International Journal of Scientific Research and Management (IJSRM)
||Volume||13||Issue||08||Pages||4265-4273||2025||
Website: https://ijsrm.net ISSN (e): 2321-3418
DOI: 10.18535/ijsrm/v13i08.el06

Assessing the Relationships between Gender, Years of Teaching Experience, and Proficiency in Test Construction among Ghanaian Senior High School Teachers

Ibrahim, Mohammed Gunu[1], Abu, Adam[2], Mustapha Abdul-Waris[3]

[1]Depratment of Educational Management and Policy Studies
Faculty of Education
University for Development Studies,
P. O. Box 1350
Tamale, Ghana
ORCID: 0000-0002-1608-0856
[2] Department of Mathematics and ICT Education,
Faculty of Education,
University for Development Studies,
Tamale, Ghana
ORCID: 0009-0005-4933-5324
[3]Deparment of Educational Foundation,
Faculty of Education,
University for Development Studies,
Tamale, Ghana

Abstract:
This study investigated the relationship between years of teaching experience, gender and proficiency in test construction skills among senior high school teachers in Ghana. Using a descriptive survey design, data was collected from 241 teachers selected through a multi-stage sampling procedure from four public senior high schools. A structured questionnaire employing a five-point Likert scale was administered via Google Forms to assess test construction practices. Analysis of Variance (ANOVA) revealed no statistically significant differences in test construction proficiency across different experience levels ($F = 1.143$, $p = 0.332$). However, chi-square analysis indicated significant gender differences in test construction practices ($\chi^2 = 5.306$, $p = 0.021$). The findings challenge the assumption that teaching experience alone enhances test construction abilities and highlight the need for structured professional development programmes. Recommendations of the research include implementing comprehensive assessment training, establishing mentoring systems, and developing standardized test construction guidelines regardless of teaching experience.

Keywords: Test Construction, Teaching Experience, Gender, Proficiency, Teachers

1. Introduction

Test construction is a critical skill for educators, serving as the cornerstone for assessing students' learning outcomes, guiding instructional decisions, and ensuring that educational objectives are met (see Quansah et al., 2019; Kissi et al., 2023; Alade & Odunsi, 2023; Mpuangnan, 2024). Effective test construction practices have serious implications for both students and teachers. However, proficiency in this area varies significantly among educators, often influenced by factors such as years of teaching experience. This study sought to explore the relationship between teaching experience, gender and proficiency in test construction skills. This topic holds importance for educational quality and policy reform (see Solheim & Lundetræ, 2018; Quansah et al., 2019; Lockwood, et al., 2021).

Globally, the role of standardized assessments in education has evolved, emphasizing the need for teachers to possess robust test construction skills. The ability to design valid and reliable assessments directly

impacts the accuracy of evaluating students' knowledge and competencies (Anderson and Krathwohl, 2001; Quansah et al., 2019; Ankomah, Amedahe, & Cobbinah, 2020; Mpuangnan, 2024). Yet, achieving this level of proficiency requires training, experience, and a thorough understanding of test development principles, including validity, reliability, fairness, and practicality (Wudthayagorn et al., 2022; Ankomah, Amedahe, & Cobbinah, 2020). While some educators acquire these skills through professional training programmes, others develop them through years of classroom practice (Ruegg et al., 2024). Consequently, teaching experience emerges as a potential determinant of proficiency in test construction.

In Ghana, the education system emphasizes periodic assessments to monitor student performance and inform curriculum delivery. The responsibility for constructing these assessments often falls on teachers, making their expertise in this area crucial. Despite its significance, limited accessible literature exists on how teaching experience and gender influence test construction proficiency among educators in Ghana. Anecdotal evidence suggests that experienced teachers may have an advantage due to their familiarity with instructional content and assessment techniques (Kini & Podolsky, 2016). However, this assumption has not been rigorously tested, creating a gap in understanding this phenomenon.

