IJEIR
International Journal of Educational Innovation and Research
Volume 4, Number 1, 2025, pp. 112-124
DOI: https://doi.org/10.31949/ijeir.v4i1.11953
ISSN: 2810-0808

Educators' Dexterity in Test Construction within Senior High Schools: Implications for Educational Administrators

Rioborue Alexander Oghenerume*
University of Benin, NIgeria
*Corresponding Author: raoghenerume@gmail.com

ABSTRACT
Educators' dexterity in constructing effectual, dependable, and valid tests is indispensable for accurately measuring learners' outcomes and supporting data-driven decisions by educational administrators. This exploration probed educators' test construction dexterity across all disciplines in senior high schools within Edo South Senatorial District, tackling two foremost lacunas in existing explorations: the focus on all disciplines and the implications for educational administrators. Applying survey design and census sampling, data were amassed from 1,596 educators across 129 public senior high schools via the validated Educators' Dexterity in Test Construction Questionnaire (EDTCQ), with a dependability coefficient of .91. The aftermaths uncovered a low dexterity level (n = 1596, M = 57.52, SD = 16.95), below the scale mean of 67.5. Male educators (n = 595, M = 61.65, SD = 16.59) significantly proved superior test construction dexterity than female educators (n = 1001, M = 55.07, SD = 16.69; t(1594) = 7.63, p = .000), and educators with over 10 years of experience (n = 725, M = 69.27, SD = 12.97) exhibited significantly greater dexterity than those with 10 years or less (n = 871, M = 47.75, SD = 13.27; t(1594) = 32.59, p = .000). No significant discrepancy in test construction dexterity between educators with bachelor's degree (n = 1074, M = 57.54, SD = 17.16) and those with master's degree (n = 522, M = 57.50, SD = 16.52; t(1594) = -0.04, p = .969). These outcomes underscored the necessity for educational administrators to tackle inadequacies in educators' assessment dexterity. Interventions should contain mandatory workshops for new educators, advanced training for experienced ones, and sex-responsive professional development. Administrators should initiate mentorship programmes to transfer test design dexterity and prioritise ongoing development in test dependability and validity. Likewise, assessment dexterity audits should be part of performance appraisals to guarantee effectual test construction.

ARTICLE HISTORY
Received 2024-11-20
Accepted 2025-01-25

KEYWORDS
Assessment
Educators' Dexterity
Test Construction
Educational Administrators

INTRODUCTION

In the educational structure, assessment operates a predominant role, functioning as a link between educating and learning outcomes. One primary assessment tool is the test, a mechanism utilised by educators to measure learners' understanding of content, mastery of aptitudes, and progress towards learning intents. Educational administrators, tasked with ensuring instructional quality and accountability, rely on the data derived from these assessments to inform decision-making, allocate resources, and refine educational practices. Consequently, the dexterity of educators to construct effectual tests—tests that are both dependable and valid—becomes a critical factor in the educational process. Given the indispensable function of tests in connecting educational intents with measurable outcomes, educators' dexterity to design

assessments that meet rigorous benchmarks of dependability and validity is essential, forming the foundation for attaining an ideal educational terrain. In an ideal educational terrain, educators possess the indispensable dexterity to create high-quality assessments that are both dependable and valid. These high-quality assessments function as a dual purpose: they allow learners to express their erudition commendably and empower educational administrators to make informed, data-driven decisions that support continuous progress in educational practices. When assessments meet rigorous benchmarks, they augment the educational experience, furthering learners' growth and refining educational excellence across institutions. Failing to attain this ideal, however, can have adverse repercussions. Perhaps, without dependable and valid assessments, educators may struggle to measure learners' feat accurately, limiting the feedback essential for learners' development. Furthermore, educational administrators may lack robust data to evaluate instructional efficacy, impeding efforts to implement improvements that facilitate profound learning outcomes. This inadequacy perpetuates a cycle where learners' learning outcomes remain stagnant and educational excellence is compromised.

The current state of test construction amid educators falls short of this ideal within senior high schools in Edo South Senatorial District. Many educators grapple with ordeals in creating assessments that align with set benchmarks of dependability and validity, resulting in tests that inadequately quantify learners' learning outcomes. This mismatch between educational practices and established benchmarks is indicative of broader systemic issues in test construction. Corroborating the foregoing observation, Agu et al. (2013) expounded that majority of the tests created by educators to measure learners' outcomes in Nigeria are void of dependability and validity, signifying a low level of test construction dexterity amid educators. These ordeals are not isolated, but rather reflect a nationwide trend that impacts educational worth. Further accentuating this ordeal, Alade and Odunsi (2023) reported that there is no clear link between educators' test construction dexterity and the quality of items. This insinuated that educators may not be harnessing the de rigueur aptitudes in test construction, signifying a low level of dexterity that could compromise the dependability of their assessments. This pervasive issue of flawed test construction dexterity not only affects the accuracy of assessments but also poses significant challenges for educational administrators, who rely on dependable data to function.

