Common item-writing flaws and psychometric properties of multiple-choice test in technical schools in Ghana

Yayra Dzakadzie

Department of Educational Foundations, University of Education, Winneba, Ghana.
Email: ydzakadzie@uew.edu.gh

ABSTRACT

High-quality classroom assessment is essential for effective teaching and learning, particularly in technical and vocational education, where assessment outcomes guide instructional decisions and skills development. This study investigated the common item-writing flaws and psychometric quality of multiple-choice tests (MCTs) constructed by Technical Drawing teachers in public technical schools in the Central Region of Ghana. A descriptive cross-sectional survey design was employed, involving all twelve Technical Drawing teachers using a census approach. Data were obtained through document analysis of teacher-developed MCTs and students' response data for psychometric evaluation. Item difficulty indices, discrimination indices, and distractor efficiency were analysed using SPSS and JMetrik software. The findings revealed recurring item-writing flaws, including identifiable answer-key patterns, ambiguous wording, unclear directions, and fairness concerns related to contextual references and linguistic complexity. Psychometric analysis showed an imbalance in item difficulty levels and a high proportion of items with poor or negative discrimination indices, indicating limited effectiveness in differentiating between high-and low-performing students. Although a substantial proportion of distractors were functional, several were either non-functional or misleading, thereby weakening the diagnostic value of the items. The study highlights gaps in teachers' application of test-construction principles and underscores the need to strengthen assessment literacy. Targeted professional development, routine item analysis, and institutional assessment quality assurance to enhance the validity, reliability, and fairness of classroom-based multiple-choice assessments in technical education.

Keywords: *Assessment quality, Classroom assessment, Ghana, Multiple-choice tests, Technical and vocational education, Test construction.*

Highlights of this paper

- The study identifies common item-writing flaws in multiple-choice tests constructed by Technical Drawing teachers in technical schools in Ghana.
- Psychometric analysis revealed imbalanced item difficulty levels and a high proportion of items with poor or negative discrimination indices.
- Strengthening teachers' assessment literacy through professional development and routine item analysis is essential for improving classroom assessment quality.

1. INTRODUCTION

Assessing students' learning is an essential component of effective teaching and learning, as instructional decisions are heavily dependent on the highly valid and quality information obtained from classroom assessment (Crocker & Algina, 2008). When assessment practices are well planned and properly implemented, they provide valuable feedback that serves as a scaffold for supporting learning activities, guides instructional management decisions and planning and fosters students' understanding of subject matter (Nitko, 2001). By contrast, poorly developed assessment has the potential of misrepresenting students' achievement, undermines motivation and leads to ineffective instruction management decisions (Agu, 2013). These negative outcomes are detrimental to the teaching and learning enterprise, thereby undermining educational success and systems.

Teachers at all levels of education play a critical role in the implementation of classroom assessment activities, particularly engaging in the construction of achievement tests, which serve both formative and summative purposes (Oppong, Nugba, Asamoah, Quansah, & Ankoma-Sey, 2023). It is therefore important that teachers possess the requisite knowledge and skills in test development in order to achieve the expected learning outcomes. In fact, this required competence is therefore an essential professional skill, requiring teachers to align assessment tasks with instructional objectives and apply sound measurement principles (Fives & DiDonato-Barnes, 2013). It is in this light that several stakeholders and international bodies have demonstrated commitment to enhancing teachers' competence and skills in developing and administering assessment tasks.

Teachers, in their efforts to develop high-quality assessments, use a wide range of assessment formats, from objective test type to essay formats (Frey & Schmitt, 2007; Quansah, 2018). However, the multiple-choice test format, a type of objective test, has been found to be widespread for certain reasons. A paramount reason for the widespread use of the multiple-choice test at local, national and international assessment contexts is its strong alignment with item and psychometric analyses (i.e., the functionality properties of the items), and its suitability in measuring broad cognitive outcomes (Baker, 2001; Nitko, 2001). Despite this strength, the effectiveness of the multiple-choice items largely depends on the quality of the item writing, including stem clarity, fairness, distractor plausibility, and appropriate difficulty and discrimination levels (Haladyna & Rodriguez, 2013).

