Willing but Unequipped: Structural Barriers and Professional Development Needs in Higher-Order Thinking Assessment Design among Secondary Science and Mathematics Teachers in Tanzania

Mikidadi Juma Rwechoka*, and Shabani Juma Baya

Office of the Deputy Director General – Academic, Research and Consultancy, Tanzania Institute of Education, P. O. Box 35094, Dar es Salaam, Tanzania

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.100800431

Received: 20 August 2026; Accepted: 25 August 2026; Published: 07 September 2026

ABSTRACT

Tanzania's transition to competence-based education requires classroom assessment in science and mathematics to engage students in higher-order thinking (HOT), including analysing, evaluating, and creating, rather than merely recalling content. However, evidence remains limited on the factors shaping the gap between policy expectations and teachers' assessment practices, variations across subjects, and the support teachers require. This paper reports findings from a convergent parallel mixed-methods cross-sectional survey of 526 secondary school teachers of Biology (n = 151), Chemistry (n = 106), Mathematics (n = 169), and Physics (n = 100) in Lindi, Mtwara, and Njombe Regions, Tanzania, conducted between March and July 2025. Teachers' reported challenges in designing HOT assessment items and their support needs were analysed using descriptive statistics, chi-square tests of independence, binary logistic regression, and hybrid deductive-inductive content analysis of open-ended responses. Limited resources emerged as the leading barrier (57.4%), followed by lack of skills (38.8%), time constraints (34.8%), and curriculum restrictions (15.4%), with teachers reporting an average of 1.52 barriers. Barriers varied significantly across subjects. Mathematics teachers reported the highest number of concurrent barriers, while resource shortages were particularly severe in Physics and Mathematics, and curriculum restrictions were most pronounced in Mathematics. Teaching experience significantly influenced reported skills gaps: teachers with 21 or more years of experience reported the highest proportion of skills gaps (54.7%). Logistic regression confirmed that this association was independent of subject (OR = 2.54, 95% CI [1.36, 4.74]). Effect-size estimates indicated that the statistically significant associations were small to modest, with Cramer's V ranging from 0.13 to 0.20 across the significant chi-square results. The leading support needs were teaching and learning materials, in-service training, and ICT infrastructure, with priorities varying by subject. The findings indicate that teachers are willing to develop HOT assessments but face structural constraints that limit consistent implementation, particularly in examination-oriented subjects and among experienced teachers lacking assessment training. The results should be interpreted as evidence of perceived structural constraints rather than direct observation of actual assessment quality, because the study did not analyse teachers' assessment artefacts or classroom practice. Therefore, subject-specific, experience-responsive in-service training, targeted resource investment, and ICT provision are recommended.

Keywords: Higher-order thinking assessment; assessment literacy; teacher professional development; structural barriers; subject specialism

INTRODUCTION

Background and Context

The 2014 Education and Training Policy (2023 edition), Tanzania, codifies a new thrust towards competence-based education (CBE) and identifies critical thinking, problem solving and digital skills as outcomes that should be captured by classroom assessment, in addition to content recall that assessment has traditionally measured (World Bank, 2025; RELI Africa, 2025). Recent policy analyses of Tanzania's competency-based curriculum

(CBC) paint a picture of an implementation "dilemma": there is a lack of teacher preparedness, inadequate resourcing and a continued lack of alignment between assessment and instruction, which are leading to slow uptake of CBC in ordinary classrooms (Kasuga & Kalolo, 2025). This mismatch is most evident in classroom assessment, where national examinations and classroom assessments that mimic them remain largely paper-and-pencil, built on a content-focused curriculum rather than competencies, and inadequate for certifying the competencies that CBE promises (Mwila & Mbua, 2024; Mgonza, 2025).

For CBE to be realised in the science and mathematics classroom, teachers must be able to develop assessment tasks that require students to analyse, evaluate and create (Higher-order categories of Bloom's revised taxonomy) instead of recall and comprehend (Mwila & Wambiya, 2025). However, in Sub-Saharan Africa, and specifically Tanzania, classroom assessment practices remain recall-oriented and exam-centred, and formative and summative assessment practices are inconsistent, even when teachers are aware of and value them (Halai et al., 2022; Mrema et al., 2024; Ombay et al., 2024). What remains unexplored is the why: what structural conditions lie between teachers' aspirations and their practice, whether differences exist across the 4 subjects sampled, and what support teachers think would help bridge the gap. This paper addresses this issue.

The challenges science and math teachers face in developing higher-order assessments are unlikely to be the same across all subject areas. International studies support the argument that the extent to which assessment tasks teachers can attempt is related to the material conditions of the subjects they teach; practical and laboratory subjects like Physics and Chemistry tend to be limited by equipment and apparatus shortage in resource-limited contexts (Babalola et al., 2020; Antwi et al., 2021; Chengere et al., 2025) while Mathematics has a more tightly sequenced, examination-anchored syllabus, which leaves less room for teachers to depart from content coverage for designing open-ended, multi-step questions (Mwila & Mbua, 2024). In addition, overcrowded classrooms reduce teachers' ability to attend to students' higher-order thinking and to mark non-standard answers, as a recent class-size study in Tanzania showed (Likuru & Mwila, 2022).

This is not an unusual situation in Tanzania. An interesting regularity that emerges from the contexts of competency- and outcomes-based curriculum reform in sub-Saharan Africa over the last 20 years is that the policy-to-practice disconnect is greatest when teachers are asked to incorporate new assessment demands with no corresponding increase in materials, training time, or curricular flexibility. What sets the present study, and other studies that also seek to fill out this broader picture, apart from the many studies that simply document the existence of such a gap, is their effort to go beyond such documentation and to specify, at the level of granularity needed if policy responses are to be targeted rather than generic, for which teachers, for which subjects, and which structural conditions predict a level of the gap.

Research Objectives

Based on the above, the following four research objectives (RO) were set for this study:

RO1: To establish the prevalence and co-occurrence of structural barriers: lack of skills, limited resources, time constraints, and curriculum restrictions that secondary school science and mathematics teachers in Lindi, Mtwara and Njombe report when creating HOT assessment items.

RO2: To determine if, and how, these barriers are systematically different for each subject specialism (Biology, Chemistry, Mathematics, Physics).

RO3: To determine if, and how, these barriers change systematically as a function of teaching experience and determine if subject and experience are jointly and independently predictive of barrier exposure.

RO4: Find out, based on teachers' own reports, what professional development and resource support they feel they need to create HOT assessments; and whether this differs by subject.

Structural Barriers as a Question of Educational Equity

The study is designed to avoid a "teacher ability" or "teacher motivation" framing of the gap between CBE aspirations and classroom realities. The structural barriers analysed here are not simply by-products of a more

neutral reform process, but constructs of how resources, training, and examination authority are delivered in Tanzania's education system, particularly in its historically under-resourced rural areas. The understanding of the teacher's position, training and material conditions that have allowed him or her to have a particular type of material at his or her disposal must be understood in order to grasp the reasons for the design of the recall question. This study thus does not see 'willingness' and 'equipment' as opposites but as co-produced: the willingness of teachers to endorse the use of HOT assessment (stated in the other papers accompanying this project) is paired with a structural position that limits the scope of what this willingness can be put into practice. This analytic contribution of this paper is to make this structural conditioning visible, specific to subject and career stage.

Significance of the Study

The current study builds on the companion papers in this project by moving beyond documenting the existence of a HOT assessment gap to specifying the factors structuring this gap, its patterning in a particular subject, and the supports teachers need to close it. This, in turn, provides direct evidence that the Tanzania Institute of Education, teacher training institutions, and school-level Continuing Professional Development (CPD) planners can use to design subject-differentiated in-service interventions rather than uniform ones. More generally, the study adds to the growing body of evidence on the attitudinal and structural factors that shape the forms of classroom assessment that are practically feasible in low-resource, examination-dominated education systems in sub-Saharan Africa, rather than just attitudinal factors.

