practices to enhance student perceptions of equity in assessment. The findings indicate an urgent need for structured mathematics assessment tools to enhance validity and reliability, aligning with

educational objectives. These align with Tan and Saligumba (2019), who stressed the importance of cognitive levels, item analysis, and distractor analysis in enhancing test quality, ultimately strengthening the moderation process and fostering student confidence in assessments.

*Well-designed tests improve students' critical thinking & apply mathematical concepts in the national examination.*

Data in Table 2 indicates that the well-designed tests improve students' critical thinking and application of mathematical concepts in exams. A moderate majority (45.9%) of the students rated it as a high extent that tests enhance their critical thinking skills, which is essential for mathematics success. Conversely, a minority (22.4%) of the students rated the improvement moderately, suggesting some find it somewhat effective. In comparison, a minority (31.7%) of the students expressed a very low extent, indicating uncertainty about the tests' efficiency in adopting analytical skills. With a mean score of 3.23 as a moderate extent, students recognised the benefits of well-structured tests but also highlighted the need for further improvements in test design. These findings suggest that enhancing test design could better encourage critical thinking skills among students.

Again, during the interview with one of the HoS-J, who had this to say:

The schools do not moderate the school-based tests, such as weekly, monthly, and other tests. When a test, such as a terminal test, is submitted to the academic master, the academic committee observes the formality of the test before the arranged timetable of the based examination. But the mock examination was moderated by the regional educational committee" (HoS-J, personal communication, 19 September 2024).

Moreover, during an interview with a ZEC subject coordinator, the coordinator said, *"Actually, the teacher prepared the test and submitted it to the academic committee without being observed by the mathematics panel before the testing"* (ZECSC-3, personal communication, 26 August 2024).

Information collected from interviews with HoS "J", ZEC subject coordinator-3, and document analysis conducted at School "K" reveals that during an interview, through interviews and document analysis guides, it was found that the school has insufficient moderation of assessments, like weekly and monthly tests. Terminal tests are checked for formality by the academic committee before scheduling, while the Regional Educational Committee moderates mock examinations. However, a slight moderate (45.9%) of the students rated it as high extent about the well-designed tests improving students' critical thinking & applying mathematical concepts in the national examination, while a minority (22.4%) of the students rated it as moderate extent, and 31.7% were low extent. This absence raises concerns about the credibility of mathematics assessments and underscores the need for a structured moderation process. The findings highlighted significant gaps in the current moderation practices that require immediate attention to ensure adequate assessment and instructional planning. These findings are related to the study by Takunyaci and Aydin (2021) in Turkey, which found that collaborative judgments among raters positively influence the reliability of an evaluation system.

*Quality Math tests help motivate students to prepare attentively for the national exams.*

Data in Table 2 indicates that the quality of Mathematics tests motivates students to prepare attentively for national exams. A moderate majority (53.4%) of students demonstrated a high to very high extent of agreement, suggesting that tests were a significant motivating factor in their examination preparation activities. The very low minority (11.2%) of the students rated it as a moderate extent (ME), indicating that a smaller fraction of the students believe the motivation provided is somewhat effective. However, a minority (35.4%) of the students rated a low extent of agreement, showing that some students may not feel adequately encouraged by the moderation process. The mean score of 3.35 reflects a moderate extent, highlighting that while many students find motivation in moderated tests, there remains an opportunity for further improvements in this area. These findings suggest that enhancing the motivational aspects of the moderation process could further boost students' preparedness for examinations. By creating a supportive environment with constructive feedback and recognition, Mathematics teachers raise student confidence and commitment, leading to better outcomes and a positive learning culture.

Moreover, during interview with one of the head of school-I who had this to say, *"Due to a lack of standard quality in ordinary secondary schools, regular tests are prepared based on what has been taught and submitted to the academic master without proper review; in contrast, mock examinations are moderated by the regional educational committee"* (HoS-I, personal communication, 08 October 2024)

Information collected from interviews with the Head of School (HoS) "A," ZEC Subject Coordinator-1, and document analysis at School "L" reveals significant gaps in the moderation processes for assessments. Through interviews and document analysis guides, it was found that tests are prepared based on the curriculum and submitted without adequate moderation, with only Mock Examinations reviewed by the Regional Educational Committee. Well-designed and moderated Mathematics tests should closely align with the FTNA. Despite those issues, a moderate majority (45.9%) of the students rated it as high extent about the Quality Math tests help motivate students to prepare attentively for the national exams, while a very low minority (11.2%) of the students rated it as moderate extent and a minority (35.4%) of the students were low extent. This information suggests that these tests' enhancements could increase fairness and better meet students' learning needs, fostering a more supportive educational environment. Additionally, quality Mathematics tests motivate students to prepare more effectively for national examinations. These findings underscore the urgent need for improved assessment practices, aligning with Puspita (2019), which emphasises the importance of enhancing moderation processes for valid assessment tools.