Notwithstanding the widespread use of multiple-choice tests, empirical studies have consistently shown that several teachers struggle to develop high-quality, reliable and valid multiple-choice assessment tasks (Attali & Bar-Hillel, 2003; Kissi, Baidoo-Anu, Anane, & Annan-Brew, 2023; Quaigrain, 1992; Quansah, Amoako, & Ankomah, 2019; Tarrant, Ware, & Mohammed, 2009). These challenges manifest as difficulties in ensuring content representativeness, selecting functional distractors, avoiding ambiguity, and interpreting assessment data effectively. Such limitations highlight the importance of assessment literacy, defined as teachers' ability to design, administer, analyse, and use assessment information appropriately to support learning (Mertler, 2016).

In Ghana, classroom assessment is conducted by both trained and untrained teachers across all levels of education, often regardless of whether they have received adequate formal preparation in educational measurement and evaluation (Gabuyo, 2012). Although teacher education institutions offer courses in assessment and test construction (Quansah & Ankoma-Sey, 2020) empirical evidence suggests that that teachers' practical application of

these skills remains inconsistent. While some studies report relatively high levels of competence in test construction (Ankomah, 2020; Oduro-Kyireh, 2008) others reveal persistent weaknesses, including structural flaws, poor content coverage, and limited adherence to item-writing principles (Kissi et al., 2023; Quansah et al., 2019). These mixed findings suggest that teachers' assessment competence is uneven and context-dependent.

Document analysis studies conducted in Ghana have further shown that teachers often reuse test items without adequate modification, introduce identifiable answer-key patterns, and include ambiguous or curriculum-misaligned questions, all of which compromise test validity (Kissi et al., 2023; Quansah et al., 2019). These weaknesses are particularly concerning in technical and vocational education, where assessment outcomes inform both academic progression and skills development.

Although earlier studies have provided useful insights into teachers' assessment practices, many have relied predominantly on self-reported questionnaires or surface-level document analysis, offering limited empirical evidence on the psychometric quality of classroom test items (Ankomah, 2020; Kissi et al., 2023; Oduro-Kyireh, 2008). Evaluating teachers' test-construction competence through questionnaires or documentary review may yield overly positive impressions of their assessment skills or replicate standardised test formats without genuinely understanding the underlying principles of test design. Moreover, these studies often failed to explore teachers' deeper perceptions of content representativeness, an essential element in ensuring content validity and constructing quality test items. A critical analysis of the literature, thus, reveals that the predominant reliance on quantitative and document analysis methods has not adequately captured teachers' detailed perspectives on test construction.

In a more recent and comprehensive research, Quansah, Quansah, and Dzakadzie (2025) adopted a triangulation research methodology to evaluate the multiple-choice test development competencies of mathematics teachers in Ghana. The authors combined evidence from a survey, documentary review and psychometric item analysis. They reported generally low multiple-choice test scores in the mathematics teachers' competencies. However, the survey showed modest competencies in test item assemblying, content validity, and test option handling. Despite the significant contribution of this study to understanding teachers' test construction competencies, its findings are contextually limited to mathematics instruction. Mathematics and Technical Drawing differ substantially in their curricular orientation, cognitive demands, instructional practices, and assessment traditions. While Mathematics assessment often emphasises symbolic manipulation, procedural fluency, and abstract reasoning, Technical Drawing requires integrating spatial visualisation, graphical interpretation, and applied technical conventions. These differences shape teachers' professional training, classroom practices, and approaches to test construction, suggesting that assessment competencies observed among Mathematics teachers cannot be assumed to generalise to Technical Drawing teachers. Consequently, an empirical investigation focused specifically on Technical Drawing teachers is necessary to capture the unique assessment challenges and item-writing demands inherent in technical education.

More importantly, relatively little attention has been given to technical schools and to subject areas such as Technical Drawing. It should be clarified that assessment outcomes in technical education carry high instructional and vocational stakes, inform students' progression into skilled trades, and influence the credibility of competency-based certificate systems. In this context, poorly constructed test items may not only misrepresent students' conceptual understanding but also distort decisions about readiness for practical training, thereby undermining the alignment between theoretical assessment and skill acquisition. Examining teachers' test-construction competencies in technical schools is therefore essential to ensuring that multiple-choice assessments validly, reliably, and fairly measure the specialised cognitive demands of technical subjects and support meaningful learning and workforce preparation.