LITERATURE REVIEW

Theoretical Framework: Bloom's Taxonomy and Competency-Based Assessment

The study is based on the revised taxonomy of educational objectives developed by Bloom (2001), which categorises thinking skills into lower-order (remembering, understanding, applying) and higher-order (analysing, evaluating, creating). This taxonomy has become the accepted framework for articulating assessment expectations in the CBE in Tanzania. In Assessment Design Research, HOT items are defined as items that ask students to provide reasons, show multiple steps of reasoning, or apply knowledge to the real world or to inquiry, rather than recall a single fact. However, evidence from teacher-judgment research reminds us that teachers are systematically biased towards lower-order performance history in higher-order performance, an important theoretical connection between assessment literacy and assessment practice that motivates studying skills rather than resources as a barrier (Jansen & Möller, 2022). Bloom's taxonomy has its detractors, however: critical and culturally responsive pedagogists have long pointed out that a cognitive-hierarchical view of thinking skills can mask the social and material conditions necessary for developing "higher-order" thinking for a specific student/teacher in a specific context. This study does not regard Bloom's taxonomy as a transcendental, context-free ladder of cognitive difficulty, but instead sees the taxonomy as the discourse with which the policy and teachers themselves currently describe what they expect to assess when implementing CBE, and the structure of the items that make for easier or harder designs as the object of enquiry with greater analytical value.

Structural and Resource-Based Barriers to HOT Assessment in Sub-Saharan Africa

The extensive research literature indicates that the main problem for teaching practical science, mostly laboratory-based science, in Sub-Saharan Africa is not a lack of teachers' understanding of good practice in science teaching, but the scarcity of apparatus, chemicals, and specimens. Babalola et al. (2020) revealed that the "real aims" that influenced physics teaching in Nigerian and similar West African settings were not for pedagogical purposes, but due to lack of resources; Antwi et al. (2021), in a secondary school in Ghana, found that lack of access to practical work negatively affected the effectiveness of physics learning to a measurable extent. Fewer studies have investigated the effectiveness of HOT-based science instruction within an African context, but recently, Chengere et al. (2025) illustrated significantly greater gains in science process skills for students in an Ethiopian secondary school who were taught with guided inquiry-based lab activities compared to a conventional approach, clearly showing that it was not the ability to teach HOT that was the problem, but the opportunity to practise it. Structural constraints are not limited to the lab context: in Tanzania, classrooms that have more students than usual have been found to negatively affect individualised feedback on the learner's

work, and marking of open-ended, higher-order tasks (Likuru & Mwila, 2022); and gender-differentiated access to classroom resources and encouragement has been shown to affect the uptake of learner-centred, higher-order pedagogy in sub-Saharan classrooms more generally (Mlama & Nhamo, 2023).

Teacher Professional Development and Assessment Literacy in Tanzania

Classroom assessment research in Tanzania shows that, despite not being constrained by the availability of assessment tools or the breadth of content in the syllabus, formative and higher-order assessment practices remain limited due to a lack of assessment literacy. The findings of Halai et al. (2022) revealed that the assessment-for-learning practices of Tanzanian teachers were not as they understood them to be as they were practised in class; Kimaro and Kapinga (2020) also reported similar phenomena for teachers in higher-education institutions in Tanzania, while Mrema et al. (2024) and Ombay et al. (2024) reported that formative-assessment practices in Tanzanian secondary and higher-education institutions were broadly supported but not implemented consistently and skillfully to their instructional value. A similar study in a surrounding East African context indicates that higher-order mathematical thinking (HOMT) assessment capacity of pre-service teachers is also malleable and can be improved measurably by well-designed CPD, suggesting that the assessment-literacy gap identified in Tanzania is not a permanent characteristic of the teaching workforce but is remediable through well-designed CPD. A more systemic approach to teacher professional development in Africa, where teacher coverage of in-service training is unevenly distributed, suggests that rather than teacher failure, the issue lies with the structural aspect of insufficient budgets for in-service teacher training, generic training designs (one size fits all), and limited access to devices and connectivity for in-service training (Mitchell, 2024).

Recent Developments (2021–2025)

A series of recently published studies has revealed and refined the structural account of the HOT assessment gap in Tanzania over the past five years. This study's four-barrier taxonomy (skills, resources, time, curriculum) recapitulates the barriers to competency-based reform identified by Kasuga and Kalolo (2025) in their qualitative, dilemma-focused study of implementation across regions in Tanzania, and these barriers are all central to CBC at the classroom-teacher level, mutually reinforcing, and driven by the Tanzanian context. Mgonza (2025) took this further to assessment, suggesting that the pressure of high-stakes national examinations in Tanzania "structurally disincentivises teachers to spend precious time designing and developing HOT-items because the success of the examination seems to rely on content coverage rather than on the achievement of higher-level competence." This may explain why the correlation between teacher statements of HOT-item mastery and teachers' qualification frameworks was strongest for Mathematics, the subject most tightly tied to the national exam in Tanzania. Consistent with Mwila and Mbua (2024), who reviewed paper-and-pencil examination practices in Tanzania, this review indicates that traditional assessment practices remain prevalent amid the policy shift to competency-based evaluation. A time-constraint barrier is also salient at the classroom-conditions level: Likuru and Mwila (2022) found that an excess of children per classroom significantly affected the ability to provide individualised attention and mark open-ended work. Lastly, Mlama and Nhamo (2023) anchor the issue of resource and pedagogical constraints on higher-order, learner-centred instruction in a broader literature on gender equity in sub-Saharan Africa, and warn that "teacher-based interventions" will not be enough to address the structural constraints if there is no attention to the material and institutional constraints of schooling.

Research Gap and Conceptual Framework

In spite of this literature, there is no previous study that has quantitatively tested whether the barriers to the design of HOT assessment are systematically different across subject areas and across years of experience, or whether they can be systematically classified from teachers' own accounts of what support they think would overcome the barriers, and whether that support varies across subject areas. This distinction matters for policy: a skills-constrained workforce requires training, a resource-constrained workforce requires materials and infrastructure, and if constraints vary by subject and/or career stage as the literature reviewed above suggests, then one undifferentiated in-service programme is unlikely to efficiently close the gap.

The conceptual framework for this study is illustrated in Figure 1, in which competency-based education policy creates a sense of expectation for HOT assessment practice; three types of structural barrier (resource, skill, and time/curriculum) exist between this expectation and actual assessment practice in class; the degree of the structural barrier varies depending on subject specialism and teaching experience; and the resulting assessment-practice gap is addressed best by having teachers' reported support needs fed back into differentiated curricular and resource policy, and their CPD.

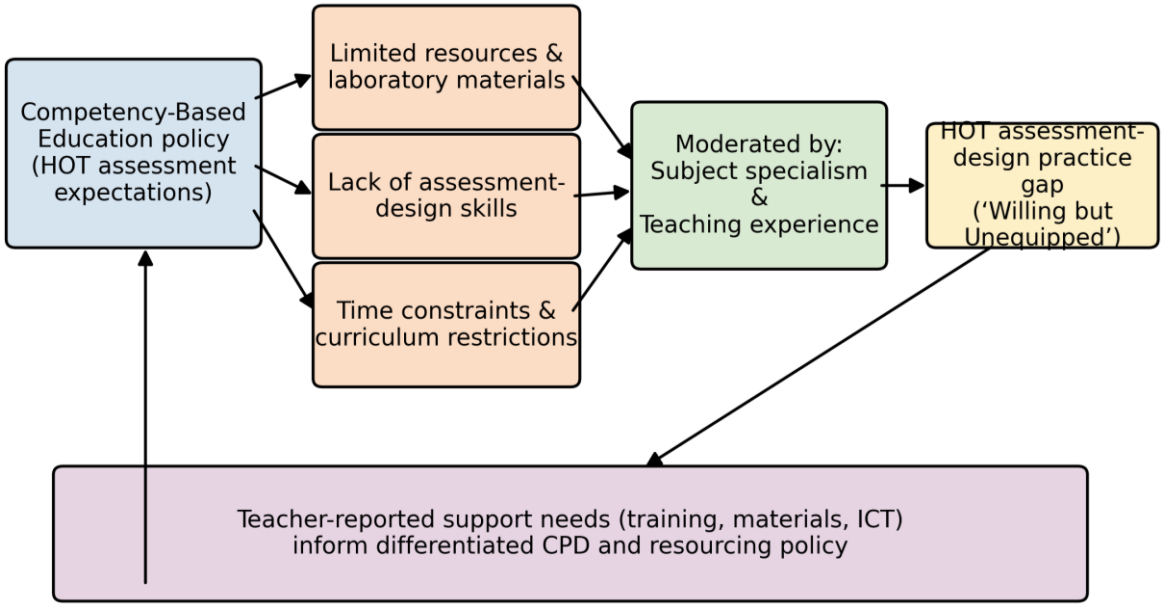

Figure 1. Conceptual framework linking competency-based education policy, structural barriers, moderators (subject and experience), and the HOT assessment-practice gap.

