E-ISSN: 1115 - 8425 P-ISSN: 3027-012X
MPRI-JAEE | INTERNATIONAL JOURNAL OF ASSESSMENT & EVALUATION IN EDU.
VOL. 09 NO. 8 – JULY. 2025

AN ITEM DIFFICULTY ANALYSIS OF THE 2023 CEP AND FSLC EXAMINATIONS IN THE CENTRE REGION OF CAMEROON: IMPLICATIONS FOR VALID ASSESSMENT PRACTICES IN BASIC EDUCATION

DR. ERNESTINE WIRNGO TANI; & DR. LAWRENCE NTAM NCHIA

Department of Sciences of Education, Higher Teacher Training College, Yaounde, University of Yaounde 1, Cameroon

Corresponding Author:

ernestine.wirngo@univ-yaounde1.cm

DOI: https://doi.org/10.70382/mejaee.v9i8.051

Abstract

This study analyzed the item difficulty of the 2023 Certificat d'Études Primaires (CEP) and First School Leaving Certificate (FSLC) examinations in the Centre Region of Cameroon, with emphasis on validity under the Competency-Based Approach (CBA). A descriptive evaluative design was adopted, using examination datasets from MINEDUB and the Department of Certification and Exams. After data cleaning in SPSS version 23, item difficulty indices (P) were calculated using Classical Test Theory, with thresholds for easy (0.76–1.00), appropriate difficulty (0.26–0.75), and difficult (0.00–0.25). A curriculum alignment review and gender-based comparisons complemented the analysis. Findings indicate that 71.4% of CEP and 66.7% of FSLC sub-disciplines were "easy," with very high pass rates in Reading Comprehension (97.27%), French (95.56%), and Production d'Écrits (93.98%). Only four subjects, CEP Dictée + Questions (53.05%), CEP Problèmes (66.73%), FSLC Guided Writing (72.68%), and FSLC Mathematics 2 (70.23%), met the "appropriate difficulty" range. Mathematics 1 (FSLC) was the sole "difficult"

Keywords: Item Difficulty, Docimology, Psychometrics, Competency-Based Approach, Cameroon, Primary Education.

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

subject, with a pass rate of 49.5%. Gender analysis showed that girls outperformed boys in language-based subjects, while boys slightly excelled in numeracy. The predominance of easy items undermines test validity, limiting discrimination between learners and neglecting higher-order skills required by the CBA. The study recommends revising exam construction to include a balanced distribution of item difficulty, integrating higher-order tasks aligned with Bloom's taxonomy, pre-testing items, conducting routine psychometric reviews, and training examiners in assessment literacy. Implementing these measures will enhance fairness, curriculum alignment, and support Cameroon's pursuit of SDG 4 on inclusive and equitable quality education.

Introduction

Item analysis is an essential component of educational measurement, evaluating test items based on difficulty, discrimination, and reliability. Among these, item difficulty is pivotal in determining whether a test appropriately assesses the range of learners' abilities. In Cameroon, national examinations such as the Certificat d'Études Primaires (CEP) and the First School Leaving Certificate (FSLC) play a crucial role in evaluating learning outcomes at the primary school level.

Recent statistics from the 2023 CEP and FSLC examinations in the Centre Region revealed significant variations in pass rates across disciplines, suggesting inconsistencies in item difficulty. These discrepancies warrant a scientific inquiry into the quality and fairness of assessment practices in these high-stakes exams.

Context and Justification

Docimology, the scientific study of examinations, emphasizes objectivity, validity, and reliability in assessment. According to Lupian (2022), test items should neither be overly easy nor excessively difficult to ensure they effectively differentiate learners' abilities. Data from the 2023 CEP and FSLC highlight high pass rates in subjects like Dictation, Science, and Reading Comprehension—indicating a potential overuse of easy items very high pass rates (above 90%). In contrast, subjects like Mathematics 1 (FSLC – 49.5%) and Dictée + Questions (CEP – 53.05%) demonstrated more appropriate difficulty levels.

The increasing adoption of the Competency-Based Approach (CBA) in Cameroon makes it imperative to ensure that assessments accurately measure learner competencies. A lack of empirical review of item quality risks undermining curriculum implementation (Nitko & Brookhart, 2014). This study provides actionable insights for curriculum designers, examination boards (e.g., DECC-MINEDUB), and educators to better align assessments

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

with learning outcomes, thus supporting improved educational outcomes and policy development.

