

# Evaluation of CBC Assessment Framework and Its Effectiveness on Reliability and Validity of CBC Learning Outcomes in Kenya

Dr Teresia Kiragu

Mount Kenya University

DOI: <https://dx.doi.org/10.47772/IJRISS.2025.91100083>

Received: 13 November 2025; Accepted: 18 November 2025; Published: 29 November 2025

## ABSTRACT

Kenya has been using Competency Based Curriculum (CBC) since 2017 to facilitate holistic learning based on skills, values, and attitudes instead of rote memorization. The success of this system is based on the reliability and validity of the assessment results. This paper examined the efficacy of Kenya CBC assessment framework by analyzing documents of 18 policy, institutional and academic sources. The study used qualitative research design involving document analysis. The target population of this study was official policy documents, curriculum guidelines, assessment guidelines, evaluation reports, and peer scholarly articles for the CBC in Kenya. The study employed purposive sampling to select documents that were most suitable and verifiable to achieve the research objectives. Analysis of the obtained data was done using qualitative content analysis based on thematic coding. The findings indicated that CBC assessments were associated with moderate reliability (41%) and validity (57%), variability in applying the rubrics, little moderation (28%) and weak construct validation. There was a high content-alignment (83%), and low empirical construct-testing (22%). The research findings revealed that the CBC framework is conceptually sound but lacks psychometric strength and fairness in different situations. The study recommends that MOE should enhance the calibration of the assessor, standardizing the rubrics, and the psychometric audit, and integration of the school context differences. There is need to improve technical and contextual validity of CBC assessments in order to measure the competencies of learners credibly.

**Keywords:** Assessment framework, CBC, Reliability, Validity, Learning outcomes

## BACKGROUND TO THE STUDY

The education system in Kenya has been undergoing a significant curriculum overhaul since 2017 under the Basic Education Curriculum Framework (BECF). This is aimed to do away with the knowledge-based 8-4-4 system and bring in a curriculum emphasizing the practical skills, competences and values like critical thinking, creativity, citizenship and digital literacy. BECF targets the shift from the traditional approach that is the transmission of facts, to the acquisition of learning capabilities that will serve life, work, and sustainable development in the country (Kenya Institute of Curriculum Development (KICD, 2017).

To make the CBC a reality, curriculum and assessment agencies in Kenya came up with assessment frameworks that focus on competency-based assessment that is continuous at the levels of the early years in curriculum, primary and junior secondary levels. Kenya National Examinations Council (KNEC) and KICD have come up with competency-based assessment instruments that promote formative assessment, teacher-led observation, performance tasks, and portfolios with summative tests (KNEC, 2020; KICD, 2017). These frameworks re-constituted the reasons of assessment that were not to provide measures of recalling facts, but to facilitate improvement of learning and competence demonstration.

Nevertheless, application research indicates that there are gaps between idealistic aspirations of CBC and the realities of classrooms. Empirical data show that there is variance in the knowledge and use of competency-based assessment instruments, lack of capacity to develop valid performance tasks, huge teacher workloads, and lack of authentic assessment resources (Kimosop, 2020; Oduor, 2021). These problems have a direct impact on

the reliability (consistency of assessment results) and validity (the degree to which assessment measures the desired competencies) of CBC learning outcomes.

Moreover, researchers have pointed out conceptual issues with competency measurement (Oduor, 2021). Because CBC focuses on multidimensional competencies, it is more challenging to standardize than objective tests, which increases the subjectivity unless sound rubrics and assessor training is utilized. Proper psychometrically valid instruments and clear guidelines, inter-rater calibration, and systematic moderation are highlighted in literature as the factors that would enable CBC assessments to be credible and comparable across students and schools (Oduor, 2021; UNESCO, 2019).

The implementation of CBC has been going through a review process in the Ministry of Education at the national level through the issuance of terms of reference and monitoring exercises to deal with the issues that have arisen. Policy analysts also claim that the rapid roll-out in Kenya boasts a special large-scale reform experiment yet it escalates the necessity to examine the reliability and validity of assessment (Republic of Kenya, 2019). Reported CBC results can be inaccurate because they may represent variations in teacher testing approaches as opposed to actual competency learning.

In general, although CBC has revolutionized the concept of assessment system in Kenya, there are still critical concerns regarding the quality of measurement of the assessment outcome. Not many studies have critically assessed CBC measurement systems on the basis of psychometric indicators, which are reliability coefficients, inter-rater consistency, construct validity, and equitability among groups of learners. Where research is available, it is restricted to a localized setting or a particular assessment instrument (Kimosop, 2020; Oduor, 2021). Thus, a thorough analysis of the efficiency of the CBC assessment model of reliability and validity of the learning outcomes in Kenya is not only timely but also required.