Against this backdrop, the present research examines common item-writing flaws and the overall quality of multiple-choice test items developed by Technical Drawing teachers in public technical schools in the Central Region of Ghana. Specifically, the study seeks to: (1) identify common item-writing flaws in the multiple-choice tests constructed by technical school teachers, and (2) evaluate the psychometric quality of these tests using indices of item difficulty discrimination and distractor efficiency. Through the integration of documentary review and analysis with empirical item analysis, this research contributes evidence-based insights into classroom assessment quality. It offers practical implications for improving test construction practices in technical education.

1.1. Theoretical Perspectives on Multiple-Choice Test Construction

Classical Test Theory (CTT) served as the primary theoretical foundation for evaluating the quality of classroom-based multiple-choice tests. According to the CTT, an observed test score is composed of a true score and measurement error, and the quality of test items is reflected in their ability to minimise errors while accurately representing the construct being measured (Crocker & Algina, 2008). Within this measurement framework, item-level statistics such as item difficulty indices, discrimination coefficients, and distractor functioning are widely used to evaluate test quality, particularly in classroom and large-scale assessments (Ebel & Frisbie, 1991).

Item difficulty represents the proportion of examinees who answer an item correctly and indicates whether an item is appropriately challenging for the target group (Mahjabeen et al., 2017). While a range of difficulty levels is desirable in a test, extremely easy or extremely difficult items contribute little to score interpretation. Item discrimination refers to the degree to which an item differentiates between high-and low-performing examinees. Items with high discrimination indices are considered effective because they are more likely to be answered correctly by students with high overall ability (Mahjabeen et al., 2017). Negative or near-zero discrimination indices, by contrast, signal potential flaws such as ambiguous wording, incorrect keys, or misalignment with instructional content. Distractor efficiency is described as the extent to which incorrect options attract examinees with lower ability levels while being avoided by higher-performing students (Crocker & Algina, 2008). Non-functional distractors that are never selected or misleading distractors that attract high-performing students weaken item quality by reducing discrimination and increasing measurement error. Notably, these three indicators provide a comprehensive and empirically grounded basis for evaluating the quality of teacher-developed multiple-choice tests.

From a validity perspective, sound item construction is essential to ensure that test scores meaningfully reflect students' knowledge and skills rather than irrelevant factors such as guessing, ambiguity or test-wiseness (Messick, 1995). Poorly constructed items introduce construct-irrelevant variance, thereby undermining both the interpretability and fairness of assessment outcomes. These theoretical perspectives highlight the importance of evaluating teachers' competence in item writing as a prerequisite for valid and reliable classroom assessment. As discussed earlier in the introduction, extensive literature has confirmed persistent, limited test construction competencies among teachers (Kissi et al., 2023; Quansah et al., 2019; Quansah et al., 2025). The literature commonly classifies these flaws into item-level writing (e.g., irrelevant difficulty, ambiguity), option-level (e.g., non-functional or implausible distractors) and test-structural flaws (e.g., inconsistent formatting, identifiable answer patterns).

In classroom-based assessments, additional structural flaws often observed include poor alignment with instructional objectives, predictable answer patterns, and the reuse of items without appropriate review. (Quansah et al., 2019). These flaws not only reduce the psychometric quality of test items but also enable test-wise examinees

38

to exploit patterns rather than demonstrate actual understanding. Consolidating these taxonomies provides a coherent lens for identifying common weaknesses in teachers' item writing.

Assessment in technical and vocational education presents distinctive challenges that may exacerbate item-writing difficulties. As already stated, technical subjects often require integrating conceptual knowledge, spatial reasoning, and applied understanding, making them particularly sensitive to poorly worded or misaligned test items. (Yusop, Rasul, Mohammad Yasin, & Hashim, 2023). In subjects like Technical Drawing, multiple-choice tests are commonly used to assess abstract concepts such as projection method, line conventions, and geometrical constructions, which demand precision in language and visual interpretation. Furthermore, disparities in instructional resources across technical schools may introduce fairness concerns when test items reference tools, equipment, or contexts not uniformly available to all students (Yusop et al., 2023).

2. METHODS AND MATERIALS

2.1. Research Design

The study adopts a descriptive, cross-sectional survey design. The choice of this design is grounded in the fact that the researcher sought to understand the current state of test construction competencies by collecting from the Technical Drawing teachers at a single point in time to describe existing conditions, practices, and challenges (Fraenkel, Wallen, & Hyun, 2012). Moreover, this design enabled an in-depth description of teachers' test-construction practices without manipulating variables, while also identifying patterns and deficiencies in item development.