The issue of inconsistent test construction practices has been highlighted in several studies, with many teachers reportedly lacking the requisite skills to create effective assessment tools (Quansah et al., 2019; Rudolph et al., 2019; Ankomah, Amedahe, & Cobbinah, 2020). Inadequate training, coupled with the absence of standardized guidelines, often leads to assessments that fail to measure students' true capabilities (Awoniyi & Fletcher, 2013; Amua-Sekyi, 2016). These issues are particularly prevalent in developing countries like Ghana, where resource constraints and limited access to professional development opportunities exacerbate the situation. Understanding the dynamics between teaching experience, gender and test construction skills can inform targeted interventions to improve educational outcomes.

This study sought to understand whether a statistically significant relationship exists between the gender, years of teaching experience, and proficiency in test construction among senior high school teachers. By addressing this question, the research aims to provide insights that could shape teacher training programmes and policy recommendations, thereby contributing to the advancement of Ghana's educational system.

The significance of this research extends beyond the immediate context of Sagnarigu Municipality in Northern Region of Ghana. It offers a framework for understanding how experience impacts skill development in educational assessment, which could be applied in other regions and settings. Moreover, the findings have the potential to guide the design of professional development programmes that prioritize skill acquisition for less experienced teachers, ensuring equity in educational delivery. Based on the above the following were Hypothesised on years of teaching experience;

$\boldsymbol{H_0}$**:** There is no statistically significant variations in teachers' test construction based on their years of teaching experience among senior high school teachers in the Sagnarigu Municipality.

$\boldsymbol{H_1}$**:** There is a statically significant variations in teachers' test construction based on their years of teaching experience among senior high school teachers in the Sagnarigu Municipality.

On gender the following were Hypothesised;

$\boldsymbol{H_0}$**:** here is no statistically significant differences in test construction practices between male and female senior high school teachers in the Sagnarigu Municipality.

$\boldsymbol{H_1}$**:** There is a statically significant difference in test construction practices between male and female senior high school teachers in the Sagnarigu Municipality.

2. Methods

2.1 Research Design

The study used a descriptive survey research design to explore the relationships between teaching experience, gender and test construction proficiency. This design was chosen for its ability to accurately capture current practices and relationships between variables. As highlighted by Asenahabi (2019), the descriptive survey design effectively mirrors reality and allows for comprehensive characterization of the present situation through respondent questionnaires.

While this design presents certain challenges, such as ensuring sufficient response rates and promoting honest responses, its primary advantage lies in the ability to collect substantial data from a significant sample of teachers with varying years of experience. To address potential challenges, the study implements direct instrumentation and establishes clear credibility measures.

2.2 Population

The study targeted all teachers in four public senior high schools within the Sagnarigu Municipality, comprising a total population of 609 teachers. The accessible population consists of teachers from these four schools, representing various levels of teaching experience ranging from novice educators with less than five years of experience to seasoned professionals with over fifteen years in the field.

2.3Sample Size and Sampling Procedure

The study employed a multi-stage sampling approach to select participants from the target population of senior high school teachers. To determine the appropriate sample size, the Krejcie and Morgan formula was applied:

$$n = \frac{N}{1+N(e^2)}$$

where n represents the required sample size, N represents the population size of 609 teachers, and *e* represents the margin of error set at 5%.

Using this formula, the calculation $n = 609/(1+609(0.05^2))$ yielded a sample size of 241 teachers. The sampling process was conducted in three distinct stages to ensure comprehensive representation. First, purposive sampling was utilized to target teachers across various experience levels, ensuring inclusion on of both novice and veteran educators. This was followed by stratified random sampling to achieve proportional representation of different experience groups, guaranteeing that teachers with varying years of service were adequately represented in the final sample. Finally, simple random sampling using the lottery method was implemented for the final selection of participants within each stratum, ensuring that every teacher within their respective experience group had an equal chance of being selected for the study. This rigorous multi-stage approach was designed to minimize selection bias while maintaining representative distribution across teaching experience levels.

2.4 Data Collection Instrument

The primary data collection tool employed in this study was a structured questionnaire, meticulously designed to assess teachers' test construction skills in relation to their years of experience and gender. The instrument utilized a five-point Likert scale ranging from Strongly Agree (5) to Strongly Disagree (1) and was comprehensively organized into seven distinct sections. These sections encompassed information including years of teaching experience, gender, test construction knowledge and skills, test administration practices, exam scoring proficiency, assessment strategies, challenges in test construction, and professional development in assessment. To ensure accessibility and facilitate efficient data collection, the questionnaire was administered through Google Forms, allowing participants to complete the survey at their convenience. This digital administration method not only streamlined the data collection process but also enhanced response rates by providing teachers with a user-friendly platform for participation. The structured nature of the questionnaire, combined with its comprehensive coverage of relevant aspects and digital delivery method, ensured systematic collection of data pertinent to understanding the relationship between teaching experience, gender and test construction proficiency.

