

Assessment of Multiple-Choice Construction Competence Among Public Junior Secondary School Teachers in Edo Central Senatorial District, Nigeria.

Ehigbor, B. O., Ph.D, Osumah, Obaze Agbonluae, Ph.D
Department of Guidance and Counselling, Faculty of Education,
Ambrose Alli University, Ekpoma, Nigeria

DOI: <https://doi.org/10.47772/IJRISS.2023.7442>

Received: 19 March 2023; Accepted: 07 April 2023; Published: 02 May 2023

ABSTRACT

This study assessed multiple-choice construction competence among public junior secondary school teachers in Edo Central Senatorial District, Nigeria. Two research questions were raised and one hypothesis was formulated and tested. This study adopted the descriptive study design. The population of the study covers the six hundred and sixteen (616) Secondary School teachers in all the 69 public secondary schools in Edo Central Senatorial District. The simple random sampling technique was used to select 125 subject head teachers who were drawn as the representative sample to measure teachers' competence. These subject head teachers cover language subject head teachers, Science subject head teachers, and Art subject head teachers in schools. These subject teachers were drawn using purposive sampling because they are like supervising teachers who work hand in hand with the management and school principals on administrative and management related duties. The instrument used was a 20-item survey questionnaire adapted from the work of Kissi (2020). The instrument was used to measure multiple-choice construction competencies and items' quality among teachers. The test-retest reliability coefficient produced an r-value of 0.74 which shows that the instrument is reliable. Research question 1 was analyzed using mean () while the independent t-test for two sample means was used to test the hypothesis at 0.05 level of significance. Findings revealed that the level of multiple-choice construction competence among public junior secondary school teachers in Edo Central Senatorial District, Nigeria is low and female teachers had higher level of multiple-choice construction competencies than male teachers in secondary schools in the district. It was concluded that sex differences in favour of female teachers, exist in teachers' multiple-choice construction competencies. It was recommended that the district education administrators from the Local Government Education Council Authorities should place more emphasis on exposing classroom teachers on more practical ways of constructing multiple-choice items especially on how to effectively handle items? alternatives. This can be achieved through workshops, or training in test construction.

Keywords: Multiple-Choice Test, Construction Competence, Assessment

INTRODUCTION

Enhancing student learning is one of assessment's main goals. The process of assessment involves educators using techniques and instruments to gauge the depth of students' learning across a variety of activities (Opatye, 2016). In order to enhance programs and learning, assessment entails the use of empirical data on student learning. To get a comprehensive grasp of what students know, comprehend, and can accomplish with their knowledge as a consequence of their educational experience, it is a process of obtaining and analyzing data from many and varied sources (Ibrahim, Ibrahim, & Amina, 2022). This indicates that evaluation considers a variety of approaches to comprehending learners' academic success. Tests are used as an evaluation tool because classroom instructors must create and administer them in order to gather certain