Educational administrators are experts who superintend the operation of educational institutions, such as high schools, colleges, universities, or educational agencies where they strategically function in shaping educational quality (Hoy & Miskel, 2013; Sergiovanni et al., 2009; Gorton & Alston, 2023). As part of their responsibilities, educational administrators depend on the accuracy and reliability of educator-generated assessment data to assess instructional efficacy, determine curricula adjustments, and identify areas requiring intervention. In this context, quality assessments provide a sound basis for making data-driven decisions, empowering administrators to effectually tackle issues related to learners' outcomes and drive overall school improvement. However, when assessments lack precision, the resulting data can mislead administrators, putatively leading to ineffectual or misguided policy decisions. For example, inaccurate assessment outcomes can lead to misunderstandings about learners' adeptness, limiting the efficacy of remedial efforts or interventions. To abate these issues, educational administrators make sure educators possess a high level of dexterity in test construction and, when vulnerabilities are spotted, offer targeted professional development.

Knowing the indispensable function that educators perform in this process, it becomes evident that their dexterity in test construction directly impacts the quality of evaluating learners. Test construction is not merely about assembling questions; it necessitates erudition of how to create items that align with curricula benchmarks, measure a range of cognitive aptitudes, and offer a fair assessment of learners' aptitudes (Cooper, 2023; Coulacoglou & Saklofske, 2017). For educators, creating an effectual test entails adhering to fundamental principles, such as clearness, neutrality, and germaneness to instructional content. An effectual test should not only measure what learners have learnt but also diagnose areas needing improvement, thus

abetting both educators and learners to adjust instructional tactics accordingly. Moreover, educators face the ordeal of creating tests that are inclusive and unprejudiced, covering a broad spectrum of learning intents while ensuring that the test is free from language or content prejudice that could unfairly impact certain learner groups. When tests are defectively constructed—whether due to unclear wording, limited scope, or lack of alignment with instructional targets—the outcomes may fail to reflect learners' true aptitudes, ultimately compromising the validity and dependability of the assessment data.

Thus, educators' dexterity in test construction encompasses the technical, analytical, and practical expertise necessitated to create tests that are valid, reliable, and pedagogically sound. Dexterity in this context implies a mastery of assessment design principles and the ability to apply these principles consistently and effectually. This skill set includes the adeptness to specify clear learning intents, create a balanced distribution of items that measure a range of cognitive levels, and ensure test items are void of vagueness and prejudice. Predominantly, educators' dexterity in test construction encompasses some indispensable aptitudes. Objectivity is paramount in item writing, as educators must create items without influencing learners' rejoinders or introducing unintended challenges. Additionally, tests should comprehensively cover all significant content areas from the instructional unit, allowing for a balanced assessment that precisely exemplifies each area without prejudice. Effectual test construction also entails validating items to verify that each item assesses the intended knowledge or skills, thereby boosting overall test validity. Beyond assessing rote knowledge, educators are tasked with creating items that challenge learners' analytical, and creative thinking prowess, fostering higher-order cognitive abilities. A blueprint or table of specifications functions as a guide in this process, abetting educators align test items with instructional targets and cover the cognitive skills range set forth in Bloom's Taxonomy (Cooper, 2023; Coulacoglou & Saklofske, 2017; Omorogiuwa, 2019).

Significantly, educators' dexterity in test construction impacts educational outcomes. Educators with high test construction dexterity can more precisely measure learners' progress, providing administrators with dependable data for making informed decisions about instructional strategies, curricular adjustments, and interventions for learners. In addition to benefiting administrators, well-constructed assessments also serve learners directly by offering a fair and accurate measure of their learning, thus promoting self-efficacy and motivation. However, when educators lack the necessary skills in test construction, there can be serious adverse repercussions. For example, tests that emphasise lower-order cognitive skills or include defectively worded items may fail to accurately assess learners' erudition, resulting in unwarranted anxiety, reduced motivation, and potential misjudgments regarding learners' abilities. These issues also affect administrators, who may struggle to make evidence-based decisions when faced with unreliable or invalid test outcomes.

The ineptness of some educators to construct effectual tests can lead to unproductive outcomes, possibly distorting the measurement of learning and leading to misleading insights about both learning outcomes and instructional success (Agu et al., 2013). Educators' ineptitude in test creation has been alluded as a significant contributor to assessment malpractice amid high school educators in Nigeria (Agu et al., 2013). This shortfall may also lead to heightened concern (Oghenerume, 2022; Omorogiuwa, 2019; Adeosun & Mokogwu, 2024) and may contribute to learners' declining feats in both internal and large-scale assessments (Omorogiuwa, 2019; Oghenerume & Uyi-Osaretin, 2024; Oghenerume & Egberha, 2024). Defectively created tests can offer inaccurate assessments of learners, with outcomes that may fail to reflect learners' levels of perspicacity from instruction. One foremost repercussion is that learners who need additional support may not receive it, as these tests do not reliably spot vulnerable learners. Moreover, learners, parents or guardians, and school authorities are often bequeathed misleading data about learners' adeptness, leading to defective verdicts and potentially harmful educational decisions. Faulty test tools created by educators may result in undependable assessments of learners' erudition.