2.5 Validity and Reliability

To ensure instrument validity and reliability, a pre-test was conducted with 182 junior high school teachers in the Sagnarigu Municipality. The pre-test results were analyzed using Cronbach's Alpha, yielding a reliability coefficient of 0.81, indicating high internal consistency. Expert review and participant feedback were incorporated to enhance the instrument's validity.

2.6 Data Collection Procedures

Prior to data collection, necessary approvals were obtained from the University for Development Studies and the Ghana Education Service in the Sagnarigu Municipality. School principals' and participants consent were secured, and participants were fully informed about the study's objectives. The researchers personally managed the questionnaire distribution and collection process, ensuring participant confidentiality and providing clear instructions for completion.

2.7 Data Analysis

The study employed a comprehensive statistical analysis approach utilizing both descriptive and inferential statistics, with all analyses conducted using SPSS software. Descriptive statistical measures, including means, frequencies, percentages, and standard deviations, were calculated to characterize the relationship between teaching experience. gender and test construction skills, providing a clear overview of the patterns and trends in the data. For deeper analytical insights, inferential statistical methods were applied, primarily employing one-way ANOVA to test for significant differences in test construction proficiency across various experience levels among teachers. Additionally, correlation analysis was conducted to examine both the strength and direction of the relationship between years of teaching experience and test construction proficiency. To ensure statistical rigor and maintain research standards, the significance level was established at 0.05 for all statistical tests, providing a consistent threshold for determining the statistical significance of the findings and relationships identified in the study.

2.8 Ethical Considerations

The study adheres to strict ethical guidelines throughout all phases. Participants' privacy and anonymity are protected through confidential handling of data. Informed consent is obtained from all participants, who are made aware of their right to withdraw from the study at any time. Data analysis and reporting focus on emerging patterns rather than individual responses, ensuring participant confidentiality is maintained throughout the research process. This methodology is designed to provide a comprehensive and reliable investigation of the relationship between teaching experience and test construction proficiency, while maintaining high ethical standards and research rigor.

3. Results and Discussion

3.1 Gender Distribution

The gender distribution of respondents was analysed to assess the potential impact of gender on test construction practices. From Table 1, out of 241 participants, 166 were male, representing 68.9% of the total, while 75 were female, accounting for 31.1% of respondents. This disparity indicates a gender imbalance in the sample, which may reflect broader trends in the teaching profession at the senior high school level in the Sagnarigu Municipality. The higher proportion of male teachers suggests that gender-based comparisons in test construction skills should account for this imbalance, which may influence the findings.

Table 1: Gender distribution

	Frequency	Percent	Cumulative Percent
Female	75	31.1	31.1
Male	166	68.9	100.0
Total	241	100.0	

3.2 Years of Teaching Experience

Participants' years of teaching experience varied significantly, providing a diverse perspective on how experience correlates with test construction skills. From Table 2, nearly half of the participants (48.5%) had taught for 5-10 years, while 17% had 11-15 years of experience. Teachers with less than 5 years of experience comprised 26.6% of the sample, and only 7.9% had more than 15 years of teaching experience. The data highlights a concentration of participants within the mid-range of teaching experience, providing a robust basis for examining how teaching tenure influences test construction proficiency.

Table 2 Years of Teaching Experience

	Frequency	Percent	Cumulative Percent
Less than 5 years	64	26.6	94.2
5-10 years	117	48.5	67.6
11-15 years	41	17.0	19.1

More than 15 years	19	7.9	100.0
Total	241	100.0	

3.3 Knowledge and Perceptions of Achievement Tests

Respondents were surveyed on their familiarity with various types of achievement tests to assess their perceptions of effectiveness in measuring student learning outcomes. The results in Table 3 underscored a strong consensus among teachers about the utility of different test formats.