2.2. Study Context

The study was conducted in the Central Region of Ghana. The metropolis contains a concentration of public technical and vocational institutions that offer both core and technical subjects, including Mathematics, English Language, Integrated Science, and Technical Drawing. Technical Drawing was selected as the focus subject for this study because it is a general technical subject, theoretical in nature, and commonly assessed through multiple-choice tests. The subject demands conceptual understanding of drawing principles, geometric constructions, and visualisation skills, all of which can be effectively evaluated through well-constructed multiple-choice items. The choice of the metropolis was based on its diversity of technical institutions and accessibility for data collection.

2.3. Participants

The target population for the study consisted of all 12 Technical Drawing teachers at four public technical schools in the Central Region: Asuansi Technical Institute, Ajumako Afranse Technical Institute, Brakwa Senior High Technical School, and Cape Coast Technical Institute. These teachers were selected because they are directly responsible for preparing and administering classroom tests that often include multiple-choice items as part of continuous assessment or end-of-term examinations. Since the number of Technical Drawing teachers in the selected schools was relatively small and manageable, the census approach was employed. Thus, all the 12 teachers who teach Technical Drawing in the identified schools participated in the study. This approach was deemed appropriate because it eliminates sampling error and ensures that the data represent the entire population of interest. The 12 teachers were put into two groups to develop 20 multiple-choice questions each.

After approval from appropriate gatekeepers, the researcher contacted the heads of the selected technical schools to agree on appropriate dates for data collection. On the scheduled days, the researcher and three trained assistants visited each school to collect copies of teachers' multiple-choice tests and administer the student test

instruments. All participants were informed about the purpose of the study, and confidentiality and voluntary participation were ensured throughout the process.

2.4. Data Collection Instruments

Two main instruments were employed for data collection: documentary analysis and a test-based instrument. First, the study utilised qualitative document analysis to review and evaluate the multiple-choice test items developed by Technical Drawing teachers. Document analysis involves systematically examining written or printed materials to obtain empirical information (Creswell & Creswell, 2014). The analysis in this study focused on identifying common item-writing flaws in test development, including content representativeness, ambiguity, item structure, and fairness. The multiple-choice tests constructed by the participating teachers were also administered to a group of students to generate response data for psychometric analysis. This approach provided quantitative evidence on the performance of individual test items and allowed the researcher to assess their reliability, discrimination, and difficulty levels.

2.5. Data Analysis

Data collected were coded and entered using the Statistical Package for the Social Sciences (SPSS) version 25.0. The document analysis focused on identifying and categorising the flaws in multiple-choice items across four dimensions: Identifiable patterns, Ambiguity, Unclear directions, and Fairness to different students. In addition, students' test response data were analysed using JMetrik software to examine the psychometric characteristics of the items, including item difficulty, discrimination index, and reliability coefficients. The use of both qualitative document analysis and quantitative item analysis provided a comprehensive understanding of the common flaws in multiple-choice test construction among Technical Drawing teachers.

3. RESULTS

3.1. Common Item-Writing Flaw in MCTs Developed by the Technical School Teachers

The first objective of this study was to identify the common flaws in multiple-choice test (MCT) items developed by Technical Drawing teachers in selected technical schools within the Central Region. To achieve this, a qualitative document analysis was conducted on teachers' end-of-term tests. The analysis focused on four key dimensions of item-writing quality: identifiable patterns, ambiguity, unclear directions, and fairness to different students. The findings revealed several recurring weaknesses that compromise the overall quality, validity, and fairness of the teachers' test items.

3.1.1. Identifiable Patterns in Answer Keys

The document analysis revealed that a considerable number of the test items exhibited identifiable patterns in the arrangement of correct options. In several test papers, the correct answers followed predictable sequences such as "A–A–B–C–C" or "D–D–B–B–A." Such predictable key patterns make it easier for test-wise students to guess answers based on perceived trends rather than actual knowledge. Moreover, the analysis showed that some teachers consistently preferred certain option positions (mostly "B" or "C") as the correct answer. This bias could stem from the teacher's unawareness of the importance of randomisation or from a lack of peer review during test preparation.