## 10. Hypothesis Testing

The researcher tested the hypothesis using a simple linear regression analysis to examine the relationship between Mathematics TMT scores and FTNA scores regarding students' academic achievement. The aim was to determine if the quality of Mathematics TMT scores affects FTNA scores.

### Hypotheses:

**Null Hypothesis (H<sub>0</sub>):** There is no significant relationship between Mathematics TMT scores and FTNA scores in academic achievement.

**Alternative Hypothesis (H<sub>1</sub>):** There is a significant relationship between Mathematics TMT scores and FTNA scores in academic achievement.

The hypothesis was tested at a significance level of 0.05, indicating a 5% probability that the results occurred by chance. A p-value less than 0.05 suggests an actual correlation between the moderated TMT scores and national academic performance scores.

### Assumptions of Simple Linear Regression:

**Linearity:** The relationship between TMT (independent variable) and FTNA (dependent variable) must be linear. ANOVA results (Table 4) indicate a p-value < 0.05, confirming linearity. **Independence:** Observations must be independent. **Homoscedasticity:** Residual variance should be constant across all independent variable levels. **No Multicollinearity:** Independent variables should not be highly correlated. **Normality:** Residuals should be approximately normally distributed. The Shapiro-Wilk test results (Table 3) confirmed normality with p-values of .153 and .231, greater than 0.05.

**Table 3**

Shapiro-Wilk Test of Normality						
Scores	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	Df	Sig.
Teacher-made tests	.288	5	.200 <sup>*</sup>	.835	5	.153
Form two national assessment	.296	5	.176	.861	5	.231
*Lower bound of the true significance.						
a. Significance Correction						

**Source:** Field data (2024)

Data in Table 3 indicates that the p-values were (.153, and .231), greater than the significance level of 0.05. Thus, the data were normally distributed, allowing continuation with hypothesis testing.

**ANOVA Results:** Table 4 summarises the ANOVA results for hypothesis testing:

**Table 4:** The ANOVA Table Summarised Students' Responses for Hypothesis Testing from Simple Linear Regression (n = 303)

Model	Sum of Squares	Df	Mean Square	F	Sig.
1 Regression	9.571	1	9.571	1.545	.302 <sup>b</sup>
Residual	18.589	121	6.196		
Total	28.160	122			

**Source:** Field data (2024)

a. Dependent Variable: National Assessment scores in academic achievement

Table 4: The hypothesis testing results indicate no significant relationship between Mathematics TMT scores and FTNA scores, with a p-value of 0.302 leading to the acceptance of the null hypothesis. The regression analysis shows a correlation coefficient of  $R = 0.583$  and an  $R^2$  value of 0.340, suggesting that TMT scores explain only 34% of the variance in academic performance. The regression equation indicates an expected increase of approximately 0.884 points in FTNA scores for each additional point in TMT, but this is not statistically significant ( $p = 0.302$ ). These findings underscore the limited effectiveness of TMT in enhancing academic achievement and highlight the need for improved assessment practices and teacher training to align TMT with national standards. In contrast, the findings of the study by Hamisi and William (2020), who found that practical teachers' training, quality moderated tests, and resources significantly contribute to academic success, emphasise the importance of a comprehensive approach to improving students' mathematics performance. Generally, the gaps in perceptions between teachers and students regarding Mathematics TMT quality are better for effective communication about assessment expectations. Overall, a holistic approach that enhances assessment quality, provides teachers training on moderation, and improves effective feedback is essential for enhancing students' academic achievement in mathematics, with professional development crucial for enhancing teachers' assessment design skills and engagement and providing valuable final national assessment results.