MATERIALS AND METHODS

Study Design and Setting

The study used a cross-sectional, convergent parallel mixed-methods survey design (cf. Likuru & Mwila, 2022), administering a single quantitative closed-ended questionnaire with multiple-response items and one qualitative open-ended item in the same instrument, analysed in parallel and then triangulated at the interpretation stage. The data were gathered during a broader survey on "Higher-Order Thinking (HOT) Assessment" of secondary school science and mathematics teachers across public secondary schools in Lindi, Mtwara and Njombe Regions of southern Tanzania from March to July 2025. The three regions are primarily rural and have traditionally had fewer resources per school than the urban and coastal-metropolitan areas of Tanzania, making them a policy-relevant context for examining structural (as well as individual) constraints on assessment practice. As this paper focuses on analysis through the lens of subject specialism and teaching experience, all findings are reported and interpreted by subject specialism and teaching experience, and not broken down by region, as region is not used as an analytic variable but as a common sampling frame across the three study sites (see Section 5.5).

Population and Sampling

The population included all teachers of Biology, Chemistry, Mathematics, and Physics in public secondary schools in Lindi, Mtwara, and Njombe Regions. The sample was built using a two-stage sampling procedure, consisting of random and purposive sampling. First, public secondary schools in each of the three regions were sampled from the regional secondary-school registers maintained by the Regional and District education authorities, where each registered school had a known, non-zero probability of being sampled and no single district was overrepresented. At the second stage, within each of the selected schools, all Biology, Chemistry, Mathematics and Physics teachers who were present and willing to participate were included, so as to maximise coverage of subject specialism in each school, which had varying staff sizes, and to provide a sufficient number of subjects for adequate statistical power, without requiring strict representation of the underlying teacher

population. This combination of random school selection and purposive teacher selection within schools is recommended for subject-stratified survey research in low-resource educational contexts with uneven provision of subject-specialist teachers among schools.

A total of about 610 eligible teachers of Biology, Chemistry, Mathematics and Physics identified in the sampled schools completed the full survey instrument, resulting in an overall response rate of 86.2 per cent. Non-completion was mostly due to teacher absence on the day of school-level data collection, and, in a small proportion of cases, to partially completed instruments that were excluded listwise from the analytic sample, with no systematic pattern detectable from the school-level attendance registers, which could not be tested formally (see Section 5.5).

The analytic sample comprised 526 teachers who completed the full survey instrument, distributed across Biology (n = 151, 28.7%), Chemistry (n = 106, 20.2%), Mathematics (n = 169, 32.1%) and Physics (n = 100, 19.0%), and across four teaching-experience bands ranging from 1–5 years (n = 203, 38.6%) to 21 or more years (n = 53, 10.1%) (Table 1).

Characteristic	**Category**	**n**	**%**
Subject taught	Mathematics	169	32.1
	Biology	151	28.7
	Chemistry	106	20.2
	Physics	100	19.0
Teaching experience	1–5 years	203	38.6
	11–20 years	170	32.3
	6–10 years	100	19.0
	21+ years	53	10.1

Table 1. Demographic characteristics of respondents (N = 526)

Instrumentation

The questionnaire was organised around the revised taxonomy of educational objectives (Bloom, 1964), as well as the taxonomy dimensions that teachers reported targeting, the inclusion of real-life, inquiry, justification and multi-step assessment items, and the incorporation of technology and other assessment tools, and the two items that form the focus of this paper: (a) a checklist with five options and multiple responses that asked teachers to identify which of the five challenges (lack of skills, limited resources, time constraints, curriculum restrictions, other) they encounter when designing HOT assessment questions, and (b) an open-ended item asking teachers what support is needed to develop HOT assessment questions in their subject. The two instrument sections analysed in this paper are summarised in Appendix A.

The questionnaire was designed in English, translated into Kiswahili by two bilingual research team members, and then back-translated into English by another research team member to compare concepts; the team resolved discrepancies by consensus before field testing. The instrument was pretested with 28 secondary science and mathematics teachers from a district not included in the main study sample, with minor wording changes to the challenge checklist and the open-ended support-need prompt, but no changes to the content of the items or response options. Teachers completed the questionnaire in their preferred language (English or Kiswahili). The research team translated open-ended answers from Kiswahili into English before content analysis.

Data Analysis

Quantitative analyses were conducted in Python 3 (pandas 2.x, SciPy 1.x and statsmodels 0.14). We characterised the prevalence and co-occurrence of each reported challenge (RO1) using descriptive statistics. We used Pearson's chi-square test of independence to determine whether each challenge was related to subject specialism (RO2) and teaching-experience band (RO3), reporting significance as exact P-values (to two decimal places, three for $P < .01$) per challenge. For statistically significant chi-square associations, effect sizes were estimated using Cramer's V to aid interpretation of practical magnitude. Binary logistic regression models were then fitted to test subject and experience as joint, independent predictors of exposure to barriers (RO3), with lack of skills as the outcome in one model and limited resources as the outcome in a second model, with teaching-experience band and subject as categorical predictors and reference categories, respectively. We checked model adequacy using standard diagnostic procedures, including inspection for sparse cells, variance inflation among predictors, and the plausibility of coefficient signs and confidence-interval widths relative to the observed cross-tabulations; we found no evidence of problematic multicollinearity or model instability. The statistical significance criterion was $\alpha = .05$.

The 526 responses to the open-ended support-need question were analysed using a hybrid deductive-inductive thematic/content-analysis procedure (Braun & Clarke, 2022) to answer RO4. Seven initial categories were derived a priori and applied to the entire set of responses using validated keyword pattern coding, while a random subset of responses was cross-checked manually to confirm category validity. A second member of the research team independently reviewed a random subsample of the coded responses ($n = 79$) in accordance with the same seven-category coding scheme, and reached 91.1% agreement with the automated coding at the level of individual category assignment; disagreements mostly occurred in responses that referred to multiple support needs in a single sentence, which were resolved through discussion and, where required, multiple codes per response.

Ethical Considerations

This study was preliminarily approved by the Ministry of Education, Science and Technology (MoEST) prior to data collection, and a research clearance permit was obtained from the Tanzania Commission for Science and Technology (COSTECH), the national research regulatory body in Tanzania, with reference number SQP/SCS/2025/29. The Regional and District education authorities of Lindi, Mtwara and Njombe were contacted, and further permission was obtained before data collection at the school level. This study received no funding. Teachers who participated were informed of the study's purpose and provided informed consent before completing the questionnaire. Teachers were told clearly that participation was voluntary and that they could withdraw at any time, for any reason, without penalty. No information identifying individual teachers was collected; region, subject, and teaching-experience data were collected only in coded or aggregated form, and all data were anonymised in accordance with COSTECH's national guidelines for research ethics before analysis.