Inadequate assessments by educators can also lead to disappointing outcomes in large-scale assessments. Learners who are perceived as high achievers within the school context, often rated as 'A' learners, may struggle to attain average grades, such as 'C,' in large-scale assessments like WASSCE and NECO SSCE. Explorations by Oghenerume and Uyi-Osaretin (2024), Akanni (2021), Oghenerume and Egberha (2024), and Agu et al. (2013) denoted that flawed test items can impede learners' understanding and rejoinder accurateness, thus altering the validity of inferences relating to their erudition. When educators lack dexterity in test construction, it can lead to learners developing either an underconfident or overinflated self-image, further contributing to poorer outcomes in large-scale assessments.

The observed discrepancy in educators' test construction dexterity based on sex, qualification, and experience insinuates notable dissimilarities in the efficacious dexterity educators divulge when creating assessments, with variations arising across various demographic factors. For instance, in terms of sex, one might observe that male and female educators diverge in their approach to crafting items, potentially resulting in variations in clarity or alignment with curricula targets. Likewise, instances related to qualification may divulge that educators holding a master's degree tend to produce items that better encourage critical thinking, while those with a bachelor's degree may follow distinct patterns in item formulation, likely influenced by their level of educational attainment. Experience is also influential, as educators with over 10 years in the field may adopt a more refined approach to test creation, drawing on years of practice compared to their less experienced peers. These observations across sex, qualification, and experience collectively unveil how educators' backgrounds can shape their dexterity in creating assessments that are meaningful, clear, and fair.

Existing explorations on educators' test construction dexterity have predominantly focused on specific disciplines and localised contexts, leaving critical lacunas in fathoming the broader, multi-disciplinary realities within larger boroughs. For instance, while Adodo (2014), Akanni (2021), and Alade and Odunsi (2023) offered invaluable insights into specific disciplines, their outcomes lacked generalisability across all disciplines. Conversely, Agu et al. (2013) explored all disciplines, but within the constrained scope of Onitsha. Adeosun and Mokogwu (2024) tried to address this by exploring all disciplines within Ovia North East Local Government Area, yet their outcomes remain restrained to that area, leaving the broader applicability within Edo South Senatorial District unanswered.

Moreover, none of the prior explorations explicitly pivoted on CTT, a basic framework for judging the dependability and validity of assessments. The absenteeism of CTT as a rule diminished the potential for their explorations to offer an all-inclusive understanding of how educators' test construction practices align with instituted psychometric benchmarks. This latest exploration bridged these lacunas by probing educators' test construction dexterity across all disciplines within senior high schools in Edo South Senatorial District, proffering an all-inclusive perspective. Furthermore, it uniquely explored the implications of these dexterities for educational administrators, an aspect previously neglected. Through pivoting this exploration on CTT, it foregrounded the germaneness of dependability and validity in assessment design—elements that had been largely overlooked in preceding reports. This tactic not only amplified the scope of existing reports but also offered indispensable acumens for educational administrators, offered pragmatic propositions to refine assessment attributes, and fostered data-driven decision-making.

LITERATURE REVIEW

This latest exploration pivoted on Classical Test Theory (CTT), initially articulated by Spearman (1904) and subsequently expanded by Lord and Novick (1968) as its theoretical framework. At the heart of CTT is the assertion that any observed test score *'X'* comprises both a true score *'T'* and a random error *'E'*, expressed in the formula: $X=T+E$ where *'X'* represents the observed score, *'T'* is the true score, and *'E'* is the error component (Cooper, 2023; Rust et al., 2021; Goldstein et al., 2019; Coaley, 2010; Crocker & Algina, 2008;

Ojerinde et al., 20014). This foundational model is central to CTT's tactic to assessing test dependability and assessing validity, accentuating the distinction between true, stable individual features and measurement error (Cooper, 2023; Rust et al., 2021). In this model, the true score reflects the consistent, underlying adeptness or attribute being measured, while the error accounts for inconsistencies arising from various sources such as test conditions, participant mood, or environmental factors (Coulacoglou & Saklofske, 2017; Coaley, 2010).

CTT is programmed on some central assumptions that simplify its application. First, it infers that the true score is a stable, unchanging construct across repeated test dispensations. Thus, any fluctuation in observed scores is attributed solely to random error rather than fluctuations in the test-taker's underlying adeptness (Cooper, 2023; Spearman, 1904; Rust et al., 2021). This assumption is indispensable for upholding the dependability of test scores over time and across divergent settings (Lord & Novick, 1968; Linn, 2010; Goldstein et al., 2019). The second assumption is that errors are random and apportioned equivalently across all test-takers, insinuating that measurement error should cancel out when amassed across a large, diverse sample of test-takers, which guarantees that discrepancies in observed scores are meaningful reflections of the true score divergences (Lord & Novick, 1968; Coulacoglou & Saklofske, 2017; Crocker & Algina, 2008).