Table 3 Results on Teachers knowledge about Classroom Test Construction

	N	M	SD
I am familiar with test construction principles.	241	3.96	1.322
I am able to develop test items that align with learning objectives.	241	4.50	.578
I can develop clear and unambiguous test items.	241	4.49	.620
I am able to develop test items that are appropriate for the target population.	241	4.48	.626
I can develop test items that assess a range of cognitive skills.	241	4.43	.616
I am able to develop test items free from bias and cultural stereotypes.	241	4.49	.620
I am able to develop reliable and valid test items.	241	4.48	.620
I can develop test items appropriate for different levels of Bloom's taxonomy.	241	4.29	.643
I am able to develop test items that are appropriate for different types of learners.	241	4.44	.668
I am able to develop test items that align with curriculum standards.	241	4.46	.658
I am able to develop test items that measure what they are intended to measure.	241	4.46	.639
I am able to develop test items appropriate for different levels of difficulty.	241	4.38	.704
I am able to develop test items that are appropriate for different types of assessment (e.g., formative, summative).	241	4.42	.673
I am able to develop test items that can be scored objectively and consistently.	241	4.41	.654
I am able to develop test items that are appropriate for different types of assessment formats (e.g., multiple choice, essay).	241	4.54	.605
I am able to develop test items that are appropriate for different types of assessment purposes (e.g., diagnostic, evaluative).	241	4.39	.656
I am able to develop test items free from errors and ambiguities.	241	4.31	.699
I am able to develop test items that are appropriate for different types of content (e.g., factual, conceptual).	241	4.40	.658
I am able to develop test items that are appropriate for different types of learning domains (e.g., affective, psychomotor).	241	4.38	.691
I believe that adherence to test construction principles is essential for ensuring achievement tests' validity and reliability.	241	4.46	.683

The highest-rated test types were multiple-choice and essay-based tests, with mean scores of 4.51 and 4.44, respectively, indicating strong teacher confidence in their utility for assessing student understanding. Conversely, adaptive achievement tests received the lowest mean score of 2.60, reflecting limited familiarity or perceived relevance among respondents.

3.4 Gender and Test Construction Practices

The study investigated potential gender-based differences in test construction practices. A chi-square test was used to explore the relationship between gender and test construction. Table 4 presents statistically significant findings regarding these relationships. The Pearson chi-square is 5.306 with one degree of

freedom and a two-sided significant value of 0.021, indicating a significant relationship between gender and test construction. This suggest that men and women construct test differently, and the observed differences is statistically significant, meaning it unlikely to be due to chance alone.

The continuity-corrected chi-square statistics is 4.664 with one degree of freedom, with a two-sided significance value of 0.031, further confirming the significant relationship that exists between gender and test construction. The adjusted chi-square statistics, being larger, reflects a low probability of the chi-square value observed occurring by chance, further providing additional support to the hypothesis.

Additionally, the likelihood ratio chi-square statistics, which compares the likelihood under the null and alternative hypothesis, is 5.465 with one degree of freedom and a two-sided significance value of 0.019. This reinforces the rejection of the null hypothesis, confirming a statistically significant relationship between gender and test construction.

Fisher's exact test, particularly suitable for smaller size or cases with low expected counts, yields an exact two-sided significance value of 0.022 and an exact one-sided significance value of 0.015. These results underscore the significant relationship, with the p-value of 0.022 being well below the 0.05 threshold for rejecting the null hypothesis.

Table 4. Crosstab between Gender and Test construction

			Test construction		Total
			No	Yes	
What is your gender?	Female	Count	20	55	75
		Expected Count	28.0	47.0	75.0
	Male	Count	70	96	166
		Expected Count	62.0	104.0	166.0
Total		Count	90	151	241
		Expected Count	90.0	151.0	241.0

3.5 Years of Teaching Experience and Test Construction

To assess the impact of teaching experience on test construction proficiency, an analysis of variance (ANOVA) was conducted and findings is presented in Table 5. The results did not show statistically significant differences in test construction practices across the various experience groups.