3.1.2. Ambiguity in Item Stems and Options

Another recurring flaw identified was ambiguity in the wording of both item stems and response options. Ambiguity occurred when questions were poorly phrased, contained double meanings, or included terms that were not clearly defined. For example, in several Technical Drawing questions, stems used vague expressions such as "the best method of projection" or "the most appropriate line type" without specifying the context or drawing standard being referenced. Similarly, some options were semantically overlapping, making it difficult to distinguish between correct and incorrect answers. For instance, two options might both be partially correct depending on the interpretation of the question.

3.1.3. Unclear Directions

The analysis further revealed that many tests lacked clear instructions guiding students on how to respond to the items. Some papers did not explicitly state the number of questions to be answered, the marking scheme, or whether multiple responses were allowed for compound items. In certain tests, inconsistent item and option numbering caused confusion. For example, question numbers were sometimes skipped or repeated (e.g., two items labelled "Q15"), and some options were mislabeled (e.g., two options marked "C"). Such presentation flaws not only confuse students but also hinder accurate scoring and interpretation of results.

3.1.4. Fairness to Different Students

The fourth dimension examined was fairness to different students, defined as the extent to which all students, regardless of background, have an equal opportunity to demonstrate their knowledge and skills. The study found several items that were unfair or biased toward certain student groups. For instance, some test items used examples, tools, or scenarios unfamiliar to all examinees, such as questions about specialised industrial machines or drawing instruments not available in all schools. This disadvantage particularly affected students from resource-constrained institutions, highlighting issues of contextual bias.

Additionally, certain items contained complex grammatical constructions that disadvantaged linguistically weaker students, even when they understood the underlying Technical Drawing concepts. In some instances, items required knowledge of topics outside the approved Technical Drawing syllabus, suggesting curricular misalignment.

3.2. Quality of Multiple-Choice Tests (MCTs) Developed by Technical School Teachers

This section presents the results of the analysis of multiple-choice test (MCT) items developed by Technical Drawing teachers in selected technical schools in the Central Region. The selected test covered content areas including orthographic projection, geometrical construction, and instrumental drawing: key theoretical domains in Technical Drawing typically assessed through multiple-choice tests. Table 1 presents the results of the item difficulty and discrimination analysis. For interpretation, item difficulty indices below 0.30 were considered as difficult, those between 0.30 and 0.70 were regarded as moderately difficult, and items above 0.70 were classified as easy (Crocker & Algina, 2008). Item discrimination was evaluated using item-total correlation coefficients, with values below 0.10 considered poor, between 0.10 and 0.29 acceptable, and 0.30 and above excellent. Negative discrimination indices were interpreted as indicators of serious item flaws (Crocker & Algina, 2008).

41

Table 1. Item difficulty and discrimination indicators.

Item No.	Difficulty (p)	Difficulty Level	Discrimination (r)	Discrimination Quality
Q1	0.19	Difficult	0.03	Poor
Q2	0.33	Moderate	0.19	Acceptable
Q3	0.54	Moderate	-0.09	Negative
Q4	0.83	Easy	0.14	Acceptable
Q5	0.77	Easy	0.43	Excellent
Q6	0.81	Easy	0.51	Excellent
Q7	0.52	Moderate	0.43	Excellent
Q8	0.27	Difficult	-0.28	Negative
Q9	0.29	Difficult	-0.01	Negative
Q10	0.92	Easy	0.23	Acceptable
Q11	0.56	Moderate	0.15	Acceptable
Q12	0.10	Difficult	-0.17	Negative
Q13	0.92	Easy	0.19	Acceptable
Q14	0.83	Easy	0.20	Acceptable
Q15	0.31	Moderate	0.13	Acceptable
Q16	0.02	Difficult	-0.27	Negative
Q17	0.25	Difficult	-0.20	Negative
Q18	0.67	Moderate	0.28	Acceptable
Q19	0.06	Difficult	0.02	Poor
Q20	0.27	Difficult	-0.03	Negative
Q21	0.40	Moderate	0.25	Acceptable
Q22	0.50	Moderate	0.58	Excellent
Q23	0.87	Easy	0.24	Acceptable
Q24	0.46	Moderate	-0.28	Negative
Q25	0.02	Difficult	-0.27	Negative
Q26	0.75	Easy	-0.06	Negative
Q27	0.12	Difficult	-0.30	Negative
Q28	0.56	Moderate	0.20	Acceptable
Q29	0.19	Difficult	-0.50	Negative
Q30	0.06	Difficult	-0.17	Negative
Q31	0.44	Moderate	0.20	Acceptable
Q32	0.15	Difficult	-0.24	Negative
Q33	0.37	Moderate	0.21	Acceptable
Q34	0.13	Difficult	-0.42	Negative
Q35	0.85	Easy	0.12	Acceptable
Q36	0.87	Easy	0.24	Acceptable
Q37	0.67	Moderate	0.06	Poor
Q38	0.63	Moderate	0.05	Poor
Q39	0.77	Easy	0.38	Excellent
Q40	0.94	Easy	0.15	Acceptable