CTT also infers that error is independent of the true score, which is vital for guaranteeing the legitimacy of dependability assessments, as it impedes systematic prejudices from affecting test scores (Coaley, 2010; Cooper, 2023; Rust et al., 2021). If measurement errors were associated with true scores, the dependability assessments would be skewed, possibly leading to inaccurate verdicts about a test-taker's adeptness (Lord & Novick, 1968; Goldstein et al., 2019; Rust et al., 2021). Furthermore, CTT infers homogeneity of error variance across test-takers, which simplifies the computation of dependability constants, for instance, Cronbach's alpha (Coulacoglou & Saklofske, 2017; Cooper, 2023; Lord & Novick, 1968; Cronbach, 1950). This homogeneity inference alludes measurement error is stable across the entire test populace, expediting a single Standard Error of Measurement (SEM) for all partakers, regardless of their true score level (Cooper, 2023; Linn, 2010; Spearman, 1904; Coulacoglou & Saklofske, 2017; Crocker & Algina, 2008).

The germaneness of CTT to this exploration, *"Educators' Dexterity in Test Construction within Senior High Schools: Implications for Educational Administrators,"* is significant because it offered a robust framework for appraising the efficacy and worth of assessments created by educators. Through strategic concepts such as dependability and validity, CTT qualified this exploration to assess how well educators are trained to create tests that accurately reflect learning outcomes. This offered an underpinning for assessing the dexterity of educators in test construction, spotting areas that may necessitate additional support or professional development. Such acumens are imperative for educational administrators intending to refine assessment practices in their institutions. Moreover, CTT approved for an analysis of the foremost aspects that influence educators' test construction dexterity. Through its principles, this exploration tested the influences underscored in the postulations—namely, whether there are significant discrepancies in educators' test construction dexterity by sex, qualification, or experience. Through leveraging CTT, this exploration was positioned to offer actionable acumens that could steer the development of training programmes and resources to refine educators' test construction dexterity. Largely, the utilisation of CTT bolstered this exploration's adeptness to offer realistic propositions that appended to more effectual testing practices, benefiting both educators and learners in high schools.

Agu et al. (2013) did an exploration titled "Measuring Teachers' Competencies in Constructing Classroom-Based Tests in Nigerian Secondary Schools: Need for a Test Construction Skill Inventory." Their exploration aimed to tackle concerns about the item attributes of classroom-based tests often utilised within high schools in Nigeria, especially following the introduction of continuous assessment in 1985. Their exploration sought to create and validate a Test Construction Skill Inventory to assess the dexterity of high

school educators in creating effectual tests. A 30-item tool was created, which underwent factor analysis, resulting in the identification of 25 items deemed factorially valid. This TCSI unveiled a dependability coefficient of 0.73, signifying suitable consistency. The outcomes divulged that high school educators acknowledged nearly all the identified skills as essential for quality test construction, highlighting the magnitude of these dexterities. Regarding the influence of sex, this exploration detected no significant discrepancy in the mean amid female and male educators, with a p-value exceeding .05, insinuating sex did not play a role in dexterities related to test construction. Nonetheless, a significant discrepancy in the mean by experience was noticed, insinuating that more experienced educators proved superiority in test construction dexterity. These researchers endorsed the TCSI as an indispensable tool for evaluating educators' dexterity in test construction within Nigeria, insinuating that it could inform professional development initiatives to boost assessment practices.

Adodo (2014) did a study evaluating high school teachers' competency in assessing learners' cognitive and psychomotor achievements in Basic Science and Technology (BST). This empirical research explored how factors such as educators' sex, experience, and qualification influence their dexterity in evaluation. The exploration opted applied a survey design and focused on high school educators within four LGAs in Ondo State. Through stratified random sampling, Adodo picked 90 educators—57 males and 33 females—from a combination of 5 private and 20 public high schools. A 25-item, self-created tool was utilised to measure educators' dexterity, predominantly in setting test aims, creating tables of specification, and evaluating learners' learning outcomes. Analytics encompassed both descriptive and inferential statistics to handle one query and three postulations. The outcomes denoted that educators' qualification and experience had no significant influence on their dexterity in evaluating learning outcomes, setting aims, or utilising tables of specification. Nonetheless, a significant discrepancy was spotted amid female and male educators, with sex influencing dexterity levels in assessing BST learners' learning outcomes.

Akanni (2021) appraised the test construction dexterity of educators and its effect on Physics achievement within Education District IV of Lagos State. The exploration expended survey, with a populace entailing all educators within the district. 21 educators were opted for, via simple stratified random sampling. The tools engaged were the TSCI and PAT, which in turn, demonstrated reliabilities of 0.82 and 0.79, during pilot testing. Three postulations were posed. Akanni's exploration outcomes proved sex significantly manipulated educators' test construction dexterity. Precisely, a significant discrepancy was spotted amid female and male educators' dexterity in test construction, whereas no significant discrepancy transpired based on qualification. Additionally, no substantial impact of educators' dexterity occurred on Physics' learning outcomes.

Adeosun and Mokogwu (2024) conducted an exploration on teachers' competency in test construction in high schools. These researchers expended a survey design for their exploration, with a populace containing 162 educators from 23 senior high schools in Ovia North East LGA. This exploration opted for the 162 educators as its sample, as the census tactic was utilised to ensure comprehensive depiction of the all-inclusive populace. Data compendium was expedited via Teachers' Competency in Test Construction Questionnaire (TCTCQ), which was validated by three authorities in psychometrics domain. To assess the dependability of the tool, Cronbach alpha method was applied, yielding a coefficient of 0.80, denoting suitable consistency. Mean and standard deviation were applied to address the query, while independent t-test and ANOVA were harnessed to verify the postulations. Their exploration's outcomes proved that educators' dexterity level in test construction is high in Ovia North East LGA. Additionally, this exploration spotted a significant discrepancy in dexterity based on sex, although no significant discrepancies were spotted concerning educators' qualifications and experience.