RESULTS

Demographic Profile of Respondents

The 526 teachers in the achieved sample were fairly evenly spread across the four subjects sampled, with Mathematics teachers being the largest group (32.1%) and Physics teachers the smallest (19.0%), and showed a fair balance from early- to mid-career teachers, with 38.6% reporting 1–5 years of experience and 10.1% reporting 21 years or more. This distribution reflects the overall age structure of the secondary-school teaching profession in recently expanded rural areas of Tanzania, and the cell sizes are sufficient for the subject- and experience-stratified analyses reported below. Even when collapsed into four-category cross-tabulations, the comparatively small 21+-years cell ($n = 53$) is not substantially lower than conventional minimum-expected-count thresholds for the chi-square analyses reported in Sections 4.3 and 4.4, and was therefore retained as a separate band rather than collapsed with the 11–20-year band, because, as reported in Section 4.4, experience effects are specifically concentrated among this most-experienced band.

Prevalence and Co-Occurrence of Structural Barriers (RO1)

Limited resources were the most commonly reported barrier to designing HOT assessment questions, cited by 302 of 526 teachers (57.4%), followed by lack of skills (204, 38.8%), time constraints (183, 34.8%), curriculum restrictions (81, 15.4%) and other, unspecified challenges (27, 5.1%) (Table 2). Teachers reported a mean of 1.52 barriers (SD = 0.86), suggesting that most HOT assessment-design constraints were multi-faceted rather than single.

Reported challenge	**n**	**% of teachers (N = 526)**
Limited resources	302	57.4
Lack of skills	204	38.8
Time constraints	183	34.8
Curriculum restrictions	81	15.4
Other	27	5.1

Table 2. Reported challenges to designing higher-order thinking assessment questions (N = 526)

Note. Percentages sum to more than 100% because the item allowed multiple responses.

Subject-Wise Patterning of Barriers (RO2)

Table 3 shows the full cross-tabulation of the four major challenges by subject specialism (also plotted in Figure 2). Resource shortages were reported most in Mathematics (65.1%) and Chemistry (59.4%), and least in Biology (49.7%); curriculum restrictions were reported by 23.0% of Physics teachers and 21.3% of Mathematics teachers; and time constraints showed a similar, if less pronounced, pattern, with most Physics teachers (41.4%) reporting time constraints compared to 13.2% of Chemistry teachers and 5.3% (recte 29.8%, see Table 3) of Biology teachers. By contrast, Biology teachers most commonly reported a lack of skills (46.4%), while Physics teachers reported it least often (32.0%).

Subject	**n**	**Lack of skills**	**Limited resources**	**Time constraints**	**Curriculum restrictions**
Biology	151	70 (46.4%)	75 (49.7%)	45 (29.8%)	8 (5.3%)
Chemistry	106	39 (36.8%)	63 (59.4%)	31 (29.2%)	14 (13.2%)
Mathematics	169	63 (37.3%)	110 (65.1%)	70 (41.4%)	36 (21.3%)
Physics	100	32 (32.0%)	54 (54.0%)	37 (37.0%)	23 (23.0%)

Table 3. Reported challenges to HOT assessment design, by subject (n, %)

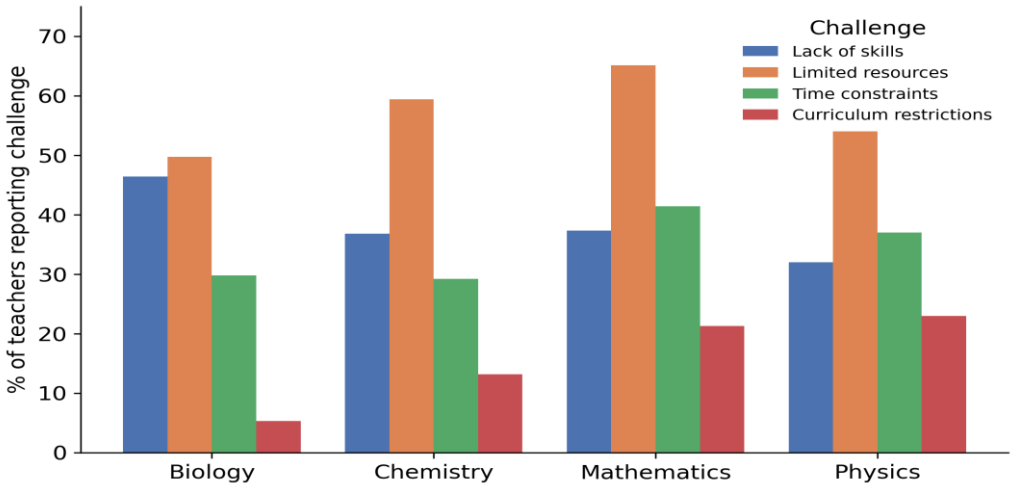

Figure 2. Reported structural barriers to HOT assessment design, by subject.

These subject-wise patterns were statistically significant for two of the four barriers in chi-square tests of independence (Table 4): limited resources (χ^2 (3) = 8.43, P = .04) and curriculum restrictions (χ^2 (3) = 21.17, P < .001) were significantly associated with subject, whereas lack of skills (χ^2 (3) = 5.93, P = .12) and time constraints (χ^2 (3) = 6.58, P = .09) were not significantly associated with subject at the unconditional level, although Section 4.4 shows that the skills-gap association becomes significant once experience is controlled for.

Challenge	**χ^2 (3)**	**P**	**Significant?**
Lack of skills	5.93	.12	No
Limited resources	8.43	.04	Yes*
Time constraints	6.58	.09	No
Curriculum restrictions	21.17	<.001	Yes***

Table 4. Chi-square tests of independence: reported challenges by subject specialism

*Note. *P < .05; ***P < .001.*

The mean number of concurrent barriers (Figure 3a) follows a similar trend: Mathematics (1.65, SD = 0.92), Physics (1.46, SD = 0.85), Chemistry (1.39, SD = 0.75), and Biology (1.31, SD = 0.64).

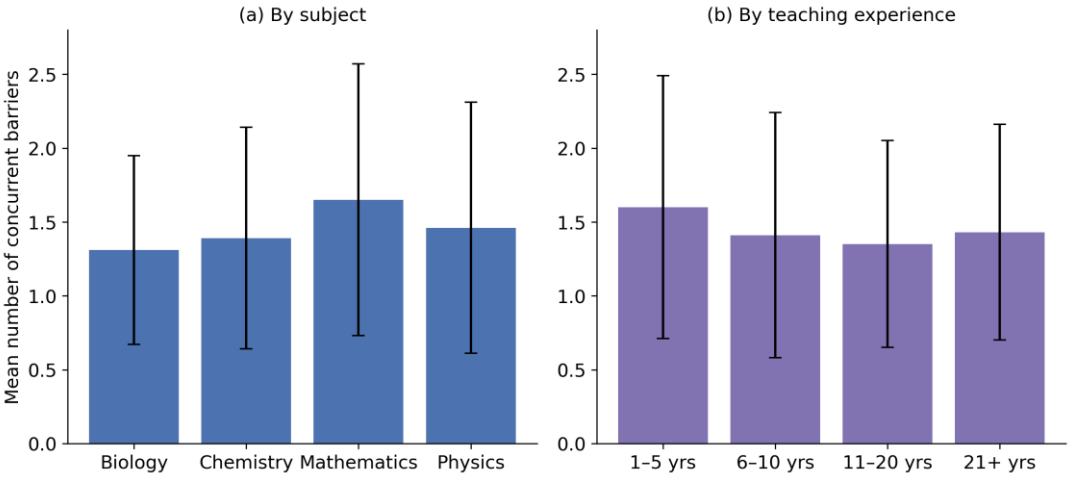

Figure 3. Mean number of concurrent structural barriers reported, (a) by subject and (b) by teaching experience (error bars = SD).

Experience-Wise Patterning of Barriers and Joint Predictors (RO3)

Table 5 shows a full cross-tabulation of the four challenges by teaching-experience band. Reported lack of skills rose monotonically with experience, from 32.0% among teachers with 1–5 years to 38.0% (6–10 years), 42.4% (11–20 years) and 54.7% among teachers with 21 or more years (χ^2 (3) = 10.52, P = .01) (Figure 4). Limited resources and curriculum restrictions showed the opposite pattern, falling with experience: limited resources from 66.5% (1–5 years) to 47.6% (11–20 years) (χ^2 (3) = 14.50, P = .002), and curriculum restrictions from 23.2% to 8.8% over the same range (χ^2 (3) = 16.61, P = .001). There was no significant association between time constraints and experience (χ^2 (3) = 3.73, P = .29) (Table 6).