Table 5

Table 13. Multiple Comparisons

Years of Teaching Experience.
Tukey HSD

(I) Years of Teaching	(J) Years of Teaching	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Less than 5 years	5-10 years	.011	.109	1.000	-.27	.29
	11-15 years	.122	.140	.820	-.24	.48
	More than 15 years	.294	.183	.376	-.18	.77
5-10 years	Less than 5 years	-.011	.109	1.000	-.29	.27
	11-15 years	.110	.127	.820	-.22	.44
	More than 15 years	.283	.173	.361	-.16	.73
11-15 years	Less than 5 years	-.122	.140	.820	-.48	.24
	5-10 years	-.110	.127	.820	-.44	.22
	More than 15 years	.172	.194	.812	-.33	.67
More than 15 years	Less than 5 years	-.294	.183	.376	-.77	.18
	5-10 years	-.283	.173	.361	-.73	.16
	11-15 years	-.172	.194	.812	-.67	.33

The ANOVA analysis provided insights into whether years of teaching experience significantly influence test construction proficiency among senior high school teachers. The test compared the mean proficiency scores across four groups: teachers with less than 5 years of experience, 5-10 years, 11-15 years, and more than 15 years.

Despite variations in teaching experience, the F-value of 1.143 and a p-value of 0.332 indicated no statistically significant differences in test construction proficiency across these groups. This result suggests that teaching experience, as measured in years, does not have a substantial impact on a teacher's ability to construct tests effectively.

Between-group variability was low, as evidenced by a sum of squares of 1.674, highlighting minimal differences in proficiency that could be attributed to teaching experience. Conversely, within-group variability was much larger, with a sum of squares of 115.703, indicating that most of the variability in test construction proficiency occurred within individual experience groups rather than between them.

Mean comparisons showed slight differences between the groups, but these variations were not statistically significant. This finding was further supported by overlapping confidence intervals for the mean scores of each group, suggesting that any observed differences could be due to random sampling error rather than genuine disparities in proficiency.

The results challenge the common assumption that teaching experience inherently improves test construction skills. Instead, they underscore the need for targeted professional development programmes that address specific skill gaps regardless of a teacher's years of service. Continuous exposure to modern assessment techniques and structured training may be more critical than experience alone in enhancing teachers' test construction capabilities.

Furthermore, these findings highlight the importance of exploring other potential factors influencing test construction proficiency, such as subject specialization, access to resources, and attitudes toward assessment practices. Future research could delve deeper into these variables to provide a more comprehensive understanding of what drives proficiency in test construction.

4 Conclusion

This study aimed to investigate the relationship between years of teaching experience, gender and proficiency in test construction skills among senior high school teachers. The research sought to determine whether more experienced teachers demonstrated higher levels of proficiency in developing and implementing assessment tools compared to their less experienced colleagues, while also examining potential gender differences in test construction practices. This finding was particularly noteworthy given the substantial within-group variability (SS = 115.703) compared to the between-group variability (SS = 1.674), suggesting that individual factors beyond teaching experience play a more crucial role in determining test construction proficiency. An unexpected but significant finding emerged through chi-square analysis ($\chi^2 = 5.306$, $p = 0.021$), which revealed substantial gender differences in test construction practices. This result was further validated through Fisher's exact test ($p = 0.022$), indicating that male and female teachers approach test construction differently. The robustness of this finding was supported by multiple statistical measures, including the likelihood ratio chi-square statistic (5.465, $p = 0.019$) and the continuity-corrected chi-square statistic (4.664, $p = 0.031$). These findings challenge conventional assumptions about the relationship between teaching experience and assessment competency. They suggest that the development of effective test construction skills may be more closely linked to specific training and professional development opportunities rather than years of classroom experience. The significant gender differences identified also highlight the need for a more nuanced understanding of how different demographic factors influence assessment practices in educational settings. The research therefore recommends that, educational authorities should establish comprehensive professional development programmes specifically focused on test construction skills for all teachers, regardless of their years of experience. Such programmes should include practical workshops, hands-on training sessions, and regular refresher courses that cover modern assessment techniques, item writing principles, and test validation methods. Educational institutions should also implement mentoring systems that pair teachers based on demonstrated assessment proficiency rather than years of experience. These mentoring relationships should be formalized within school structures, with dedicated time allocated for collaboration, peer review of assessment materials, and shared learning experiences. Given the significant gender differences identified in test construction practices, schools should develop gender-sensitive assessment guidelines that promote

consistency while acknowledging and leveraging diverse approaches to test development. Regular monitoring and evaluation of these standards should be conducted to ensure compliance and identify areas needing improvement. This research is not based on longitudinal studies; further research should be examined with the longitudinal research design. Longitudinal.