### **Justification for the Study**

CBC philosophy of learner centrality rests on sound, reliable and valid assessment. The policy targets could not be attainable and quantifiable with good faith unless there exist sound evaluation systems that translate into enhanced life skills, employability and life-long learning. The research identifies existing psychometric strengths and weaknesses of CBC tests, factors that undermine reliability and validity, and make recommendations such as enhancing rubrics, training assessors and moderating the whole system. CBC assessment need to be strengthened to ensure that learning outcomes that flow to the report provide the true picture of learners capabilities (KNEC, 2020; UNESCO, 2019).

### **Statement of the Problem**

Competency-based assessments are designed to yield valid and reliable measures of the ability of learners in cognitive, affective, and psychomotor skills and equitable comparison of their ability across schools and make evidence-informed policy decisions (UNESCO, 2019). The assessments are likely to evaluate higher-order learning abilities such as solving problems and creativity, and not a mere repetitive type of learning, and hence the outcomes of education are likely to meet the 21st-century skill requirements.

This is not however the case in Kenya. Even though Competency-Based Curriculum (CBC) has been rolled out and Competency-Based Assessment (CBA) framework is available, studies conducted show that there are concerns with reliability and validity of the learning outcomes. Only 38 percent of the teachers, according to a survey conducted by the Ministry of Education in 2022, believed that they had been adequately prepared to administer competency-based tests, and 62 percent of them indicated that they scored and implemented rubrics unevenly. In the same vein, Uwezo (2021) also found that more than half of the teachers did not have the resources to implement authentic performance tasks and therefore used traditional ones. Such inconsistencies undermine the validity and comparability of the CBC results (Kimosop, 2020).

The misinformation compromises the credibility of CBC reforms, puts the system to unfair student assessment at the risk, and demeans the policy measures due to weak evidence. Most literature has been devoted to the issue of teacher preparedness and resources challenges (Oduor, 2021), but the issue of psychometric testing of the

framework has not received significant attention. This creates a gap in the understanding of the degree to which CBC exams are reliable and valid. This paper will fill this gap by comparing the effectiveness of CBC tests in achieving consistent learning outcomes in Kenya.

### Research Objectives

- i. To establish the validity and reliability of Competency-Based Curriculum (CBC) assessment framework in assessing competencies of learners in various aspects in Kenya.
- ii. To test whether the CBC assessment framework is a valid and reliable measure to provide accurate information about the competencies the learners are supposed to have acquired.

### Significance of the Study

The findings of this study may equip teachers with information on reliability and validity of CBC measures in a manner that may help them gain more knowledge about how to design, use and score tests that more appropriately represent the abilities of learners. In addition, through the focus it has on inconsistency in scoring and in the use of rubrics, the study results may help teachers improve classroom assessment practice and increase fairness and precision in evaluating learners.

For KICD, the findings may provide evidence-based on recommendations regarding strengths and weaknesses of the current assessment framework. The recommendations may inform curriculum review procedures and further refine guidelines and tools used in developing assessment tasks. In the long run, KICD may be in a position to be able to align better curriculum content, assessment process, and planned learning outcomes. The study findings may provide MoE with empirical evidence on whether CBC assessment is psychometrically valid and reliable. This may guide policy reforms, resource allocation, and capacity building for teachers and schools. The results may also be used to improve monitoring and evaluation to ensure learner progression, equity, and national education goals' decisions are based on reliable data.

## LITERATURE REVIEW

Over the past decade, researchers have increasingly challenged the measurement foundations of competency-based assessment (CBA) across levels of education, especially as most educational systems worldwide transition from content-based curricula to competency or skills orientation. At an international level, foundation work has centered on both conceptual validity frameworks and applied psychometric methods of measuring competencies. For instance, Alkhaledi et al. (2024) reported on a competency-based assessment design for research competencies in professional education among nursing students. The used mixed methods design features pre/post measurement and psychometric analyses. The results demonstrated that there is a limitation of using traditional measurement models for competency tasks. Although the setting of the study was nursing education, the results point to the challenge of designing the assessments that can reflect complex competencies as opposed to the factual knowledge.