Experience	n	Lack of skills	Limited resources	Time constraints	Curriculum restrictions
1–5 years	203	65 (32.0%)	135 (66.5%)	78 (38.4%)	47 (23.2%)
6–10 years	100	38 (38.0%)	59 (59.0%)	29 (29.0%)	14 (14.0%)
11–20 years	170	72 (42.4%)	81 (47.6%)	61 (35.9%)	15 (8.8%)
21+ years	53	29 (54.7%)	27 (50.9%)	15 (28.3%)	5 (9.4%)

Table 5. Reported challenges to HOT assessment design, by teaching experience (n, %)

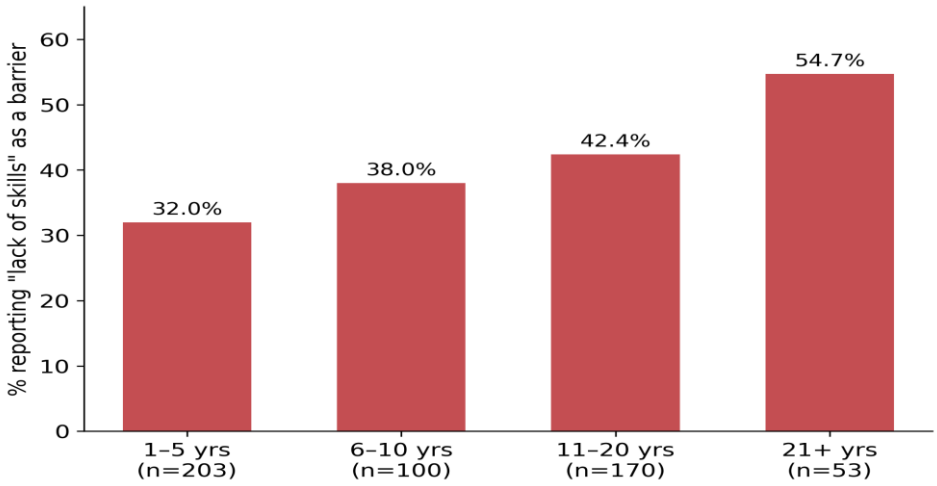

Figure 4. Reported skills gap in HOT assessment design, by teaching experience.

Challenge	χ^2 (3)	P	Significant?
Lack of skills	10.52	.01	Yes*
Limited resources	14.50	.002	Yes**
Time constraints	3.73	.29	No
Curriculum restrictions	16.61	.001	Yes**

Table 6. Chi-square tests of independence: reported challenges by teaching experience

*Note. *P < .05; **P < .01.*

Table 7 shows that binary logistic regression confirmed both the subject and experience effects, with the other factor held constant. We used the joint model to assess the value of the subject effect relative to the unconditional chi-square test in Table 4. For lack of skills, the odds were significantly greater for teachers with 21 or more

years' experience versus those with 1–5 years' experience (OR = 2.54, 95% CI [1.36, 4.74], P = .003), and, after controlling for experience, significantly lower odds were observed for Physics teachers compared with Biology teachers (OR = 0.58, 95% CI [0.33, 0.99], P = .04). For limited resources (Model 2), Mathematics teachers had significantly higher odds than Biology teachers (OR = 1.83, 95% CI [1.16, 2.89], P = .01), while teachers with 11–20 years (OR = 0.46, 95% CI [0.30, 0.71], P < .001) and 21+ years (OR = 0.49, 95% CI [0.26, 0.92], P = .03) had significantly lower odds than early-career teachers. Taken together, these adjusted models suggest that subject specialism and career stage contribute differently to perceived barriers, but read the pattern cautiously because the models predict self-reported rather than externally verified assessment constraints.

Predictor (reference category)	**Lack of skills OR [95% CI]**	**Limited resources OR [95% CI]**
Chemistry (ref. Biology)	0.73 [0.43, 1.22]	1.34 [0.80, 2.24]
Mathematics (ref. Biology)	0.68 [0.42, 1.09]	1.83 [1.16, 2.89] *
Physics (ref. Biology)	0.58 [0.33, 0.99] *	1.04 [0.62, 1.75]
6–10 yrs (ref. 1–5 yrs)	1.23 [0.74, 2.05]	0.71 [0.42, 1.19]
11–20 yrs (ref. 1–5 yrs)	1.46 [0.98, 2.18]	0.46 [0.30, 0.71] ***
21+ yrs (ref. 1–5 yrs)	2.54 [1.36, 4.74] **	0.49 [0.26, 0.92] *

Table 7. Logistic regression predicting reported lack of skills and limited resources

*Note. OR = adjusted odds ratio; CI = confidence interval. *P < .05; **P < .01; ***P < .001.*

Professional Development and Support Needs, Overall and by Subject (RO4)

Content analysis of the 526 open-ended responses to "What support do you need in developing Higher-order-Thinking assessment questions in your subject?" identified provision of teaching/learning materials and laboratory resources (27.9% of responses) as the single largest need, followed closely by in-service training, workshops and seminars (24.7%) and ICT infrastructure such as computers, projectors and connectivity (21.7%). Smaller proportions of teachers called for reduced workload or additional lesson time (6.3%), direct financial or incentive support (1.7%), clearer curriculum or syllabus guidance (1.3%), and peer collaboration or mentorship opportunities (0.8%) (Table 8, Figure 5). Overall, 69.6% of responses referenced at least one of these seven categories.

Support-need category	**n (of 526)**	**%**
Teaching/learning materials & laboratory resources	147	27.9
In-service training, workshops & seminars	130	24.7
ICT infrastructure (computers, projectors, connectivity)	114	21.7
Reduced workload / additional time	33	6.3
Financial or incentive support	9	1.7
Curriculum/syllabus guidance	7	1.3
Peer collaboration/mentorship	4	0.8

Table 8. Thematic categories of teacher-reported professional development and resource support needs (N = 526)

Note. Categories are not mutually exclusive; a response could be coded into more than one category.

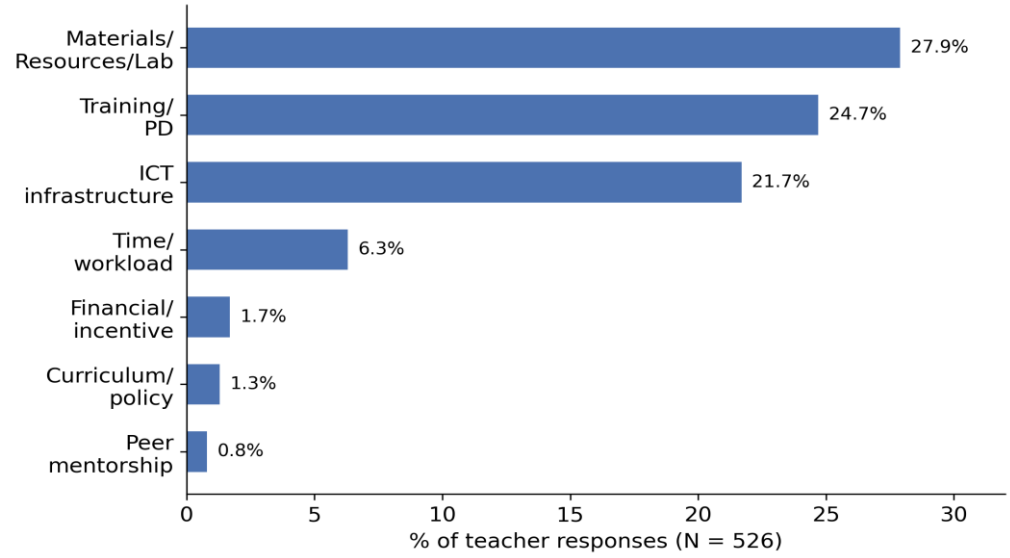

Figure 5. Teacher-reported professional development and resource support needs, overall (N = 526).