Problem Statement

In a competency-based education framework, assessment is expected to include a balanced mix of easy, moderate, and difficult questions that can effectively capture the full spectrum of learners' abilities. Classical Test Theory emphasizes that test items within a moderate difficulty range (0.30–0.75) are the most informative since they distinguish stronger learners from weaker ones while still aligning with curriculum expectations (Gulliksen, 1950; Nitko & Brookhart, 2014). Ideally, examinations in Cameroon should embody the principles of the Competency-Based Approach (CBA), which stresses not only the recall of knowledge but also the development of higher-order thinking, problem-solving, and application of learning in real-life situations (Bloom, 1956; Biggs, 1996).

The reality, however, shows a different picture. Analysis of the 2023 Certificat d'Études Primaires (CEP) and First School Leaving Certificate (FSLC) examinations in the Centre Region revealed that the majority of test items were overly simple, with 71.4% of CEP subjects and 66.7% of FSLC subjects classified as "easy." Pass rates were particularly high in Reading Comprehension (97.27%), French (95.56%), and Production d'Écrits (93.98%), suggesting that these tests may not be measuring the intended cognitive depth. Only a few sub-tests such as Dictée + Questions (53.05%) and Mathematics 2 (70.23%), fell within the desirable range of difficulty, while Mathematics 1 (49.5%) was the only subject identified as difficult. These findings resonate with research across West Africa showing that poorly designed items can distort learners' competencies and weaken the validity of assessment outcomes (Omo-Egbekuse, 2017; Afang & Ekpoh, 2020). In addition, regional studies such as PASEC 2014 and 2019 report that over 60% of pupils in Cameroon finish primary school without mastering foundational skills in literacy and numeracy, underscoring persistent learning challenges despite seemingly high pass rates.

This discrepancy between inflated success rates and weak competency attainment highlights a fundamental disconnect between curriculum expectations, teaching practices, and national assessments. The predominance of easy questions provides an illusion of success while masking deeper gaps in learning achievement. Moreover, the absence of routine psychometric reviews of national exams raises concerns about fairness, validity, and the effective implementation of the CBA. Addressing this gap requires a systematic analysis of item difficulty to guide the construction of fairer and more reliable examinations that reflect true learner competencies, advance equity in education, and align with Cameroon's commitment to SDG 4 on quality education.

Main Research Question:

- To what extent do the test items in the 2023 CEP and FSLC examinations reflect appropriate levels of difficulty in accordance with psychometric standards?

Specific Questions:

- What are the item difficulty levels across sub-disciplines in the 2023 CEP and FSLC?
- How do item difficulty levels correlate with learner performance?
- Are there significant differences in item difficulty by subject or gender?
- What are the implications of current item difficulty levels for teaching and learning?

Aim:

- To analyze the item difficulty levels in the 2023 CEP and FSLC examinations and assess their implications for valid and equitable measurement of learners' competencies in the Centre Region of Cameroon.

Specific Objectives:

- To determine the item difficulty indices for sub-disciplines in the 2023 CEP and FSLC exams.
- To analyze the relationship between item difficulty and learner performance across subjects.
- To compare item difficulty levels by gender and discipline.
- To find out the implications of current item difficulty levels for teaching and learning outcomes.

Theoretical Framework and Empirical Literature Review

A sound understanding of item difficulty is fundamental to the construction of valid and reliable assessments in basic education. This study draws on Classical Test Theory (CTT), Bloom's Taxonomy, and Constructive Alignment Theory to analyze item difficulty.

According to CTT (Gulliksen, 1950; Crocker & Algina, 2006), item difficulty is determined by the proportion of students answering an item correctly. Ideally, items should be neither too easy nor too difficult to maintain discriminatory power and fairness. CTT posits that an observed test score consists of a true score and an error component. CTT introduces the item difficulty index (P), which is the proportion of students who answer an item correctly. This index helps developers classify items as easy, appropriately difficult, or difficult.

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

Beyond psychometric properties, the cognitive complexity of questions, as described by Bloom's Taxonomy (1956), also impacts item difficulty. According to Nitko and Brookhart (2014), the effective design of assessment items requires matching these levels of cognitive demand to the expected outcomes of instruction. Simplistic items that focus solely on factual recall may fail to measure the deeper understanding and problem-solving skills students are expected to develop. Popham (2008) similarly argues that well-constructed assessments should challenge learners and align with the higher-order thinking processes embedded in the curriculum.

Constructive Alignment Theory, proposed by Biggs (1996), further reinforces the need for coherence between learning objectives, instructional strategies, and assessment tasks. Misalignment, such as when assessments are too easy or too difficult relative to what was taught, can distort the measurement of student learning and lead to invalid conclusions.