Query

What is the dexterity level of test construction amid Edo South Senatorial District educators in senior high schools?

Postulations

1. No significant discrepancy exists in educators' test construction dexterity by sex.
2. No significant discrepancy occurs in educators' test construction dexterity by qualification.
   No significant discrepancy transpires in educators' test construction dexterity by experience.

METHODOLOGY

This exploration methodically expended survey research as its design. This design was opted for, due to its adeptness to offer in-depth acumens into the populace's attributes, permitting the researcher to attain an all-encompassing and accurate picture of educators' dexterity in test construction. According to Edo State Ministry of Education (2024), this exploration's populace included 1,596 educators within 129 senior high schools in Edo South Senatorial District. Given this relatively manageable populace size, census sampling tactic was expended to incorporate every educator within this populace. The use of census sampling tactic was chiefly apt as it eradicated sampling error and amplified the dependability of outcomes by offering an all-inclusive dataset. This tactic was indispensable for attaining a thorough understanding of test construction dexterity across the district, guaranteeing outcomes that are more precisely reflective and generalisable. Likewise, this tactic also allowed the researcher to tackle potential variations across diverse educators without the restrictions associated with sample-based estimations. Conversely, it is imperative to note the flaws of applying census sampling tactic. While it offers an all-inclusive dataset from the intact populace, it could be resource-intensive and time-consuming. Besides, the outcomes are constrained to the precise populace explored and may not be applicable to populaces with dissimilar features. Notwithstanding these considerations, the benefits of expending census tactic—such as greater accuracy and the adeptness to generalise outcomes to the precise populace of interest—outweighed the flaws. Ethically, this exploration adhered to strict rules to guarantee the credibility and veracity of the research process. All participation was not involuntary, and informed consent was gotten from every educator entailed in this work. The partakers were fully instructed about this study's intent, their function in this exploration, and their rights, including the proclivity to retract at any time without penalisations. Additionally, stringent confidentiality measures were applied throughout the data compendium process to safeguard the privacy and security of educators' data. These ethical precautions guaranteed that this exploration adhered to high benchmarks of integrity and reverence for educators.

Predominantly, this exploration applied the Educators' Dexterity in Test Construction Questionnaire (EDTCQ) as its core tool for data compendium. This EDTCQ was adapted from previously validated tools created by Agu et al. (2013), Armah (2018), and Adeosun and Mokogwu (2024), with the items from Agu et al. (2013) being the dominant framework during this adaptation. These tools have all been successfully applied in similar explorations on educators' dexterity. This tool was formed into two sections: Section 'A' and Section 'B.' Section 'A' amassed demographic facts on the educators, including sex, experience, and qualification. Section 'B' entailed 27 items precisely created to assess educators' dexterity in harnessing test construction principles. These items were rated on a four-point Modified Likert Scale, with choices ranging from Strongly Agree 'SA' to Strongly Disagree 'SD', allowing for nuanced rejoinders and offering a robust framework for analysing levels of agreement on diverse features of dexterity in test construction. This EDTCQ's validity was meticulously verified via expert judgment by a panel of seven specialists in psychometrics (5) and educational administration (2). Utilising their vast experience, these experts rigorously assessed each item's germaneness, clearness, and inclusiveness in the context of educators' dexterity in test construction. Their criticism, containing proposed amendments to wording and construct alignment, was assiduously incorporated to

boost the tool's clarity and aptness for this exploration's intents. The dependability of the EDTCQ was verified via a pilot test dispensed to 50 educators from schools outside this exploration's populace. The rejoinders from this pilot test were analysed expending Cronbach's alpha to assess internal consistency, resulting in a dependability coefficient of .91. This high coefficient connoted that the tool's items were dependable and sufficiently internally consistent, making it apt for this latest exploration. Data garnered were meticulously scrutinised utilising mean and standard deviation, and the independent samples t-test to address the query and verify the postulations, respectively. To facilitate clarification, predefined score classifications were applied in line with the benchmarks recommended by Omorogiuwa (2016) and Owie (2013): Low, typifying scores far below the mean of the scale; moderate, signifying scores near the scale's mean; and high, expressing scores well above the mean of the scale. Conversely, for the postulations, the independent samples t-test was expended, as sex implied female and male educators, experience typified educators with over 10 years of experience compared to their peers with less experience, and qualifications connoted educators with a master's degree compared to their peers with a bachelor's degree.

RESULTS

Query 1: What is the dexterity level of test construction amid Edo South Senatorial District educators in senior high schools?