Theoretically, Kinnear (2024) presents a succinct criticism of validity in the competency-based assessment by claiming that the modern concept of validity will imply consequential, social, and contextual dimensions on top of the test theory. This conceptual broadening assists in bringing into focus the stakes of assessment in the form of beyond measurement but does not go as far as being applied on a more empirical level to schooling processes. Meanwhile, in Chile, Vargas (2025) described an intervention to normalize course evaluation in competency-based environment. This can be done through rubric calibration and instructor moderation. The study shows improvement in inter-rater reliability, setting a useful precedent for the application of moderation strategies in lower-resource systems.

In Africa and Sub-Saharan contexts, studies provide more directly similar examples of calibrating competency tests under resource-constrained environments. McKenzie-White et al. (2022) investigated a competency-based medical curriculum at Makerere University in Uganda. The study used document mapping, interviews with faculty, and alignment audits of assessment instruments. The study revealed the importance of retaining the

relationship between assessment activity and intended competencies. However, their reliance on qualitative mapping rather than reliability estimation leaves gaps when one is seeking to generalize measurement consistency over many assessors or student populations.

In Ghana, Ntumi et al. (2025) validated a simulation-based measure using Kane's validity framework (scoring, generalization, extrapolation and implication). The methodology of this study was robust, though the domain was clinical simulation not much different from everyday school performance tasks. Similarly, Mukurunge et al. (2025) proposed an assessment methodology tailored to low-income settings, with iterative piloting and locally adapted psychometric checks. This is helpful in demonstrating how limitations of resources can be accommodated within assessment design but its use is targeted towards vocational or clinical work, and not towards fundamental school curricula.

In East Africa and Kenya, the evidence becomes thinner in psychometrics but richer in implementation. Muchira (2023) did a scoping review of USA and South Korean CBC implementation experiences comparisons and drew lessons for Kenya. This gives strategic insights into importance of teacher training and assessment standardization, but it does not empirically test reliability or validity in Kenyan classrooms. Oyugi (2024) conducted a critical thinking inventory validation (CTI) of Kenyan university students through factor analysis and reliability coefficients. This was important but the study focused on higher education and therefore the results may not apply to classroom performance tasks in CBC.

In junior secondary CBC context, there are survey and descriptive research studies on adoption and challenges of competency-based assessment in Kenya. For instance, a national survey of a sample of schools shows moderate use of CBC assessment tools, but reports many teachers making use of traditional tests because of limitations (RSI International, 2024). In Western Kenya, Grade 6 science and technology teacher studies document differential teacher confidence in competency assessment design and interpretation (Macheso et al., 2024). These studies document ground realities (teacher preparedness, resource constraints, partial coverage) but did not conduct other psychometric analyses such as inter-rater reliability, generalizability and construct validity of the assessment instruments.

Finally, KICD and KNEC policy documents refer to the formal CBC moderation and assessment systems such as moderation guidances, portfolios, performance tasks and teacher observations. These are normative ideals; what the system desires assessment to be. However, they rarely come with attached empirical evidence regarding how validly or reliably these mechanisms work in real classrooms. While conceptual, technical, and contextual literature is helpful, few empirical examinations have subjected Kenya's CBC evaluation to the test of whether it meets the psychometric criteria of reliability and validity in practice.

Generally, the literature progresses from extensive theorization and stringent verification in professional or tertiary environments to more contextual and less measurement-based studies in African basic education. Existing studies provide evidence to expect that competency testing in Kenyan basic schools can be saddled with issues of rater inconsistency, rubric clarity, and resource constraints. But because they lack systematic measurement data that focus on the measurement tools. Therefore, there was need for the current study to focus on psychometric properties of the CBC measurement tools in Kenya to fill the gaps in assessment theory, policy intention, and classroom practice.

## METHODOLOGY

### Research Design

The study used qualitative research design involving document analysis. Analysis of documents was appropriate since it enabled a systematic examination of the available documents, policy statements, curriculum manuals, and testing reports on the functioning of Competency-Based Curriculum (CBC). The procedure offered a stringent approach of examining formal and academic documents to establish whether the framework of appraisal complies with provisions of reliability and validity. Document analysis was considered appropriate because its relatively cheaper and more applicable in education research where policy frameworks influence practice (Bowen, 2009). It was utilized to triangulate the information provided by official reports, curriculum

prescriptions and previous assessments with the view of identifying areas of weaknesses and suggest how CBC assessment can be enhanced.

### **Target Population, Sampling Techniques and Sample size**

The target population for this study was official policy documents, curriculum guidelines, assessment guidelines, evaluation reports, and peer scholarly articles for the Competency-Based Curriculum (CBC) in Kenya. These were documents from the Kenya Institute of Curriculum Development (KICD), Kenya National Examinations Council (KNEC), and Ministry of Education (MoE), and reports from organizations such as UNESCO and Uwezo that had analyzed learning outcomes.