Empirical research (Afang & Ekpoh, 2020; Lupian, 2022) confirms that balanced item difficulty enhances the validity of standardized assessments. In West Africa, Omo-Egbekuse (2017) found that poorly constructed items could distort student competency evaluation. Similarly, Suen and Ary (1989) argue that assessment must consider both psychometric quality and curriculum goals. The Ministry of Basic Education (MINEDUB, 2023) data for the Centre Region's 2023 exams, showing significant variations in pass rates, further justifies this study.

Methodology

This study uses a descriptive evaluative research design that combines quantitative analysis of examination data with a review of curriculum documents. The primary goal is to describe and evaluate the difficulty levels of the 2023 CEP and FSLC examination items in relation to the national curriculum's intended learning outcomes.

The study was conducted in the Centre Region of Cameroon, which is a representative area due to its diverse school types and demographics. The target population included all students who took the 2023 CEP and FSLC examinations in this region, as well as the examination papers, marking guides, and relevant curriculum documents. A purposive sampling strategy was used to select specific subjects for in-depth analysis, such as Dictation, Mathematics 1, and Science, due to their varied performance patterns.

Data were collected using examination result datasets from MINEDUB and the Department of Certification and Exams (DCE). Raw examination datasets from MINEDUB and the Department of Certification and Exams were imported into SPSS version 23 and subjected to data cleaning procedures, including double-entry verification, consistency checks across subjects and candidate codes, detection and removal of missing or duplicate entries, and recoding of categorical variables to ensure accuracy and reliability prior to analysis.

The primary metric for evaluating item difficulty was the item difficulty index (P), calculated as the number of correct responses divided by the total number of candidates who attempted the item. Based on standard psychometric thresholds, items were categorized as: difficult (P = 0.00–0.25), of right difficulty (P = 0.26–0.75), and easy (P = 0.76–1.00).

Descriptive statistics were used to compare item difficulty across disciplines and gender. Qualitative content analysis was also used to assess the alignment between test items and curriculum objectives, particularly regarding cognitive demand based on Bloom's Taxonomy.

Data Presentation and Analysis

This section presents and analyzes the data collected from the 2023 CEP and FSLC examination results in the Centre Region of Cameroon, focusing on the item difficulty levels across specific sub-disciplines. The analysis aims to determine how appropriately test items were constructed in terms of difficulty, their alignment with curriculum goals, and the implications for educational measurement and evaluation.

The table below summarizes candidate performance by discipline, showing the number of pupils who passed, the percentage pass rate, and corresponding item difficulty levels.

Table 1: Presentation of CEP 2023 Results by Discipline and Item Difficulty

No	*Discipline*	*% Admitted*	*Difficulty Index*	*Difficulty Level*
1	Dictée + Questions	53.05%	0.5305	Right Difficulty
2	Production d'Écrits	93.98%	0.9398	Easy
3	Calcul Rapide	92.40%	0.9240	Easy
4	Problèmes	66.73%	0.6673	Right Difficulty
5	Sciences	76.73%	0.7673	Easy
6	Grammaire & Vocabulaire	91.20%	0.9120	Easy
7	Reading Comprehension	97.27%	0.9727	Easy

From observation, most subjects fall within the "easy" category, especially Reading Comprehension and Directed Writing. Only two disciplines (Dictée + Questions and Problèmes) reflect a more balanced level of difficulty, falling in the "right difficulty" range.

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

Table 2: Presentation of FSLC 2023 Results by Discipline and Item Difficulty

No	Discipline	% Admitted	Difficulty Index	Difficulty Level
1	Dictation	91.20%	0.9120	Easy
2	Guided Writing	72.68%	0.7268	Right Difficulty
3	Mathematics 1	49.50%	0.4950	Difficult
4	Mathematics 2	70.23%	0.7023	Right Difficulty
5	Sciences	89.20%	0.8920	Easy
6	French	95.56%	0.9556	Easy

Similarly, most FSLC items were categorized as "easy." Mathematics 1 stands out as a "difficult" discipline with a difficulty index of 0.4950, indicating poor performance. Guided writing and Mathematics 2 fall within the "right difficulty" range.

Table 3: Comparative Analysis of Item Difficulty Across Disciplines

Category	CEP (% of disciplines)	FSLC (% of disciplines)
Easy	71.43% (5/7)	66.67% (4/6)
Right Difficulty	28.57% (2/7)	33.33% (2/6)
Difficult	0%	16.67% (1/6)

The data shows a predominance of "easy" test items in both exams. According to standard item analysis practices (Nitko & Brookhart, 2014), overuse of easy items compromises the discriminatory power of the test and may inflate pass rates without reflecting true learning. Only FSLC has one "difficult" sub-discipline, suggesting better variability than CEP.