Table 1. Analytics of Educators' Dexterity Level of Test Construction within Senior High Schools in Edo South Senatorial District

Variable	*n*	Sum	*M*	*SD*	Scale Mean	Verdict
Educators' Dexterity	1596	91808	57.52	16.95	67.5	Low

Table 1 expounded the analytical data for the educators' dexterity level of test construction within senior high schools in Edo South Senatorial District. With n=1596 educators, the total score was 91808, yielding a mean score of 57.52 ± 16.95. As this mean is evidently below the scale's mean of 67.5, the dexterity level of test construction amid these educators is clarified as low.

Postulation 1: No significant discrepancy exists in educators' test construction dexterity by sex.

Table 2. Educators' Test Construction Dexterity by Sex

Sex	*n*	*M*	*SD*	*DF*	*t*	*p*	Verdict
Male	595	61.65	16.59				
				1594	7.63	.000	Significant
Female	1001	55.07	16.69				

α = .05.

Table 2 outcomes divulged a significant discrepancy, $t(1594) = 7.63$, $p = .000$, with male educators ($M = 61.65$, $SD = 16.59$, $n = 595$) proving superior test construction dexterity than female educators ($M = 55.07$, $SD = 16.69$, $n = 1001$). Given the p-value falls short of the significance level, unproven postulation—expressing no significant discrepancy in educators' test construction dexterity by sex—was refuted. This proved male educators have superior test construction dexterity compared to their female counterparts.

Postulation 2: No significant discrepancy occurs in educators' test construction dexterity by qualification.

Table 3. Educators' Test Construction Dexterity by Qualification

Qualification	*n*	*M*	*SD*	*DF*	*t*	*p*	Verdict
Master's	522	57.50	16.52				
				1594	-.04	.969	Insignificant
Bachelor's	1074	57.54	17.16				

α = .05.

Table 3 outcomes alluded no significant discrepancy, $t(1594) = -.04$, $p = .969$. Educators with master's degree ($M = 57.50$, $SD = 16.52$, $n = 522$) divulged an equivalent dexterity in test construction as educators with bachelor's degree ($M = 57.54$, $SD = 17.16$, $n = 1074$). Given the p-value exceeds the significance level, unconfirmed postulation—averring no significant discrepancy in educators' test construction dexterity by qualification—was endorsed. This confirmed qualification does not significantly impact educators' test construction dexterity.

Postulation 3: No significant discrepancy transpires in educators' test construction dexterity by experience.

Table 4. Educators' Test Construction Dexterity by Experience

Experience	***n***	***M***	***SD***	***DF***	***t***	***p***	**Verdict**
Over 10 Years	725	69.27	12.97				
				1594	32.59	.000	Significant
10 Years & Below	871	47.75	13.27				

α = .05.

Table 4 outcomes verified a significant discrepancy, $t(1594) = 32.59$, $p = .000$. Educators with over 10 years of experience ($M = 69.27$, $SD = 12.97$, $n = 725$) exhibited significantly superior test construction dexterity than those with 10 years or less of experience ($M = 47.75$, $SD = 13.27$, $n = 871$). Given the p-value falls short of the significance level, unverified postulation—declaring no significant discrepancy in educators' test construction dexterity by experience—was repulsed. This proved educators with more than 10 years of experience demonstrated significantly stronger test construction dexterity compared to those with 10 years or less of experience.

DISCUSSION

Query one proved educators' dexterity level of test construction within senior high schools in Edo South Senatorial District is low. This report is in nexus with those of Alade and Odunsi (2023) and Agu et al. (2013) who reported no clear link between educators' test construction dexterity and the quality of items. They also noticed that numerous tests created by educators to measure learners' learning outcomes in Nigeria are void of dependability and validity, alluding a low level of test construction dexterity amid educators. Oppositely, this result contradicts that of Adeosun and Mokogwu (2024) who explored educators' dexterity in test construction within senior high schools, and reported a high level of test construction dexterity amid educators. This contradiction may be ascribed to divergences in the methodologies applied, as Ovia North East LGA, the area explored by Adeosun and Mokogwu (2024) typifies only a subset of the larger district in Edo State assessed in this recent work. This latest result underscored the urgent necessity for deliberate interventions intended for equipping educators with introductory psychometric aptitudes grounded in CTT. Such interventions would refine the dependability and validity of their assessments, guaranteeing they realistically measure learning outcomes. Also, the outcome alluded that educational policies should prioritise hands-on training and incessant professional development to tackle this shortcoming systematically.

Postulation one proved male educators possess superior test construction dexterity compared to their female counterparts. This new discovery is equivalent to those of Akanni (2021), Adodo (2014), and Adesoun and Mokogwu (2024) who all identified significant discrepancy amid female and male educators' dexterity in test construction in assessing learners' outcomes within Lagos, Ondo, and Edo States, correspondingly. Oppositely, this current discovery refutes the outcomes of Agu et al. (2013), who reported no significant discrepancy in the test construction dexterity of female and male educators. This refutation may have stemmed from the methodological advancements in this current exploration, which applied refined tools adapted from those created by Agu et al. (2013), Armah (2018), and Adeosun and Mokogwu (2024). While the framework from Agu et al. (2013) served as the foundation, these tools underwent rigorous refinement

by a panel of seven specialists—five in psychometrics and two in educational administration. Leveraging their vast expertise, these specialists meticulously assessed each item's germaneness, clearness, and inclusivity concerning educators' dexterity in test construction. Their critiques, which contained proposed amendments to wording and construct alignment, were strictly incorporated to intensify the tools' precision and aptness for this exploration's intents. These refinements likely triggered the divergence in outcomes. Additionally, the sex discrepancies observed may reflect broader societal or institutional elements that influence male and female educators inversely, such as variations in professional development opportunities, workload distribution, or potential prejudices within educational terrains.