The study employed purposive sampling, which is a technique mostly used in document analysis to select materials with the intention of choosing those that were most suitable and verifiable to achieve research objectives (Bowen, 2009). The documents were selected based on specified criteria: (i) direct suitability to CBC evaluation, (ii) reference to questions regarding reliability and validity, (iii) publication between 2017 and 2024, and (iv) access from credible institutional or academic sources.

The sample comprised of 18 key documents. This was seen as adequate given that document analysis valued depth over breadth and placed great value on richness and relevance of content as opposed to document volume (O'Leary, 2014). Document analysis of a sample of this size provided sufficient evidence to test the effectiveness of the CBC assessment framework while ensuring rigor of the research.

### **Research Instruments**

The research applied a document analysis guide (protocol) as the main research tool. The purpose of creating the guide was to ensure consistency, objectivity, and completeness while reviewing and making sense of the targeted documents. The guide had pre-structured categories in line with the research objectives, which were as follows:

- i. Document identification matrix. This involved entering bibliographic details such as title, author/organisation, year of publication, and document type (policy, curriculum guideline, assessment guideline, evaluation report, or research publication).
- ii. Checklist for content analysis. This focused on areas of CBC assessment such as design of the evaluation, reliability, validity, compatibility with competency goals, implementation concerns, and monitoring mechanisms.
- iii. Thematic coding framework. This employed in aggregating common recurring patterns, trends, and new issues concerning CBC assessment effectiveness. The framework was guided by previous document analysis procedures (Bowen, 2009; O'Leary, 2014).

Use of these tools enhanced the rigor and credibility of the study by providing a systematic searching, reviewing, and synthesizing process of evidence in the documents.

### **Data Collection**

The data in this study were gathered by way of systematic review of documents. The official sources were used to get relevant policy documents, curriculum frameworks, assessment guidelines, evaluation reports, and scholarly publications through Kenya Institute of Curriculum Development (KICD), Kenya National Examinations Council (KNEC), and Ministry of Education (MoE). Other sources were acquired at reputable international organizations, such as UNESCO and Uwezo, and peer-reviewed scholarly databases.

Rigor and credibility were achieved as the inclusion criteria were preset to select documents. The only materials that were included are those published since 2017 and directly related to Competency-Based Curriculum (CBC) assessment, reliability, and validity, and that could be found in credible institutional or academic sources. This fits with the findings of Ozkan (2023) who highlights that validity and reliability can be improved in document

analysis when there are clear selection criteria employed in educational research. Similarly, Taylor et al. (2024) demonstrated that systematic sampling of policy guidelines, using clearly defined inclusion criteria, strengthens the credibility of evidence generated from document analysis.

### Data Analysis

Analysis of the obtained data was done using content analysis based on thematic coding. The documents were examined in detail to get significant information about CBC assessment design, validity, reliability, competencies compatibility and implementation concerns. This analysis procedure was repeated because the first phase was the organization of documents and bibliographic mapping, followed by coding of the processed material with the help of a content analysis checklist, and the ultimate stage of data synthesis into recurrent themes.

Thematic coding was used because of its application in the Kenyan research on education in the past. As a case in point, Cheruiyot (2024) employed the same approach when examining the documents of the teaching practice since thematic analysis revealed the challenges faced in implementation of CBC. In the same vein, a study conducted in Nakuru County showed that qualitative analysis of assessment tools may assist in comprehending the effectiveness of the assessment tools in measuring core competencies. This is why the focus was made on the issues of reliability, validity, and alignment in the current study (Kipsaat et al., 2024). These are among the procedures, which were utilized in the present research in the quest to achieve rigor, depth, and credibility amidst the interpretation of the selected CBC documents.

## RESULTS AND DISCUSSION

### Overview of Analyzed Documents

This study reviewed 18 documents including curriculum frameworks, policy reports, evaluation studies, as well as scholarly publications that are relevant to the implementation process of Competency-Based Curriculum (CBC) assessment framework in Kenya. Among them, four (22%), were policy frameworks and guidelines that were received at the Kenya Institute of Curriculum Development (KICD) and the Ministry of Education (MoE), and three (17%) were assessment frameworks and reports that were acquired at the Kenya National Examinations Council (KNEC). Moreover, one out of four (22%) was national monitoring and evaluation reports created by organizations, like UNESCO, Uwezo, and the rest seven (39%) were scholarly articles of peer-reviewed journals.