3.2.1. Item Difficulty and Discrimination

The results show that 14 items (e.g., Q1, Q8, Q9, Q16, Q17, Q19, Q20, Q25, Q27, Q29, Q30, Q32, Q34) were classified as difficult, with difficulty indices below 0.30 (see Table 1). Another 14 items fell within the moderate range (0.30–0.70), while 12 items were easy (above 0.70). This distribution suggests that the test lacked a balanced spread of item difficulty, as a well-constructed test should contain a mix of easy, moderate, and difficult items. In terms of discrimination, 15 items (e.g., Q3, Q8, Q9, Q12, Q16, Q17, Q20, Q24, Q25, Q26, Q27, Q29, Q30, Q32, Q34) exhibited negative discrimination indices, indicating potential flaws where low-performing students answered correctly more often than high-performing students. Conversely, 16 items showed acceptable discrimination (0.10–0.29), while only 5 items (Q5, Q6, Q7, Q22, Q39) achieved excellent discrimination (≥0.30). This pattern reflects

weaknesses in test validity and item writing, as many items failed to differentiate between high- and low-ability learners.

3.2.2. Distractor Analysis

Distractor functionality was evaluated using option-total corrections. Distractors with negative correlations were considered functional, as they appropriately attracted lower-performing examinees. Distractors with zero coefficients were classified as non-functional, indicating that no examinees selected the option. Distractors with positive correlations were regarded as misleading, as they disproportionately attracted higher-performing students, suggesting ambiguity or poor option construction (Haladyna & Rodriguez, 2013). Accordingly, items containing one or more non-functional or misleading distractors were considered psychometrically weak, as ineffective distractors reduce item discrimination and increase measurement error. Distractor efficiency was assessed using option-total correlations, summarised in Table 2.

Table 2. Outcome of distractor efficiency analysis.

Indicator	Observation	Interpretation
Functional distractors	13 items (e.g., Q2, Q4, Q5, Q6, Q7, Q11, Q13, Q14, Q20, Q21, Q22, Q31, Q33)	Negative coefficients indicated that distractors successfully attracted low scorers.
Non-functional distractors (No selection)	6 items (e.g., Q1D, Q3D, Q10D, Q23D, Q28D, Q35D)	Distractor coefficient of 0.00 − no student selected the option, reducing item quality.
Misleading distractors (positive correlation)	7 items (e.g., Q3B, Q16B, Q17B, Q18B, Q19B, Q24B, Q37B)	Positive coefficients suggest that high-scoring students were distracted by incorrect options.

The distractor analysis revealed that approximately one-third of the distractors were non-functional, meaning that some options were never chosen by any examinees (see Table 2). Such distractors fail to contribute to the diagnostic quality of the item. Furthermore, several distractors showed positive correlations, suggesting that high-performing students selected incorrect options, indicating ambiguous or misleading item phrasing. These findings demonstrate weaknesses in item construction and review, where distractors may have been poorly worded, implausible, or inconsistent with the stem.

4. DISCUSSION

The study revealed that approximately 75% of the distractors were functional, which aligns with previous research emphasising the importance of well-constructed distractors in improving test quality. For instance, Tarrant et al. (2009) found that approximately 70% of distractors in their analysis of multiple-choice questions were functional, reinforcing the idea that well-developed distractors enhance the discriminative power of test items. Similarly, Attali and Bar-Hillel (2003) reported that over 80% of distractors on standardised tests were functional due to rigorous item-review processes. The relatively high proportion of functional distractors in the present study suggests that Technical Drawing teachers, despite limited formal training in test development, can construct plausible distractors that engage examinees effectively.

However, the presence of 25% non-functional distractors, those that were never selected or had zero coefficients, still indicates areas for improvement. Such distractors fail to contribute to the item's difficulty or discrimination, as argued by Nitko (2001), who emphasised that non-functional distractors weaken test validity by reducing the range of plausible options. This pattern is consistent with Rodriguez (2005), who found that in

43

classroom-based assessments, only about half of the distractors functioned effectively, largely due to limited item-writing expertise and insufficient item review.