Postulation two proved qualification does not significantly impact educators' test construction dexterity. This proof associates with explorations by Adodo (2014), Akanni (2021), and Adeosun and Mokogwu (2024), which empirically reported similar verdicts. This insinuated that dexterity in test construction may be more influenced by practical experience or exposure to effectual assessment practices rather than formal qualification alone. This outcome underscored the aftermath of prioritising skill-based training and ceaseless professional development over reliance on formal qualification as a measure of dexterity in test construction. Likewise, it exposed potential lacunas within educator training programmes, which may inadequately address psychometric rules such as those embedded in CTT. For instance, while qualifications are critical for foundational knowledge, they may lack emphasis on the realistic usage of CTT tenets— predominantly those related to dependability and validity. Given this, professional development initiatives should focus on bridging this gap by equipping educators with the dexterity to design assessments that meet rigorous psychometric benchmarks. Moreover, the outcomes insinuate a pressing need for educational policymakers to recalibrate recruitment and career progression frameworks, warranting that skill-based dexterity, rather than formal qualifications alone, are prioritised in evaluating educators' readiness for assessment obligations.

Postulation three established that experience is significantly influential on educators' test construction dexterity, insinuating that educators with more experience are likely to possess better-developed dexterity in creating assessments that accurately measure learning outcomes compared to those with less experience. This evidence is in nexus with that of Agu et al. (2013), who reported a significant discrepancy in test construction dexterity amid the more experienced and the less experienced educators, with the former demonstrating superior dexterity. This outcome underscored the critical function of applied experience in refining educators' adeptness to create dependable and valid assessments, alluding that accumulated years of practice may foster a deeper erudition of test construction rules and their usage in varied instructional contexts. Contrariwise, this evidence opposes the verdicts of Adeosun and Mokogwu (2024), and Adodo (2014), who conveyed no significant discrepancy of educators' dexterity in test construction by experience within senior high schools within Edo and Ondo States, correspondingly. This opposition may have transpired due to the methodological advancements in this latest exploration, which applied refined tools adapted from those created by Agu et al. (2013), Armah (2018), and Adeosun and Mokogwu (2024). While the framework from Agu et al. (2013) served as the underpinning, these tools underwent meticulous adjustment by a panel of seven authorities—two in educational administration and five in psychometrics. Hinging on their vast expertise, these authorities rigorously assessed each item's germaneness, clearness, and inclusivity concerning educators' dexterity in test construction. Their analyses, which contained refined rewording and construct alignment, were strictly integrated to strengthen the tools' precision and aptness for this exploration's intents. Perhaps their refined analyses likely triggered the divergence in outcomes. Furthermore, the nuanced methodological changes, containing applied CTT rules to guarantee the dependability and validity of the refined tools, may have amplified the accuracy of the data amassed.

CONCLUSION AND RECOMMENDATIONS

Given the outcomes of this exploration, it was resolved that educators' dexterity level in test construction within senior high schools in Edo South Senatorial District was low. Significantly, discrepancies were observed in educators' test construction dexterity by sex and experience. Precisely, male educators significantly possessed superior test construction dexterity compared to their female counterparts, while educators with more experience exhibited greater dexterity in test creation. Nonetheless, no significant discrepancy was found in educators' test construction dexterity based on qualification. These outcomes have critical implications for educational administrators, underscoring the necessity to tackle inadequacies in educators' assessment dexterity. However, this exploration was constrained to one borough, and did not account for national discrepancy. As such, the outcomes may not fully exemplify broader conditions across Nigeria. A replica of this exploration in other boroughs of Nigeria is recommended to assess the broader applicability of these discoveries and resolve whether borough discrepancies exist. Notwithstanding these constraints, this exploration rendered significant contributions to the domains of educational assessment and educational administration. In educational assessment, it underscored inadequacies in educators' test construction dexterity and uncovered areas for upgrading, such as tackling discrepancies by sex and experience. In educational administration, it accentuated the necessity for systemic policies to refine educators' dexterity and guarantee equitable assessment practices.

Ergo, this quantitative researcher proposes that, to upgrade educators' test construction dexterity, targeted interventions should include mandatory assessment development workshops for newly recruited educators, advanced training on test construction rules for experienced educators, and sex-responsive professional development to tackle observed discrepancies. With the proclivity of CTT in guaranteeing the dependability and validity of assessments, workshops must also introduce educators to uncomplicated psychometric rules, accentuating CTT's function in refining test construction practices. Administrators should initiate mentorship programmes that pair early-career educators with experienced educators to foster dexterity transfer in test design, integrating CTT rules in the mentoring process. Policy interventions must also prioritise ongoing professional development via school-based workshops fixed on test dependability, validity, item and statistics. Likewise, educational leaders should incorporate assessment dexterity audits into performance appraisals, guaranteeing that educators are held accountable for harnessing effectual test construction routines instituted on CTT framework. Through these policy actions, educational administrators can refine educators' dexterity levels, upgrade the quality of assessments expended to measure learners' outcomes, and foster equitable, data-driven decision-making. By tackling these inadequacies and executing systemic changes, this exploration advanced the discourse in both educational assessment and educational administration, furthering sustainable capacity-building endeavours and refined educational outcomes in Nigeria.