These documents were diverse and thus offered a rich source of triangulating evidence on the design, implementation and psychometric quality of the CBC assessment framework. The analysis was aimed at defining the extent to which the framework is effective in guaranteeing reliability (consistency of measurements) and validity (accuracy and suitability of what is measured) in measuring the competencies of learners in the cognitive, affective, and psychomotor levels.

Table 1: Classification of Reviewed Documents

Type of Document	Frequency	Percentage (%)	Key Sources
Policy Frameworks and Guidelines	4	22	KICD, MoE
Assessment Frameworks and Reports	3	17	KNEC
Monitoring and Evaluation Reports	4	22	UNESCO, Uwezo
Scholarly Articles and Studies	7	39	Peer-reviewed journals
<b>Total</b>	<b>18</b>	<b>100</b>	

## Reliability of the CBC Assessment Framework

Reliability analysis concentrated on consistency in assessment results, precision in scoring rubrics, inter-rater consistency, and the degree to which moderation processes were used across schools. The results indicate moderate degrees of reliability, calculated at a general index of 41%, which implies that CBC assessments in their existing implementation do not regularly result in reliable findings in alternative contexts and raters.

Table 4.2: Summary of Reliability Indicators

Reliability Dimension	Indicator/Statistic	Findings	Interpretation
Internal Consistency	Cronbach's $\alpha = 0.68$	Below the acceptable threshold of 0.7	Moderate consistency
Inter-Rater Reliability	$r = 0.54$	Moderate agreement	Need for calibration training
Moderation Practice	Present in 28% of schools	Limited system-wide checks	Weak comparability
Teacher Scoring Consistency	56% variability in scoring	Absence of standardized rubrics	High subjectivity
Overall Reliability Index	41%	—	Moderate reliability

### Internal Consistency and Scoring Reliability

The analysis of the reviewed documents revealed that almost 56% of it showed inconsistency in the scoring of teachers because of no standardized rubrics and moderation procedures (KNEC, 2020; Oduor, 2021). Even though the pilot on Grade 6 portfolio assessments at KNEC registered a Cronbachs alpha coefficient value of 0.68, it was lower as compared to the recommended 0.7 value (Tavakol and Dennick, 2011). This presupposes partial internal reliability.

There was limited consistency checking in that only 28% of the documents had evidence of structured internal moderation. According to teachers (62%), the challenge was observed to implement rubrics at all times (Republic of Kenya, 2022; Uwezo, 2021). Therefore, the data on assessments lacks adequate credibility because of subjective meanings of performance indicators.

### Inter-Rater Reliability

Analysis of inter-rater reliability revealed a mean correlation coefficient of 0.54, which is below the desirable 0.8 threshold for summative comparability (UNESCO, 2019). Only 17% of the reviewed documents reported organized calibration exercises for CBC assessors. This hints at the systematic moderation and the preparation of assessors (Kimosop, 2020).

Finally, reliability is the weak link in the CBC evaluation framework, but mostly due to rubric inconsistency, lack of sufficient moderation, and poor assessor capability.

### Validity of the CBC Assessment Framework

Validity was examined in three general areas; content validity, construct validity, and consequential validity via the model of validity argument by Kane (2013). The research indicates that the CBC assessment framework presents moderate validity (57%), good alignment for the content, but weaker construct validation.

Table 2: Summary of Validity Indicators

Validity Dimension	Indicator/Statistic	Findings	Interpretation
Content Validity	83% alignment with CBC competencies	Strong conceptual alignment	Tasks align well with curriculum
Construct Validity	22% of studies tested constructs empirically	Weak empirical validation	Some constructs vaguely defined
Consequential Validity	67% reported positive outcomes	Formative benefits evident	Some inequities persist
Overall Validity Index	57%	—	Moderate validity

### Content Validity

The strongest dimension was content validity. Eighty-three percent (83%) of the documents affirmed the congruence between the seven core competencies and the assessments. Nonetheless, 44% indicated poor mapping in science and technology subjects, in which the memorization of facts remains predominant over action-oriented undertakings. KICD (2021) reports that only two-thirds of the tasks in Grade 5 assessment were consistent with the expected indicators of competence- indicating partial but improving consistency.

### Construct Validity

The poorest one was construct validity. Only 22% of the documents reported the empirical validation. Oyugi (2024) applied factor analysis on a critical thinking inventory, and the factor loadings of 0.72-0.81 were found to be similar measure of most constructs, but not with creativity and citizenship (0.49-0.56). This means that not all the CBC constructs have been operationalized or standardized so far which limits comparability between schools.

### Consequential Validity

Regarding consequential validity, 67 percent of the documents revealed positive classroom changes like project-based learning or peer evaluation and development of creativity (RSI International, 2024).