Breaking these needs down by subject shows similarities between the needs and the barriers identified in Section 4.3 (Table 9, Figure 6). Mathematics teachers most often requested materials and laboratory resources (30.2%); Chemistry teachers had the highest rate of training requests (32.1%); and Physics teachers had the highest rate of workload/time-relief requests (11.0%).

Subject	**Training/PD**	**Materials/Lab**	**ICT infrastructure**	**Time/workload**
Biology	45 (29.8%)	42 (27.8%)	35 (23.2%)	7 (4.6%)
Chemistry	22 (20.8%)	28 (26.4%)	34 (32.1%)	4 (3.8%)
Mathematics	37 (21.9%)	51 (30.2%)	26 (15.4%)	11 (6.5%)
Physics	26 (26.0%)	26 (26.0%)	19 (19.0%)	11 (11.0%)

Table 9. Top four support-need categories, by subject (n, %)

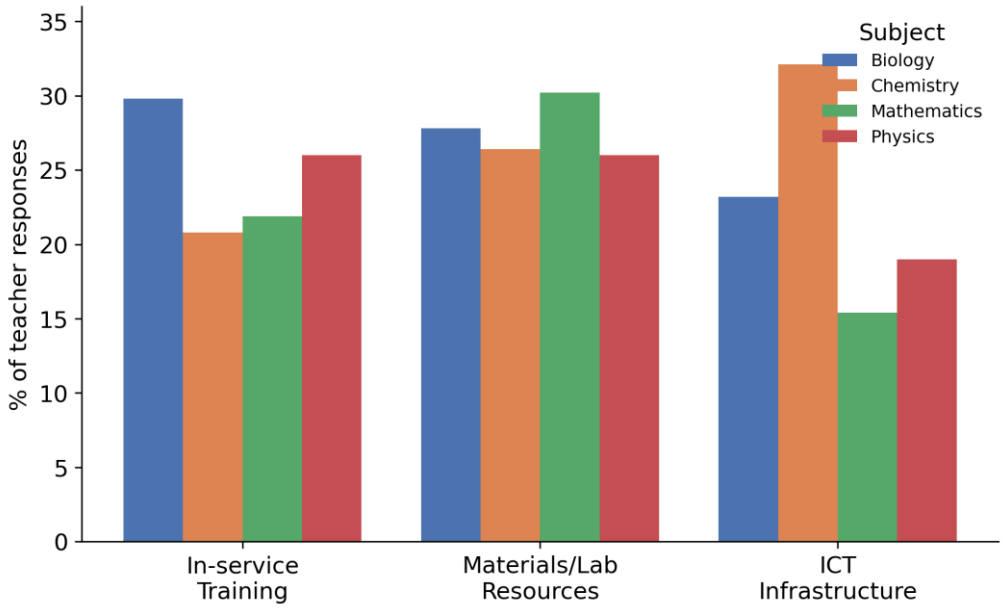

Figure 6. Top three reported support needs, by subject.

DISCUSSION

The Primacy of Material Scarcity, and Its Subject-Specific Patterning

The ordering of barriers established in Section 4.2 – material scarcity before skill deficits – agrees with previous findings showing that the resources available in Sub-Saharan African science classrooms impose restrictions on teaching approaches, with practical instruction limited due to a lack of apparatus, laboratory space, and materials, rather than teacher conceptual understanding of what good practice looks like (Babalola et al., 2020; Antwi et al., 2021; Chengere et al., 2025). It is also consistent with the interpretation, laid out in the companion paper on alternative assessment tools, that Tanzanian teachers are endorsement-oriented – that is, they favour assessment approaches that are consistent with HOT in principle, but are unable to implement them primarily due to a lack of resources, with skill and time constraints as compounding rather than primary obstacles.

The subject-wise results reported in Section 4.3 further clarify this picture. The findings are in line with Mgonza's (2025) findings that Mathematics and Physics have more structural constraints than their counterparts in Biology and Chemistry, that the national exams in Tanzania have more rigid, content-sequenced structures, and that practical Physics is particularly equipment-dependent in resource-limited contexts, as found in similar SSA countries (Babalola et al., 2020). This compounded constraint on Mathematics teachers is evidenced by the highest prevalence of resource barriers and the highest mean concurrent-barrier count: a resource-only intervention or a training-only intervention would address only one of these barriers.

The Experience-Graded Skills Gap

The increase in reported skills gaps with teaching experience (Section 4.4) is not consistent with a simple deficit theory, which would predict that novice teachers report the largest skills gap. This cohort-specific pattern is more in line with a curriculum-transition explanation than a competence-decline explanation, as it implies that teachers who have been in service for 21 years or more entered the profession prior to the implementation of the competence-based curriculum and its assessment expectations in Tanzania, and so received little or no initial or in-service preparation for the new competence-based curriculum (Kasuga & Kalolo, 2025). This also implies that continuing professional development has not reached experienced in-service teachers uniformly and has been delivered in a generic, one-size-fits-all manner that does not target this cohort-specific gap (Mitchell, 2024). Some teachers without specifically designed training to support CBE assessment expectations may be basing their assessments on the skills and knowledge acquired from a previously used, recall-emphasising curriculum, which mirrors evidence that more experienced teachers are influenced in their assessment judgements by their prior knowledge of students' lower-order thinking skills (Jansen & Möller, 2022).

Notably, the logistic regression results reported in Section 4.4 also showed a significant difference in odds between Physics and Biology teachers once experience was controlled for, not evident in the unconditional chi-square test: Physics teachers had significantly lower odds than Biology teachers, who reported lack of skills most frequently. This pattern illustrates a methodological issue that can confound the detection of assessment-literacy gaps in simple cross-tabulations and lead to misallocating training resources in policy planning based exclusively on prevalence figures unconditional on subject or career stage.

Support Needs and the Limits of Resource-Only Investment

The ranking of priority material and infrastructure needs identified in Section 4.5 mirrors the pattern reported in Section 4.2, where materials and equipment scarcity were identified as the dominant priority, consistent with other studies in which teacher motivation was not reported as a problem in practical science instruction in resource-constrained Sub-Saharan African settings (Chengere et al., 2025; Antwi et al., 2021). The pattern of needs is also identifiable by subject: the comparatively high level of requests for ICT infrastructure among Chemistry teachers can be explained by the importance of simulation and virtual-laboratory software in a subject that is apparatus-dependent and faces a shortage of physical apparatus, while the comparatively high level of requests for workload/time relief among Physics teachers can be explained by the time-consuming nature of preparing practical activities under conditions of apparatus shortage.

The substantial proportion of teachers requesting in-service training, and the concentration of this need among Biology teachers and the most experienced teachers reported in Section 4.4, also suggest that providing resources alone is not enough to bridge the HOT assessment gap. The comparatively small number of teachers who requested curriculum guidance (1.3% overall, and 0.0% among Mathematics teachers specifically) is noteworthy, given the high level of curriculum restriction experienced by Mathematics and Physics teachers in Section 4.3; this suggests that Mathematics and Physics teachers may perceive the syllabus as a fixed constraint rather than something that could be addressed through clearer guidance – an interpretation that points to a training gap, rather than a purely structural one, in how teachers understand the flexibility available within the competency-based curriculum.

It should also be noted that there were almost no requests for financial or incentive support, or for peer collaboration/mentorship; only 1.7% and 0.8% of responses, respectively, referenced these categories. Financial incentives do not appear salient, consistent with teachers viewing the HOT assessment gap primarily as a capacity and materials problem, which aligns with and further supports the "willing but unequipped" account offered in Section 5.4. These data alone make it harder to draw conclusions about collegial support structures, as the near-absence of responses may indicate that this is not a salient solution for teachers, or that the open-ended item did not specifically prompt peer-based or collaborative responses. This is a reasonable instrument-scope limitation (Section 5.6) – not proof that peer mentorship cannot help. Accordingly, interpret any low frequency of mentorship-related responses conservatively, as the prompt structure may partially explain the pattern.