REFERENCES

Adeosun, P. K., & Mogokwu, G. T. (2024). Teachers' competency in test construction in secondary schools in Ovia North-East Local Government Area, Edo State. *Ilorin Journal of Education, 44*(1), 237-250.

Adodo, S. O. (2014). An evaluation of secondary school teachers' competency in evaluating students' cognitive and psychomotor achievement in Basic Science and Technology (BST). *Journal of Emerging Trends in Educational Research and Policy Studies, 5*(3), 48-53.

Agu, N. N., Onyekuba, C., & Anyichie, A. C. (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.

Akanni, O. T. (2021). Assessment of teachers' test construction competencies on achievement in physics among adolescents in Education District IV of Lagos State, Nigeria. *Review of Education, 33*(2), 1-12.

Alade, O. M., & Odunsi, O. N. (2023). Multiple-choice construction competencies and items quality: Evidence from selected senior high school subject teachers in Kosofe Local Government Area, Lagos State, Nigeria. *International Journal of Educational Research, 12*(2), 212-224.

Armah, C. (2018). Test construction and administration practices among lecturers and staff of the examination unit of the University of Cape Coast. *University of Cape Coast Institutional Repository.* https://ir.ucc.edu.gh:8080/xmlui/bitstream/handle/123456789/10333/COMFORT%20ARMAH.pdf?sequence=3&isAllowed=y

Coaley, K. (2010). *An introduction to psychological assessment and psychometrics.* SAGE. https://doi.org/10.4135/9781446221556

Cooper, C. (2023). *An introduction to psychometrics and psychological assessment: Using, interpreting, and developing tests (2nd ed.).* Routledge.

Coulacoglou, C., & Saklofske, D. H. (2017). *Psychometrics and psychological assessment: Principles and applications (1st ed.).* Elsevier.

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

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555

Edo State Ministry of Education (2024). *Statistical data of educators and senior high schools in Edo South Senatorial District.*

Goldstein, G., Allen, D. N., & DeLuca, J. (Eds.). (2019). *Handbook of psychological assessment (4th ed.).* Elsevier. https://doi.org/10.1016/C2014-0-01970-3

Gorton, R. A., & Alston, J. A. (2023). *School leadership and administration: Important concepts, case studies, and simulations (11th ed.).* McGraw-Hill.

Hoy, W. K., & Miskel, C. G. (2013). *Educational administration: Theory, research, and practice (9th ed.).* McGraw-Hill.

Linn, R. L. (2010). *Educational measurement: Overview.* In International Encyclopedia of Education (3rd ed., pp. 45–49). Elsevier. https://doi.org/10.1016/B978-0-08-044894-7.00243-8

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Addison-Wesley.

Lunenburg, F. C., & Ornstein, A. C. (2012). *Educational administration: Concepts and practices (6th ed.).* Wadsworth Cengage Learning.

Oghenerume, R. A. (2022). *Comparative analysis of item statistics of WASSCE and NECO SSCE 2022 Data Processing multiple choice tests using item response theory* [Unpublished master's thesis]. University of Benin. http://dx.doi.org/10.13140/RG.2.2.33718.31046/1

Oghenerume, R. A., & Egberha, F. (2024). Comparative analysis of item statistics of WASSCE and NECO SSCE 2023 Data Processing multiple choice tests using item response theory. *International Journal of Educational Researchers, 15*(1), 58–67. https://doi.org/10.29329/ijer.2024.652.4

Oghenerume, R. A., & Uyi-Osaretin, S. I. (2024). Analytical comparison of item statistics employing IRT 3PLM in WASSCE and NECO SSCE 2023 English Language multiple choice tests. *KIU Journal of Education, 4*(1), 69–78. https://doi.org/10.59568/KJED-2024-4-1-07

Ojerinde, D., Popoola, K., Ojo, F., & Ariyo, A. (2014). *Practical applications of item response theory in large-scale assessment.* Marvelous Mike Press Limited.

Omoroguiwa, O. K. (2016). *SPSS guide book for statistical analysis.* Perfect Touch Prints.

Omorogiuwa, O. K. (2019). *An introduction to educational measurement and evaluation (2nd ed.).* Mase-Perfect Printing Press.

Owie, I. (2013). *Fundamentals of statistics in education and the sciences (4th ed.).* National Books.

Rust, J., Kosinski, M., & Stillwell, D. (2021). *Modern psychometrics: The science of psychological assessment (4th ed.)*. Routledge.

Sergiovanni, T. J., Kelleher, P., McCarthy, M. M., & Fowler, F. C. (2009). *Educational governance and administration (6th ed.)*. Pearson.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15*(1), 72–101.