## 10. Conclusions

Based on the findings, the following are the conclusions:

The findings highlight the critical need for effective moderation practices in Mathematics TMT, revealing notable disparities in perceptions between teachers and students regarding assessment quality. While teachers recognise a strong alignment of well-moderated tests with learning objectives, students show only moderate confidence in these tests' ability to identify their weaknesses. This finding underscores the urgent need for enhanced moderation strategies to improve assessment quality, ensure equitable practices, and boost student performance in national assessments. Additionally, the correlation between Mathematics TMT scores and FTNA scores emphasises the importance of ongoing support and active teacher involvement in the moderation process, aiming to create a more effective and motivating assessment environment that aligns with national curriculum standards.

## 11. Recommendations

Based on the conclusions, it is recommended that educational stakeholders, including teachers, donors, and academic authorities, prioritise the development of comprehensive moderation practices for Mathematics TMT. Regular teacher training on effective moderation techniques is essential to ensure consistency across schools. School administration should implement assessment procedures for teachers, school leaders, and students to enhance the final assessment in FTNE. Integrating student feedback can also inform improvements and foster a more supportive learning environment. Implementing these strategies in public schools will enhance Mathematics assessment quality and improve student motivation and academic outcomes in national assessments.

## References

1. Akendita, P. A., Arthur, Y. D., & Asare, B. (2025). Moderated effect of teachers' efficacy on the relationship between students' perception of mathematics and students' mathematics achievement. *International Journal of Didactical Studies*, 6(1), 28932 <https://doi.org/10.33902/ijods.202528932>
2. Asare, B. (2025). Moderating effect of teacher–student relationship and mediating role of self-efficacy on the relationship between peer-assisted learning and mathematics performance. *Front. Educ.* 9:1450499. doi: 10.3389/feduc.2024.1450499.
3. Baker, F. B., & Kim, S. H. (2018). *Item Response Theory: Parameter Estimation Techniques*. New York: CRC Press.
4. Black, P., & Wiliam, D. (2021). Assessment Framing: An Introduction. *Assessment in Education: Principles, Policy, Practice*, 28(1), 1-10. <https://doi.org/10.1080/0969594X.2020.1860630>
5. Cronin, Anthony, Mullen, Claire, Taylor, Laura K., Liu, Chang Uri (2021). evaluating the impact of mathematics support using moderation, <http://hdl.handle.net/10197/25988> 2024-05-21T09:48:15Z

6. Cohen, A., Glick, P., & Barlow, J. (2018). "The Role of Inclusion in Intergroup Relations: A Study of Ethnic Minorities," *Journal of Social Issues*, 10.1111/josi.12270, 345-364.
7. Creswell, J. W., & Creswell, D. (2023). *Research Designing, Qualitative, Quantitative, and Mixed Methods Research Approaches*. (6<sup>th</sup> Ed.). Sage Publications. In
8. Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative and Mixed Methods Research* (5<sup>th</sup> ed.). Sage Publications Inc
9. CIRA (2023) ZEC, Candidates' Items Responses Analysis in evaluation report for Form Two National Examinations in Basic Mathematics Subject (044), Zanzibar. Retrieved from [URL]
10. EDZ, (2019). (PDF) Education Policy Implementation: A Mechanism for Enhancing Primary Education Development in Zanzibar. Retrieved from researchgate.net
11. Eman, M. A. E. (2023). Relationship between examination anxiety and computer experiences.
12. Embretson, S. E., & Reise, S. P. (2021). *Item Response Theory for Psychologists*. Psychology Press. <https://doi.org/10.4324/9781315779530>
13. Fosu, A., Osei, A., & Agyemang, E. (2023). The mediation and moderation effect of mathematics interest and teaching quality between self-concept and mathematics achievement. *Journal of Educational Research*, 116(2), 123-135.
14. Fosu, M. A.; Boateng, Y. D., & Adu-Obeng, B. (2023). mediation and moderation effect of mathematics interest and teaching quality on the relationship between self-concept and mathematics achievement. *Journal of Mathematics & science teacher*, 3(1), em 024. <https://doi.org/10.29333/mathsciteacher/12622>
15. Hamisi, M., & William, F. (2020). Contributions of the Acquired Literacy and Numeracy Education Support (Lanes) Professional Development Skills on Enhancing Literacy Skills to Std I and II Pupils in Tanzania. *Malaysian Online Journal of Education*, Vol. 4(No. 1), 59–68.
16. Harahap, D. H., Sahrani, R., & Dewi, F. I. R. (2023). The Efficacy of Mathematics as a Moderator on the Relationship between Teacher Support and Middle School Students' Academic Engagement. Universitas Tarumanagara, Indonesia. Email: devihaerani88@gmail.com (\*Corresponding Author), rianas@fpsi.untar.ac.id, [fransiscar@fpsi.untar.ac.id](mailto:fransiscar@fpsi.untar.ac.id).
17. Jiang, H. (2023). Expectancy Theory and Student Motivation: A Review. *Journal of Educational Psychology*, 115(2), 245-260. <https://doi.org/10.1037/edu0000587>
18. Johnson, L. (2022). Impact of teaching quality on student achievement: student evidence. *Frontiers in Education*. [Link]
19. Kauffman, J. M., & Landrum, T. J. (2022). *Assessment and Learning: A Comprehensive Guide*. Routledge. <https://doi.org/10.4324/978100319123>
20. Lee, J., Kim, H., & Park, S. (2023). The impact of moderated assessments on student engagement and performance in mathematics. *International Journal of Educational Research*, 112, 101-110.
21. Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
22. Mhando, L. & Mgaiwa, B. (2021). Assessment practices in Tanzanian schools: Challenges and perspectives. *Journal of Education and Practice*
23. Mhando, L., Morrow, J., & Kapinga, A. (2020). Quality of Assessment in Tanzanian Education: A Review of Literature. *African Educational Research Journal*
24. Müller, A. (2023). Effect of test anxiety on students' academic performance in mathematics at the senior high school level. *Discover Education. European Journal of Psychology of Education*.
25. Ogula, P. A., Ogoti, E. O., & Maithya, P. M. (2018). *Educational and Social Science Research Advanced Short Courses: A Guide for Facilitators*. Kenya, Society of Educational Research and Evaluation
26. Owusu, A., Boateng, S., & Adomako, E. (2021). Mathematics interest and teaching quality mediate the relationship between self-concept and mathematics achievement. *Educational Review*.
27. Pang, M. F., Shimizu, H., & van der Sanden, J. (2019). Collaborative assessment moderation: Exploring perspectives and practices. *Assessment & Evaluation in Higher Education*.
28. Puspita, R. (2019). Analysis of validity and reliability of a teacher-made test: Case study at XI Grade of SMA N 6 Bengkulu. *Journal of Education and Practice*, 10(12), 45-52.
29. Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: University of Chicago Press