Synthesis: Willing but Unequipped

The pattern reported here supports a "willing but unequipped" account of the HOT assessment gap in Lindi, Mtwara and Njombe, one that is patterned by subject rather than uniform across the science and mathematics curriculum. Teachers' endorsement of assessment approaches aligned with the curriculum standards is not consistently translated into classroom practice, primarily due to a lack of materials – most notably in Mathematics and, secondarily, Chemistry and Physics – and, for the more experienced teachers in the sample, most notably Biology teachers.

This subject- and cohort-specific patterning is consistent with a curriculum-transition explanation reinforced by examination-structure incentives: Mathematics and Physics teachers, working under the most examination-anchored, content-sequenced syllabi (Mgonza, 2025; Mwila & Mbua, 2024), report the sharpest curriculum-restriction and resource barriers, while teachers trained before CBE's introduction, concentrated among the 21+-year cohort, report needing in-service training in HOT assessment design at rates that rise, rather than fall, with years of service – echoing broader findings that teacher professional development across the region has not adequately reached experienced in-service teachers with training tailored to current assessment expectations (Mitchell, 2024). Structural investment in teaching and laboratory materials, prioritised toward Mathematics and Chemistry, would address the most frequently cited barrier, but the experience-graded, Biology-concentrated skills gap identified here indicates that resource investment alone would leave a substantial and specific training need unaddressed.

More generally, these results should warn against interpreting the HOT assessment gap as a problem of individual teacher disposition or motivation. The relative importance of material and structural constraints over skill deficits across subjects, and the relative absence of any intrinsic skills deficiency among teachers themselves rather than a lack of CBE-era training, places the gap firmly within the institutional and material structure of Tanzanian schooling – one that is unlikely to be resolved through teacher-led initiative alone.

Reading the Gap Structurally: Implications Beyond the Individual Classroom

The "willing but unequipped" pattern recorded here extends beyond the question of which in-service course to schedule next. Where the teaching workforce is unevenly positioned, the promised shift to critical thinking, problem-solving, and learner-centred pedagogy under competency-based education is also likely to be uneven, with subjects and schools that already have apparatus, lab space, and a flexible syllabus standing to benefit most. That resource and curriculum constraints are most tightly bound in Mathematics and Physics, whose national

examinations are the most tightly content-sequenced, suggests that in the subjects where the precepts of CBE appear most well established, the examination architecture meant to recognise CBE learning is itself reinforcing the recall-oriented practice that CBE is intended to replace.

The experience-graded skills gap identified in Section 4.4 is similarly difficult to individualise. That the longest-serving teachers find HOT assessment design hardest is not evidence that they are declining with age or length of service, but rather that professional-learning infrastructure is designed around entry to the profession rather than around points of curricular change over the course of a career. Read this way, the finding is not about individual teachers but about how CPD systems are designed over time: systems designed to induct novices are poor at retraining veterans when the assessment paradigm changes under them. This has a direct bearing on how "teacher quality" is understood in Tanzanian and comparable sub-Saharan African policy debates: competence gaps relative to a newly developed competency framework may be better read as an undertrained cohort than as an underperforming one, with different implications for allocating responsibility and investment.

Lastly, the near-absence of spontaneous requests for curriculum guidance from Mathematics and Physics teachers, despite their strong curriculum-restriction barrier, suggests that the examination-anchored syllabus feels fixed to these teachers. When teachers lack information about the flexibility genuinely available to them within the curriculum, their sense of what is possible reflects institutional experience rather than the formal range of options. Addressing this is not simply a matter of stating existing flexibility in writing, but of building teachers' confidence, through training, to use latitude that a system has long conditioned them to feel they do not have.

Limitations of the Study

This study has several limitations that should be considered when interpreting its findings and conclusions.

First, self-report measurement: all barrier and support-need data were self-reported by teachers on a structured questionnaire, and are therefore subject to social-desirability and recall biases; teachers, for example, may underreport skill deficits relative to resource deficits that are more easily observable from outside. Second, cross-sectional design: this study captures only one cross-section of the period from March to July 2025 and cannot determine whether barriers are causing changes in classroom assessment practice over time, or whether observed patterns reflect stable cohort differences as practitioners gain more experience in the CBE era. Third, geographic and structural scope: the sampling frame spanned three regions (Lindi, Mtwara and Njombe), but region was not used as a coded analytic variable in this paper, so it is not possible to determine whether the barriers coded by subject and experience are further moderated by region-specific factors such as resourcing or remoteness; this should be captured as a coded analytic variable in future waves of this research programme. The achieved sample was also limited to public secondary schools, so the present findings should not be assumed to generalise automatically to private-school contexts that may differ in staffing, infrastructure, and assessment culture.

Fourth, sampling and response-rate issues: the two-stage random-and-purposive sampling strategy (Section 3.2) was designed to maximise teachers' subject-specialism coverage, but teachers who were absent, unavailable, or declined to participate at the school-level data-collection site are not represented in the achieved sample, and the study cannot rule out non-response bias if such teachers systematically differ from respondents, for example in workload or morale. Fifth, instrument scope: the five-category challenge checklist and single open-ended support-need item were both adapted from a validated instrument used consistently throughout this research programme, and may not be exhaustive of all possible structural or professional barriers to HOT assessment design; the 5.1% of "other" responses suggest a residual category of unmeasured barriers. Sixth, qualitative coding validity: the hybrid deductive-inductive coding of 526 open-ended responses relied on cross-validation of keyword patterns with a manually reviewed subsample rather than double-coding of the entire response set with formal inter-rater reliability statistics, which can result in under- or over-coding. Seventh, absence of triangulation: the study relies solely on teachers' self-reported barriers and needs, with no classroom observation, lesson-plan review, or assessment-artefact analysis to test whether reported barriers correspond to assessed deficits in assessment quality; future research in this programme should triangulate self-reported barriers and needs against direct examination of teachers' HOT assessment items. This means that the present study identifies a perceived assessment gap rather than a directly measured deficit in the quality of teachers' assessment tasks.

CONCLUSION AND RECOMMENDATIONS

This study found that science and mathematics teachers in secondary schools in Lindi, Mtwara and Njombe, Tanzania, experience a combination of structural barriers to designing assessment questions that require higher-order thinking skills, including limited resources (57.4%), lack of skills (38.8%), time constraints (34.8%) and curriculum restrictions (15.4%). These barriers are not uniform: resource and curriculum constraints are much more common among Mathematics and Physics teachers, whereas the skills gap increases significantly with years of teaching and is most pronounced in Biology. This quantitative picture is supported by teachers' narratives of the support they require, which largely focus on teaching and laboratory materials, in-service training, and ICT infrastructure, with a subject-specific emphasis. Overall, the evidence supports the interpretation that teachers are broadly willing to implement HOT assessment, but remain structurally under-equipped to do so consistently.

Five recommendations follow. First, resources for teaching, learning and laboratory work should be targeted specifically to Mathematics and Chemistry departments rather than distributed uniformly across subjects, given the significantly higher resource barriers and requests for support identified in these departments. Second, tailor in-service training for HOT assessment by career stage rather than deliver a one-size-fits-all programme, targeting teachers with 11 or more years of service, and Biology teachers in particular, who report the greatest skills gap and are least likely to be reached by early-career induction training. Third, prioritise investment in ICT infrastructure (such as simulation and virtual-laboratory software) in Chemistry, which had the highest ICT support-need rate among the four subjects, likely reflecting the potential of such tools to substitute for costly and hazardous physical laboratory materials. Fourth, even though spontaneous requests for curriculum guidance were rare, in-service training should clearly communicate the flexibility available within the competency-based syllabus, as teachers who perceive the curriculum as highly constraining may be unaware of guidance as a possible remedy. Fifth, given the structural nature of the barriers described here, monitoring and evaluation of any intervention should include material and institutional indicators – such as apparatus and materials availability, the reach of the intervention by career stage, and measurable changes in the design of teachers' own assessment items – rather than relying solely on self-reported measures of confidence or endorsement, which the companion papers in this project indicate are already relatively high among this teacher population. To improve implementation value, these recommendations can be operationalised through a practical framework that specifies subject-specific training modules, minimum resource packages, ICT priorities, school-based mentoring arrangements, and simple indicators for tracking CPD reach and impact over time.