Table 4: Gender-Based Comparison of Performance (Selected Highlights)

Discipline	% Male Success	% Female Success	Observation
Dictée + Questions (CEP)	50.26%	55.88%	Girls outperform boys
Problèmes (CEP)	64.95%	68.52%	Girls outperform boys
Mathematics 1 (FSLC)	50.69%	48.44%	Boys slightly outperform girls
Mathematics 2 (FSLC)	69.12%	71.21%	Girls outperform boys

While the performance differences are not extreme, a trend was observed. Girls generally outperformed boys in language and written tasks, such as CEP Dictée + Questions (55.88% vs. 50.26%) and CEP Problèmes (68.52% vs. 64.95%). Conversely, boys slightly outperformed girls in numeracy subjects, such as FSLC Mathematics 1 (50.69% vs. 48.44%).

Curriculum Alignment Analysis

A review of the national primary curriculum for French, Mathematics, and Science reveals clear learning objectives that require:

- Mastery of basic numeracy and problem-solving (Math)
- Application of scientific knowledge in real life (Science)
- Development of expressive language and critical thinking (French)

However, the overrepresentation of easy items, especially in language subjects, suggests that the assessments may not fully challenge learners to demonstrate higher-order cognitive skills as intended in the Competency-Based Approach (CBA).

Interpretation and Discussion of Results

This section interprets the key findings from the 2023 CEP and FSLC examination results, focusing on item difficulty levels, gender disparities, and alignment with curriculum objectives. The results are discussed in light of existing theories and empirical studies in educational measurement and evaluation.

Summary of Key Findings

- Over 70% of test items in CEP and FSLC are classified as easy, raising validity concerns.
- Only 2 out of 13 sub-disciplines meet the "right difficulty" criterion with balanced challenge.
- Mathematics 1 (FSLC) is the only domain rated as "difficult."
- Females slightly outperform males in most language-based sub-tests.
- There is partial misalignment between item difficulty and the intended competency targets of the curriculum.

The findings of this study, grounded in educational measurement theories, raise several key points regarding the quality of the 2023 CEP and FSLC examinations.

Predominance of Easy Test Items

The data shows that over 70% of items in CEP and 66.67% in FSLC were classified as “easy,” with difficulty indices above 0.76. Subjects like Reading Comprehension (97.27%), French (95.56%), and Production d’Écrits (93.98%) recorded extremely high pass rates. This trend is problematic from a docimological perspective, as excessively easy items fail to distinguish between high and low achievers, thereby reducing the validity and reliability of the assessment (Nitko & Brookhart, 2014). This may also suggest that the assessments fail to adequately measure higher-order cognitive skills as outlined in Bloom's Taxonomy, giving a potentially false sense of success. These findings corroborate with PASEC (2014 & 2019) indicates that, despite slight gains in early reading, a majority of pupils continue

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

to leave primary school without mastering minimum proficiency in language and mathematics, indicating persistent learning deficits and inequities over time.

Limited Use of Items with Right Difficulty

Only a few subjects such as Dictée + Questions (53.05%) and Problèmes (66.73%) in CEP, and Guided Writing (72.68%) and Mathematics 2 (70.23%) in FSLC fall within the "right difficulty" category. According to Classical Test Theory (CTT), items with difficulty indices between 0.30 and 0.75 are ideal because they provide meaningful information about student performance while allowing differentiation (Gulliksen, 1950; Lupian, 2022).

The limited presence of these items suggests a misalignment between curriculum expectations and assessment practices, which, as per Biggs' Constructive Alignment Theory, can hinder the achievement of deep learning outcomes.

Presence of Difficult Items and Student Weaknesses in Numeracy

The identification of FSLC Mathematics 1 as the only "difficult" subject, with a pass rate of 49.50%, highlights a significant challenge in students' foundational numeracy and problem-solving skills. This could be due to issues like poor classroom delivery of abstract concepts, overreliance on rote memorization, or a misalignment between teaching methods and assessment focus. This finding is consistent with other studies in West Africa, which have also found that mathematics often underperforms due to curriculum overload and limited teacher capacity. This finding aligns with those of Wirngo T. (2021), which indicated that, while many test items were too easy, some mathematics items (especially in measurement, geometry, and applied numeracy) were too difficult, exposing pupils' weaknesses in core numeracy skills and raising concerns about the quality and validity of the national achievement test.