30. URT. (2021). National Examinations Council of Tanzania (NECTA), Guidelines on Assessment Procedures for Secondary Schools and Professional Level. The National Examinations Council of Tanzania: Dar es Salaam, Tanzania
31. VSO International. (2019). Leading Learning: A report on effective school leadership and quality education in Zanzibar. Retrieved from vsointernational.org
32. Schmidt, L., Garcia, R., & Thompson, P. (2023). Moderation processes in educational assessments: A European perspective. *Assessment in Education: Principles, Policy & Practice*.
33. Shimizu, Y. (2025). Learning engagement as a moderator between self-efficacy, math anxiety, use of diagrams, and complex plane problem-solving. Volume 21, Issue 2, em2586  
<https://doi.org/10.29333/ejmste/15956>
34. Setiawan, R. (2023). Analysing the validity and reliability of teacher-made tests in Indonesia. *Asia Pacific Journal of Education*.
35. Takunyaci, M., & Aydin, M. (2021). Implementation of a Group Moderation Assessment Model in an open-ended Math examination. *Journal of Mathematics Education*, 14(1), 1-15.
36. Tan, M., & Saligumba, R. (2019). Development of valid and reliable teacher-made tests for Mathematics. *Philippine Journal of Education*, 98(1), 23-34.
37. The WCED (2024) Internal Moderator report provides a per-question qualitative analysis of the learners' NSC examination question paper. Attribution-NonCommercial-NoDerivatives
38. Wedel, M. (2021). Instruction time and student achievement: The moderating role of teacher qualifications. *Educational Studies*, 47(3), 345-362
39. Wilson, M., & Allen, D. (2020). *Evaluating Educational Assessment: A Multi-Method Approach*. New York: Routledge.
40. World Bank. (2020). Zanzibar Improving Quality of Basic Education Project. World Bank Documents and Reports
41. Yılmaz, M. (2022). Group moderation in mathematics assessments: Enhancing reliability through collaborative evaluations. *International Journal of Education and Pedagogy*.
42. ZEC (2019-2023) Zanzibar Examination Council basic statistical report of form two national examinations in Zanzibar
43. ZEC (2023) Candidates' Items Responses Analysis in the evaluation report of form two national examinations in Zanzibar