Interestingly, despite most distractors being functional, several items in this study still showed poor discrimination indices. This apparent contradiction can occur under specific conditions related to item design, alignment, and examinee characteristics. Ambiguous or unclear keys, for example, can confuse both high- and low-performing students, leading to low discrimination even when distractors function as intended (Kissi et al., 2023). Similarly, items that are misaligned with instructional objectives may fail to reward actual mastery, resulting in random response patterns (Ebel & Frisbie, 1991). Another factor that may explain the weak discrimination observed is guessing behaviour, which is common in multiple-choice testing. When students guess, response patterns become less correlated with ability level, thereby reducing the discriminatory power of items (Quansah et al., 2025). This issue is particularly likely in Technical Drawing, where abstract spatial reasoning and visualisation skills may make some questions more susceptible to guessing.

In summary, while most distractors were functional, an encouraging sign of improving item-writing skills, issues of ambiguous keys, weak alignment with objectives, and guessing behaviours contributed to poor item discrimination. Addressing these weaknesses through targeted teacher training in test construction and item analysis would enhance the overall validity, reliability, and fairness of multiple-choice tests used in Ghanaian technical schools.

4.1. Implications for Policy and Practice

The findings of this study have important implications for both classroom practice and educational policy, particularly in the context of technical education in Ghana. The study revealed several item-writing flaws (such as identifiable answer patterns, ambiguity, unclear directions, and fairness concerns) that undermine the validity and reliability of classroom assessments. To address these issues, technical school teachers must strengthen their assessment literacy and item construction skills. Teachers should receive ongoing professional development on the principles of test construction, particularly in developing multiple-choice items that are clear, balanced, and unbiased. Training workshops should emphasise avoiding predictable patterns, ensuring clarity of language, and creating functional distractors that differentiate between high- and low-achieving students.

Teachers should be encouraged to conduct routine item analysis after each examination using basic software (e.g., JMetrik or SPSS). Such analyses would help identify non-functional distractors, ambiguous stems, or misaligned items, enabling teachers to refine their tests before reuse. Establishing departmental or inter-school assessment review committees can improve item quality through peer evaluation. Collaborative test construction encourages sharing of expertise and helps reduce personal bias and design errors common in single-author test items. Teachers should adopt the table of specifications (TOS) tools when developing tests. This ensures adequate coverage of content and a proportional representation of key topics, promoting fairness and validity in assessments.

At the policy level, the findings point to systemic issues in teacher preparation and professional support related to educational measurement and evaluation. The Ghana Education Service (GES) and the Commission for Technical and Vocational Education and Training (CTVET) should ensure that teacher training institutions incorporate compulsory, practical courses on educational assessment and psychometrics. This will equip teachers with the competencies to design valid, reliable, and fair classroom tests.

Technical schools should develop school-based assessment quality assurance frameworks to guide test construction, moderation, and review processes. Moderation panels at the district or regional level could be tasked with vetting end-of-term tests before they are administered to students. The Ministry of Education should make

assessment-focused CPD programs mandatory for all technical school teachers. Regular refresher courses on test design and evaluation will bridge the gap between policy intent and classroom practice, ensuring consistency in assessment standards across schools. Policymakers should support schools in accessing digital tools and statistical software for test analysis. Such resources will enhance teachers' capacity to perform data-driven evaluation of test quality, supporting evidence-based decision-making.

5. CONCLUSION

The study examined the common flaws and quality of multiple-choice tests (MCTs) developed by Technical Drawing teachers in selected technical schools in the Central Region of Ghana. The findings revealed several weaknesses in test construction, including identifiable answer patterns, ambiguous wording, unclear directions, and fairness issues. These flaws suggest limited application of sound item-writing principles. These findings highlight persistent weaknesses in the construction of multiple-choice tests by Technical Drawing teachers in technical schools. It underscores the need to strengthen teachers' competence in test design and validation through continuous professional development. Ensuring effective assessment practices requires institutional support, quality assurance mechanisms, and policy attention to promote fair, valid, and reliable evaluation of students' learning outcomes.