References

1. 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.
2. Ankomah, F., Amedahe, F. K., & Andrews Cobbinah. (2020). The Issue of Non-Adherence to Test Construction Principles: Do Teachers' Commitment to Teaching or Attitude by the Teacher Really Matter? *International Journal of Social Sciences & Educational Studies*, *7*(4), 109–121. https://doi.org/10.23918/ijsses.v7i4p109
3. Amua-Sekyi, E. T. (2016). Assessment, Student Learning and Classroom Practice: A Review. *Journal of Education and Practice*, *7*(21), 1-6.
4. Asenahabi, B. M. (2019). Basics of research design: A guide to selecting appropriate research design. *International Journal of Contemporary Applied Researches*, *6*(5), 76-89.
5. Awoniyi, F. C., & Fletcher, J. A. (2013). Senior high school mathematics teachers' use of school-based assessment guidelines and test scores in the Cape Coast Metropolis of Ghana. *Journal of Educational Development and Practice*, *4*(1), 19-40.
6. Kini, T., & Podolsky, A. (2016). Does Teaching Experience Increase Teacher Effectiveness? A Review of the Research. *Learning Policy Institute*, 1-8.
7. Kissi, P., Baidoo-Anu, D., Anane, E., & Annan-Brew, R. K. (2023). Teachers' test construction competencies in examination-oriented educational system: Exploring teachers' multiple-choice test construction competence. In *Frontiers in Education* (Vol. 8, p. 1154592). Frontiers Media SA.
8. Lockwood, A. B., Farmer, R. L., Bohan, K. J., Winans, S., & Sealander, K. (2021). Academic Achievement Test Use and Assessment Practices: A National Survey of Special Education Administrators. *Journal of Psychoeducational Assessment*, *39*(4), 436–451. https://doi.org/10.1177/0734282920984290
9. Mpuangnan, K. N. (2024). Determining skills of basic schoolteachers in test item construction. *Review of Education (Oxford)*, *12*(2). https://doi.org/10.1002/rev3.3478
10. 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
11. Rudolph, M. J., Daugherty, K. K., Ray, M. E., Shuford, V. P., Lebovitz, L., & DiVall, M. V. (2019). Best practices related to examination item construction and post-hoc review. *American journal of pharmaceutical education*, *83*(7), 7204.
12. Ruegg, R., Hoang, H., & Petersen, N. (2024). The influence of English language proficiency test scores on the academic success of ESL undergraduate students. *Educational Research and Evaluation*, *29*(3-4), 111-129.
13. Solheim, O. J., & Lundetræ, K. (2018). Can test construction account for varying gender differences in international reading achievement tests of children, adolescents and young adults? - A study based on Nordic results in PIRLS, PISA and PIAAC. *Assessment in Education : Principles, Policy & Practice*, *25*(1), 107–126. https://doi.org/10.1080/0969594X.2016.1239612
14. Wudthayagorn, J., Piamsai, C., & Thanompongchart, P. (2022). The Development of a Local Online English Proficiency Test for Young Learners: Key Considerations and Challenges. *PASAA: Journal of Language Teaching and Learning in Thailand*, *64*, 126-137.

Authors Profile

Ibrahim M. Gunu received the B.Ed., M.A and PhD Degrees from University of Cape Coast, University of Ghana and Keele University in 2001, 2007 and 2017 respectively. He has published extensively in Education.

Abu Adam received the B.Ed (Psychology) and MPhil in Measurement and Evaluation from University of Cape Coast in 2007 and 2013 respectively. He is currently pursuing his PhD (Measurement and Evaluation) at the University of Cape Coast.

Mustapha Abdul-Waris received the B.Ed (Mathematics) and MPhil (Measurement in Evaluation) from Valley View University and University for Development Studies in 2010 and 2025 respectively.