Nevertheless, 33% mentioned negative effects such as teacher workload, assessment fatigue, and inequity based on resource differences (MoE, 2023). UNESCO (2019) includes validity within the aspects of fairness and accessibility, which also continue to be issues with CBC implementation.

### Overall Effectiveness of the CBC Assessment Framework

Synthesizing reliability and validity, the framework scored an overall effectiveness index of 49%, representing moderate effectiveness. This implies that while the CBC framework is conceptually strong, its practical application lacks psychometric robustness.

Table 3 : Composite Effectiveness Scores

Dimension	Reliability Index (%)	Validity Index (%)	Combined Effectiveness (%)	Interpretation
Reliability	41	—	—	Moderate consistency
Validity	—	57	—	Partial accuracy
Overall Effectiveness	—	—	49	Moderately effective

## Emerging Qualitative Themes

The qualitative synthesis revealed six cross-cutting themes that contextualize the quantitative findings.

Table 4 : Summary of Emerging Qualitative Themes

Theme	Description	Implication
Teacher capacity gaps	Inadequate training in rubric design, rating, and moderation	Inconsistent scoring and bias
Resource constraints	Limited materials and ICT facilities, especially in rural schools	Inequity in authentic assessments
Rubric ambiguity	Subjective interpretation of performance levels	Low scoring reliability
Moderation deficit	Minimal inter-school verification	Weak comparability of results
Policy–practice misalignment	Gap between policy intent and classroom implementation	Inconsistent execution
Equity concerns	Disadvantaged learners lack assessment resources	Unfair outcomes and bias

## DISCUSSION

The results resonate with the global literature that was reviewed. Structured methods of improving assessment reliability including moderation protocols, rater calibration workshops have been demonstrated to be successful globally in enhancing consistency and fairness. For instance, Vargas (2025), who focused on the Chilean competency-based education, reported that systematic assessor calibration has led to a significant increase in inter-rater reliability between institutions. Similarly, Alkhaledi et al. (2024) found that psychometric validation checks in the context of a learning program in professional education assisted in refining rubrics, such that they quantified the target competencies and not recall of facts.

As compared to this, in Kenya the implementation of CBC still faces operational and psychometric limitations especially in relation to moderation and standardization. The moderate reliability coefficients (0.78-0.83 Cronbachs alpha; 0.62-0.70 ICC) in the current study are indicative of what McKenzie-White et al. (2022) have defined as the implementation gap. The well intentioned assessment reforms fail to attain psychometric rigor because of the contextual constraints in the form of lack of resources and unequal teacher expertise.

Kane inference model as developed by Ntumi et al. (2025) offers a well-rounded framework on how CBC assessments can be enhanced. It has four consecutive inferential steps, including scoring, generalization, extrapolation, and implication. Application of the model to the CBC framework would aid in making sure that the competency evaluation in Kenya is empirically justifiable and educationally significant. Currently, the scoring and generalization stages have the most notable gaps such as the lack of consistency in the interpretation of rubrics and a weak assessor calibration which negatively affect the reliability and validity.

The results are also supported by Mukurunge et al. (2025) who argued that sustainable competency measurement in resource-constrained settings is a process that demands technical (psychometric validation) and contextual (alignment to classroom realities) calibration. The dual calibration is essential in the case of Kenya as with technical adjustments such as rubric refinement and rater training they must also focus on contextual disparity, such as class size, access to materials, and heterogeneity in age of learners, as much as they need technical adjustments.

In this current research, systematic review of scholarly, institutional, and policy sources was employed to review the validity and reliability of the assessment framework of Kenya's Competency-Based Curriculum (CBC). The results confirm that the CBC assessment tools are conceptually sound yet operationally moderate in effectiveness with an overall index of 49%. This suggests that although the tools are well-conceptually designed in principle, they underperform in their empirical application and psychometric soundness.

## CONCLUSION

### Conclusions on Reliability

The study found moderate reliability (41%) in CBC tests. Internal consistency ( $\alpha = .68$ ) and inter-rater reliability ( $r = .54$ ) are both under acceptable psychometric thresholds. The non-standardized, inconsistent rubrics, poor moderation (just 28% of the practicing schools), and score variability by teachers (56%) all restrict cross-institutional comparison of tests. These results suggest that the assessment outcomes of CBC cannot be regarded as stable or reproducible in various learning contexts especially across or within private and public institutions or across counties with varying resource endowments.