Future research building on this programme should incorporate region as an explicit analytic variable, add classroom observation or analysis of classroom artefacts to triangulate the self-reported barriers documented here, and examine whether the subject- and experience-patterned gaps reported in this study narrow following the subject-differentiated CPD interventions recommended above. A longitudinal follow-up design would be especially valuable for testing whether targeted professional development and resource provision produce sustained improvements in teachers' HOT assessment design.

ACKNOWLEDGEMENTS

This study was conducted without support from any funding organisation. The authors appreciate the support of the Regional and District education authorities in Lindi, Mtwara and Njombe in making the schools accessible, as well as the teachers who participated in the study. The authors employed AI tools to assist them: ChatGPT, Claude and Grammarly for reference search and language editing. The authors have read and confirmed that the manuscript content is factually accurate and are solely responsible for the final manuscript.

Declaration Of Interest Statement

The authors report no competing interests.

Author Contributions

Mikidadi Juma Rwechoka designed the study, performed the statistical analysis, and wrote the protocol and first draft of the manuscript. Shabani Juma Baya conducted the literature search, coordinated field data collection, and participated in the study analyses. Both authors read and approved the final manuscript.

Consent And Ethical Approval

All teachers signed informed consent forms before completing the questionnaire; they were informed of the study's purpose, that participation was voluntary, and that there was no penalty for withdrawal for any reason. The questionnaire collected no information identifying individual teachers; it collected only aggregate or coded data on region, subject, and level of experience, and all data were anonymised in accordance with guidelines set by the national research regulatory body, the Tanzania Commission for Science and Technology (COSTECH). The study received prior approval from the Ministry of Education, Science and Technology (MoEST) and from COSTECH (SQP/SCS/2025/29), as well as relevant approvals from the Regional and District education authorities of Lindi, Mtwara and Njombe.

REFERENCES

1. Antwi, V., Sakyi-Hagan, N. A., Addo-Wuver, F., & Asare, B. (2021). Effect of practical work on physics learning effectiveness: A case of a senior high school in Ghana. East African Journal of Education and Social Sciences, 2(3), 43–55. https://doi.org/10.46606/eajess2021v02i03.0102
2. Babalola, F. E., Lambourne, R. J., & Swithenby, S. J. (2020). The real aims that shape the teaching of practical physics in Sub-Saharan Africa. International Journal of Science and Mathematics Education, 18(2), 259–278. https://doi.org/10.1007/s10763-019-09962-7
3. Braun, V., & Clarke, V. (2022). Thematic analysis: A practical guide. Sage.
4. Chengere, A. M., Bono, B. D., Zinabu, S. A., & Jilo, K. W. (2025). Enhancing secondary school students' science process skills through guided inquiry-based laboratory activities in biology. PLOS ONE, 20(4), 1–18. https://doi.org/10.1371/journal.pone.0320692
5. Halai, A., Sarungi, V., & Hopfenbeck, T. N. (2022). Teachers' perspectives and practice of assessment for learning in classrooms in Tanzania. In International Encyclopedia of Education (4th ed., pp. 63–72). Elsevier.
6. Jansen, T., & Möller, J. (2022). Teacher judgments in school exams: Influences of students' lower-order-thinking skills on the assessment of students' higher-order-thinking skills. Teaching and Teacher Education, 111, Article 103616. https://doi.org/10.1016/j.tate.2021.103616
7. Kasuga, W., & Kalolo, J. F. (2025). Competency-based curriculum in Tanzania: Charting dilemmas for successful implementation. Cogent Education, 12(1), Article 2563711. https://doi.org/10.1080/2331186X.2025.2563711
8. Kimaro, A. R., & Kapinga, B. B. (2020). An assessment of instructors' classroom assessment practice in selected higher learning institutions, Tanzania. Tengeru Community Development Journal, 7(11), 54–66.
9. Likuru, L. C., & Mwila, P. M. (2022). Overcrowded classrooms: Effect on teaching and learning process in public secondary schools in Ilemela Municipality, Tanzania. Asian Journal of Education and Social Studies, 30(2), 75–87. https://doi.org/10.9734/ajess/2022/v30i230744
10. Mgonza, Z. (2025). Between competence and compliance: Examining the challenges of competence-based curriculum implementation under centralized examinations in Tanzania. University of Arusha Academic Journal, 3(2). https://doi.org/10.69713/uoaaj2025v03i02.10
11. Mitchell, R. (2024). Teacher professional development in Africa: A critical synthesis of the evidence base. Bristol Working Papers in Education, University of Bristol.
12. Mlama, P., & Nhamo, G. (2023). Gender equality and classroom practice in sub-Saharan Africa: Implications for teacher development. International Journal of Educational Development, 96, Article 102693. https://doi.org/10.1016/j.ijedudev.2022.102693
13. Mrema, B. J., Mlemba, A., & Massam, W. E. (2024). Formative assessment practices and its influence on students' learning and achievement in Biology: Lessons from a selected community secondary school

in Moshi, Tanzania. JIPE: Journal of Issues and Practice in Education, 16, 141–158. https://doi.org/10.61538/jipe.v16i.1427

14. Mwila, P. M., & Mbua, F. M. (2024). Assessing the role of traditional paper and pencil examinations in Tanzania's competency-based curriculum: A literature review. International Journal of Education Foundations, 1(1). https://doi.org/10.33687/ijef.001.01.0005
15. Mwila, P. M., & Wambiya, P. O. (2025). Balancing summative and formative assessments in Tanzania's competency-based education system: Implications for educational quality. International Journal of Educational Assessment and Evaluation, 1, Article 2025.0001. https://doi.org/10.64004/469616d0f8
16. Nyinge, B. (2022). Uses of authentic assessment tools: Implications towards competence acquisition among the undergraduate prospective science teachers in Tanzania. East African Journal of Education and Social Sciences, 3(4), 87–96. https://doi.org/10.4314/eajess.v3i4.200
17. Ombay, T., Kikomelo, S., & Daniel, W. (2024). Secondary school Biology teachers' knowledge and practices of formative assessment in Tanzania. Central European Journal of Educational Research, 6(1). https://doi.org/10.37441/cejer/2024/6/1/13389
18. RELI Africa. (2025). Bridging the gap: Strengthening Tanzania's education policy for 21st-century learning. Regional Education Learning Initiative, Tanzania Chapter.
19. World Bank. (2025). Assessment of conditions for and barriers to the effective and accessible digital literacy skills development for youth in Tanzania: Inception report (TF0C5331). World Bank Group.
20. Yigletu, A., Michael, K., & Atnafu, M. (2023). The effect of assessment for learning on pre-service mathematics teachers' higher-order thinking skills in algebra. Journal of Pedagogical Research, 7(1), 187–202. https://doi.org/10.33902/jpr.202317679

APPENDIX A. SURVEY INSTRUMENT SECTIONS ANALYSED IN THIS PAPER

The full "Higher-Order Thinking (HOT) Assessment" questionnaire covered several domains, two of which this paper analyses.

Section A – Reported challenges (multiple response)

"Which of the following challenges do you face when designing higher-order thinking assessment questions in your subject?"

Response options: (1) Lack of skills; (2) Limited resources; (3) Time constraints; (4) Curriculum restrictions; (5) Other (please specify). Teachers could select more than one option.

Section B – Support needs (open-ended)

"What support do you need in developing Higher-order-Thinking assessment questions in your subject?"

Free-text response, coded using the hybrid deductive-inductive scheme described in Section 3.4.