REFERENCES

Agu, N. (2013). Measuring teachers' competencies in constructing classroom-based tests in Nigerian secondary schools: Need for a test construction skill inventory. *Educational Research Review, 8*(8), 431–439. https://doi.org/10.5897/ERR12.219

Ankomah, F. (2020). *Predictors of adherence to test construction principles: the case of senior high school teachers in Sekondi-Takoradi Metropolis.* Master's Dissertation, Cape Coast, Ghana: University of Cape Coast.

Attali, Y., & Bar-Hillel, M. (2003). Guess where: The position of correct answers in multiple-choice test items as a psychometric variable. *Journal of Educational Measurement, 40*(2), 109–128. https://doi.org/10.1111/j.1745-3984.2003.tb01099.x

Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

Creswell, J. W., & Creswell, J. D. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches.* Thousand Oaks, CA: SAGE Publications.

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory.* Mason, Ohio: Cengage Learning.

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

Fives, H., & DiDonato-Barnes, N. (2013). Classroom test construction: The power of a table of specifications. *Practical Assessment, Research & Evaluation, 18*(3), n3.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed.). New York: McGraw-Hill.

Frey, B. B., & Schmitt, V. L. (2007). Coming to terms with classroom assessment. *Journal of Advanced Academics, 18*(3), 402–423. https://doi.org/10.4219/jaa-2007-495

Gabuyo, Y. A. (2012). *Assessment of learning I. Manila.* Philippines: Rex Book Store. Inc.

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items.* New York: Routledge.

Kissi, P., Baidoo-Anu, D., Anane, E., & Annan-Brew, R. K. (2023). Teachers' test construction competencies in an examination-oriented educational system: Exploring teachers' multiple-choice test construction competence. *Frontiers in Education, 8*, 1154592. https://doi.org/10.3389/feduc.2023.1154592

Mahjabeen, W., Alam, S., Hassan, U., Zafar, T., Butt, R., Konain, S., & Rizvi, M. (2017). Difficulty index, discrimination index, and distractor efficiency in multiple-choice questions. *Annals of PIMS-Shaheed Zulfiqar Ali Bhutto Medical University,* *13*(4), 310–315.

Mertler, C. A. (2016). *Classroom assessment: A practical guide for educators.* New York: Routledge.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741.

Nitko, A. J. (2001). *Educational assessment of students.* New Jersey: Prentice Hall.

Oduro-Kyireh, G. (2008). *Testing practices of senior secondary school teachers in the Ashanti region of Ghana.* Unpublished Master's Thesis, Cape Coast, Ghana: University of Cape Coast.

Oppong, S., Nugba, R. M., Asamoah, E., Quansah, N., & Ankoma-Sey, V. R. (2023). Teachers confidence of classroom assessment practices: A case of basic schools in upper denkyira west district, Ghana. *European Journal of Education Studies, 10*(11), 148–159. http://dx.doi.org/10.46827/ejes.v10i11.5063

Quaigrain, A. K. (1992). *Teacher competence in the use of essay tests: A study of secondary schools in the Western Region of Ghana: unpublished* Master's Thesis, University of Cape Coast, Ghana.

Quansah, F. (2018). Traditional or performance assessment: What is the right way to assessing learners. *Research on Humanities and Social Sciences, 8*(1), 21–24.

Quansah, F., Amoako, I., & Ankomah, F. (2019). Teachers' test construction skills in senior high schools in Ghana: Document analysis. *International Journal of Assessment Tools in Education, 6*(1), 1-8. https://doi.org/10.21449/ijate.481164

Quansah, F., & Ankoma-Sey, V. R. (2020). Evaluation of pre-service education programme in terms of educational assessment. *International Journal of Research in Teacher Education, 11,* 56–69.

Quansah, N., Quansah, F., & Dzakadzie, Y. (2025). Multiple-choice test development competencies of junior high school mathematics teachers in Ghana: A triangulation methodology. *Frontiers in Education, 10,* 1658971. https://doi.org/10.3389/feduc.2025.1658971

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice, 24*(2), 3–13. https://doi.org/10.1111/j.1745-3992.2005.00006.x

Tarrant, M., Ware, J., & Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. *BMC Medical Education, 9*(1), 40. https://doi.org/10.1186/1472-6920-9-40

Yusop, S. R. M., Rasul, M. S., Mohammad Yasin, R., & Hashim, H. U. (2023). Identifying and validating vocational skills domains and indicators in classroom assessment practices in TVET. *Sustainability, 15*(6), 5195. https://doi.org/10.3390/su15065195