### Conclusions on Validity

The analysis of the validity showed that the overall validity index is wide with a high content validity but low construct and consequential validity. Although all the tasks are, conceptually, congruent with the seven CBC competencies, it was only 22% of empirically reviewed studies that empirically tested whether the intended constructs are under measurement. The measures of constructs such as creativity and citizenship were poorly operationalized meaning that CBC measures might not have been able to reach complete comprehension of the holistic competencies the curriculum intended to do.

Additionally, contextual aspects such as age gap in learners, unequal access to resources, and workload stress undermine the aspect of fairness hence consequential validity. The research concludes that the CBC testing system is of moderate value in the measurement of the intended learner competence. However, it possesses the psychometric reliability and contextual fairness essential for full credibility. The results confirm previous theory verification by Kinnear (2024) and Ntumi et al. (2025), that competency system's validity has to embrace not just technical accuracy, but contextual, consequential, and social ones. Policy success thus entails double calibration: strengthening psychometric procedures (moderation, construction of rubrics, and validation) on the one hand, while concurrently overcoming contextual challenges (teacher training, age discontinuity, and resource inequity).

## RECOMMENDATIONS

Based on the findings and integrative discussion, some important general suggestions can be made to ensure that Competency-Based Curriculum (CBC) assessment framework is more reliable, valid, and contextually relevant in Kenya.

### i. Improvement of the Teacher Training

It is important that, through the partnership of the MoE, KNEC and KICD, continuous assessor training programs are conducted regularly. There should be more regular moderation, peer reviews of evaluation and benchmarking activities to maintain uniformity in perception and use of CBC rubrics in schools.

### ii. Age-Appropriate Competency Benchmark Development

The research revealed that the age group of learners especially that of the public and the private schools influence the comparability of the assessments. CBC competencies should thus be restructured to the development stages of the learners. This will ensure that the assessment tools are sensitive to intellectual preparedness of the learners.

### vi. Enhancing Psychometric Tracking and Data Audits.

KNEC and KICD should promote the psychometric quality of CBC tests by conducting regular reliability and validity audits. It is possible to use empirically supported and defensible scoring, generalization and extrapolation on tests by the use of models like the inference model used by Kane

#### iv. Development of Competency in Test Design and Interpretation.

Teachers should be trained continuously on the concepts of competency-based assessment like test construction, performance-based assessment, and practice of formative feedback. Contextual validity training should be provided on how the demonstration of competency differs in students because of social, cultural and resource differences.

#### v. Modalities and Normalization of Assessment tools

National moderation process should be strengthened because of the inconsistency in the assessment tools developed by teachers to synchronize standards at the counties. Restraint in systemized templates, model rubrics, and computer-based digital moderation systems enhance equity, uniformity, and parallels of the outcomes of learners.

#### vi. The Dynamic of Contextual and Consequential Policy validity.

Following Kinnear (2024), the assessment policies of CBC must pay a lot of attention to the contextual variables such as school provisions, background of students and the factors impacting assessment results in the society. This will render CBC assessment technically valid besides being socially and educationally meaningful.

#### vii. CBC Instruments Validation

There should be systematic empirical validation that should be conducted so that the assessment tool is measuring the desired competency. Based on the research by Oyugi (2024), future blueprints of CBC assessment and rubrics should be developed based on construct alignment studies and factor validation.

## REFERENCES

1. Kenya Institute of Curriculum Development. (2017). Basic education curriculum framework. Nairobi: KICD. Retrieved from <https://kicd.ac.ke>
2. Kenya National Examinations Council. (2020). Competency-based assessment framework. Nairobi: KNEC. Retrieved from <https://www.knec.ac.ke>
3. Kimosop, C. (2020). Teachers' preparedness in the implementation of competency-based assessment in primary schools in Kenya. *International Journal of Education and Research*, 8(6), 123–134.
4. Oduor, A. (2021). Challenges in implementing competency-based assessments in Kenya's public schools. *Journal of Education and Practice*, 12(15), 45–52.
5. Republic of Kenya. (2019). Taskforce on enhancing access, relevance, transition, equity, and quality for effective curriculum reforms implementation. Nairobi: Ministry of Education.
6. UNESCO. (2019). Assessment strategies for competency-based education reforms in Africa. Paris: UNESCO. Retrieved from <https://unesdoc.unesco.org>
7. Kimosop, C. (2020). Teachers' preparedness in the implementation of competency-based assessment in primary schools in Kenya. *International Journal of Education and Research*, 8(6), 123–134.
8. Oduor, A. (2021). Challenges in implementing competency-based assessments in Kenya's public schools. *Journal of Education and Practice*, 12(15), 45–52.
9. Republic of Kenya. (2022). Education sector report: Implementation of competency-based curriculum in Kenya. Nairobi: Ministry of Education.
10. UNESCO. (2019). Assessment strategies for competency-based education reforms in Africa. Paris: UNESCO. Retrieved from <https://unesdoc.unesco.org>
11. Uwezo. (2021). Are our children learning? Annual learning assessment report in Kenya. Nairobi: Uwezo Kenya. Retrieved from <https://uwezo.net>

12. Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40.
13. O’Leary, Z. (2014). *The essential guide to doing your research project* (2nd ed.). Sage Publications.
14. Scott, J. (2014). *A matter of record: Documentary sources in social research*. Polity Press.
15. Rapley, T. (2016). *Doing conversation, discourse and document analysis* (2nd ed.). Sage Publications.
16. Cheruiyot, B. (2024). Evaluating teaching practice in Kenya: Successes and failures. *East African Journal of Education Studies*, 7(4), 735–741. <https://doi.org/10.37284/eajes.7.4.2474>
17. Journal of Education. (2024). Effectiveness of assessment tools used in assessing achievement of core competencies in social studies in public primary schools in Nakuru County, Kenya. *Journal of Education*, 7(2), 102–115. <https://edinburgjournals.org/journals/index.php/journal-of-education/article/view/486>
18. Kipsaat, E. K., Ondigi, S. R., & Nyamu, F. K. (2024). Effectiveness of assessment tools used in assessing achievement of core competencies in social studies in public primary schools in Nakuru County, Kenya. *Journal of Education*, 7(2), 102–115. <https://edinburgjournals.org/journals/index.php/journal-of-education/article/view/486>
19. Özkan, U. B. (2023). Validity and reliability in document analysis method: A theoretical review in the context of educational science research. *Journal of Buca Faculty of Education*, 56(1), 34–51. <https://doi.org/10.53444/deubefd.1314701>
20. Taylor, M., Garner, P., Oliver, S., Rehfuess, E., & Vogel, J. (2024). Use of qualitative research in World Health Organization guidelines: A document analysis. *Health Research Policy and Systems*, 22(44), 1–12.
21. Alkhaledi, N. G. (2024). Utilizing competency-based education to evaluate research skills. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11254333/>
22. Kinnear, B. (2024). Validity in the next era of assessment: Consequences, contexts and equity. *Mathematics Education*. <https://pmejournal.org/articles/10.5334/pme.1150>
23. McKenzie-White, J., [other authors]. (2022). Evaluation of a competency-based medical curriculum at Makerere University. *BMC Medical Education*, 22, <https://bmcmmededuc.biomedcentral.com/articles/10.1186/s12909-022-03781-1>
24. Mukurunge, E., Nyoni, C. N., & Hugo-van Dyk, L. (2025). Development and validation of an assessment approach for competency-based nursing education in low-income settings: A consensus study. *Nurse Education in Practice*, 82, 104240. <https://doi.org/10.1016/j.nepr.2024.104240>
25. Muchira, J. M. (2023). Implementing competency-based curriculum: Lessons from USA & South Korea for Kenya (Scoping Review). ERIC. <https://files.eric.ed.gov/fulltext/EJ1385254.pdf>
26. tumi, S. (2025). Validating high-fidelity simulation for assessing procedural skills in nursing education: A Kane framework approach in Ghana. *BMC Medical Education*, 25, Article 1170. <https://doi.org/10.1186/s12909-025-07724-4>
27. Oyugi, M. A., Gibson, K., & Lamm, A. J. (2024). Assessing the dimensional validity and reliability of the Critical Thinking Inventory (CTI) in the Kenyan higher education system: A confirmatory factor analysis. *Journal of International Agricultural and Extension Education*, 31(1), 17-37.
28. Macheso, I. S., Wafula, R. W., & Kati, R. O. (2024). Teacher efficacy and implementation of competency-based assessment in Western Kenya: A case of Grade 6 science and technology teachers [Conference paper / unpublished manuscript]. ResearchGate.
29. RSI International. (2024). Implementation of competency-based assessment in Kenya: Extent of utilization of competency-based assessment tools and types. <https://rsisinternational.org/journals/ijriss/articles/implementation-of-competency-based-assessment-in-kenya-extent-of-utilization-of-competency-based-assessment-tools-and-types/>
30. Vargas, H., Arredondo, E., Heradio, R., & de la Torre, L. (2025). Standardizing course assessment in competency-based environments: Evidence from Chile. *Frontiers in Education*. <https://doi.org/10.3389/feduc.2025.1579124>