Gender-Based Performance Differences

While minor, the gender differences with females performing better in language-based subjects and males slightly better in numeracy reflect underlying socio-cultural dynamics and gendered perceptions of subject preferences. It calls for a more gender-sensitive approach to test construction and teaching methods that avoid reinforcing stereotypes and provide equitable learning opportunities (UNESCO, 2021).

Misalignment with Competency-Based Curriculum

An analysis of the national curriculum objectives shows a strong focus on developing competencies rather than simply recalling knowledge. However, the prevalence of easy items, especially in high-scoring disciplines, suggests an overemphasis on lower-order cognitive tasks. This directly aligns with research conducted by Wirngo (2021), identifying

several issues, including a predominance of easy items with low discriminatory power, which may not effectively assess the competencies outlined in the CBA framework. She opines that the tests often fail to measure higher-order thinking skills and problem-solving abilities, which are central to the CBA and recommends that test constructors review and revise items to better align with the cognitive demands of the curriculum and the competencies expected of learners.

If assessments continue to prioritize simple recall or recognition tasks, teaching and learning will be oriented towards exam success rather than meaningful competence development. To align assessments with the CBA, it is essential to develop test items that challenge students to apply their learning in complex, real-life contexts, thereby ensuring that assessments truly reflect the competencies intended by the curriculum. This will ensure the quality of basic education and the alignment of national goals with SDG 4 - Quality Education.

Conclusion

This study's analysis of the 2023 CEP and FSLC exams reveals that while many subjects were deemed "easy," a few, particularly FSLC Mathematics 1, presented significant challenges. These results underscore the importance of using item difficulty indices as objective indicators for improving test construction (Lupian, 2022; Crocker & Algina, 2006). The data suggests a mismatch between curriculum expectations and learner outcomes in certain subjects, necessitating a systematic revision of exam design, validation processes, and teacher training in assessment literacy.

Recommendations include revising test construction processes, integrating balanced difficulty levels, pre-testing items, and training educators in assessment design. Addressing these gaps will enhance the alignment between curriculum, instruction, and evaluation supporting the national goal of delivering competency-based, quality education (SDG 4).

By ensuring test items are appropriately difficult, educational authorities can foster fairer and more meaningful assessments. This research contributes to the discourse on educational quality in Cameroon by demonstrating how psychometric insights can improve the validity, reliability, and fairness of national assessments. Future studies should expand this analysis to other regions and years and examine how socio-demographic factors influence test performance, thus strengthening the link between test data, classroom instruction, and curriculum development to advance the goals of quality education for all, as envisioned in SDG 4.

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

References

Afang, E. A., & Ekpoh, U. I. (2020). Evaluation of item difficulty and discrimination indices of multiple-choice questions in educational measurement. *International Journal of Educational Research, 18*(2), 112–124.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Prentice Hall.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*(3), 347–364. https://doi.org/10.1007/BF00138871

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain.* David McKay.

Cameroon Ministry of Basic Education (MINEDUB). (2023). *Statistiques des résultats CEP et FSLC 2023 par sous-discipline dans la région du Centre*. Yaoundé: Department of Exams and Certification.

Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Cengage Learning.

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Prentice Hall.

Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). McGraw-Hill.

Gulliksen, H. (1950). *Theory of mental tests*. Wiley.

Lupian, V. (2022). Item analysis: Evaluating the difficulty and discrimination indices in test construction. *Journal of Educational Measurement & Evaluation, 30*(1), 55–67.

Nitko, A. J., & Brookhart, S. M. (2014). *Educational assessment of students* (7th ed.) Pearson.

Omo-Egbekuse, G. (2017). Analysis of examination questions using item response theory in West African basic schools. *International Journal of Educational Testing, 17*(4), 310–328. https://doi.org/10.1080/15305058.2017.1281122

Popham, W. J. (2008). *Classroom assessment: What teachers need to know* (5th ed.). Allyn & Bacon.

Programme d'Analyse des Systèmes Éducatifs de la CONFEMEN (PASEC). (2014 & 2019). *PASEC Assessment Reports: Cameroon country results.* CONFEMEN.

Suen, H. K., & Ary, D. (1989). *Analyzing quantitative behavioral observation data*. Psychology Press.

UNESCO. (2021). *Global education monitoring report 2021: Gender equality in education.* Paris: UNESCO Publishing.

Wirngo Tani, E. (2021). Test quality and students' performances: An appraisal of the national achievement test in English and mathematics for Cameroon primary schools. *European Journal of Education Studies, 8*(8), 299–313. https://doi.org/10.46827/ejes.v8i8.3861

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]