crucial information about what pupils learn throughout the teaching and learning process (Opie, Oko-Ngaji, Eduwem, & Nsor, 2021). Testing not only gives instructors and students insight into how much has been learned, but it also leaves opportunity for further learning to occur. For instructors to be able to accurately track their students' development, it is imperative that their assessments be of higher quality (Opataye, 2016). Tests created by instructors in schools have an effect on students' lives since they are utilized to make judgments that have an influence on them. Assessments should also be given appropriate respect since they help pupils prepare for standardized achievement tests. According to Opie, et al. (2021), tests provide insight into whether learning really occurred in the classroom, enhance instruction, identify issues with the way they are being taught, and highlight students' strengths and limitations. They emphasized that testing offers helpful data for selecting kids, instructors, and a whole program. This suggests that assessments might be considered a tool for assessing whether or not learning goals have been met. Achievement exams are used to assess a person's prior knowledge. It is intended to assess a person's degree of competence, success, or understanding in a certain field. An accomplishment exam is a kind of aptitude test designed to assess what a person has learnt to perform as a consequence of deliberate training (Quansah, Amoako, & Ankomah, 2019). Educator-made or standardized assessments may be used as achievement tests. The technique, equipment, and scoring of a Standardized Achievement Test (SAT) have been predetermined so that the exact same testing procedures may be used at many times and locations. Achievement tests created by classroom instructors are known as "Teacher Made Tests" (TMTs). TMTs may be written or spoken assessments that are neither mass-produced nor standardized (Quansah *et al.*, 2019). It may be an organized written or spoken assessment of pupils' progress. Multiple-choice questions, matching questions, true-false questions, fill-in-the-blank questions, and essay questions are all common exam elements used by teachers. The specific characteristics, usability, validity, and applicability of these two categories of accomplishment assessments are different. The validity and reliability of these accomplishment tests must be increased since TMTs are crucial to the Nigerian educational system and are heavily used in secondary school placement, ongoing evaluation, and final exams (Kissi, 2020). The design of the test has a significant role in determining how well the students have understood the material. It is said to be the process used to manufacture test products. Test construction, according to the National Centre for Assessment in Higher Education (2015), involves using practical and scientific principles before, during, and after each item before it is eventually included in the test. The ability to build valid and reliable test items is part of the knowledge of test construction, which may be learned via experience or formal study. Teachers may be certain that test results accurately represent students' mastery of fundamental abilities or information by developing qualitative test items. The creation of high-quality test items takes time, and test building expertise is required. With the assistance of the Table of Specifications (TOS), instructors may create tests with a high level of content validity thanks to their understanding of test construction. The TOS is a table that lists subjects that will be studied or have already been covered in a curriculum. It also includes the goals that must be met (Kissi, 2020). It serves as a strategy or guide for test preparation since it specifies the amount of questions that will be asked on each subject or unit of instruction, as well as the questions' format, content, and other factors. The expertise of senior secondary school instructors in the study area about test creation is the major subject of this investigation. One important aspect in the educational system that deserves careful thought is the creation of exams. One might attribute the low SAT performance of kids in Edo State on the use of subpar TMTs to prepare them. Students often take a series of TMTs at different schools, with the goal of adequately preparing them for the SATs. The results of SAT tests, such as the Basic Education Certificate Examination (BECE), are crucial to students' plans to continue their education. Based on their evaluated talents, it chooses whether to allocate students to the sciences, technical/vocational, art, or commercial fields. The effectiveness of the TMTs utilized on the children has an impact on how well they do on the SATs. Therefore, it is crucial to carry out a research to evaluate teachers' proficiency in creating multiple-choice tests in secondary schools in Edo Central Senatorial District of Nigeria.

OBJECTIVES

The main objective of this investigation is to assess multiple-choice construction competence among public junior secondary school teachers in Edo Central Senatorial District, Nigeria. Specifically, the study sought to:

1. examine the level of multiple-choice construction competence among public junior secondary school teachers in Edo Central Senatorial District, Nigeria
2. Determine whether sex differences exist in the level of multiple-choice construction competence among public junior secondary school teachers in Edo Central Senatorial District, Nigeria

RESEARCH QUESTIONS

The following research questions were raised to guide the study

1. What is the level of multiple-choice construction competence among public junior secondary school teachers in Edo Central Senatorial District, Nigeria?
2. Is there any sex difference in the level of multiple-choice construction competence among public junior secondary school teachers in Edo Central Senatorial District, Nigeria?

HYPOTHESIS

The hypothesis formulated and tested in this study:

1. There is no significant sex difference in the level of multiple-choice construction competence among public junior secondary school teachers in Edo Central Senatorial District, Nigeria.

LITERATURE REVIEW

The sort of information a test may offer on students' results directly affects how well it is designed by a teacher (McCoubrie, 2004). A well-written exam enables the instructor to precisely and consistently assess the level of student understanding of a particular subject matter covered in class. The results of these assessments enable instructors to gauge, to some extent, the effectiveness of their teaching (Karimi, 2014). On the other hand, badly designed test questions may result in erroneous assessments of learning and provide misleading information about both student performance and the success of training (Van der Merwe, 2015). The usefulness of a test item is diminished by any features that draw the test taker's attention away from the main idea or subject (Salihu, 2019). Any question that is rightly or erroneously answered due to unrelated aspects in the question will provide both the examinee and the examiner false feedback (McCoubrie, 2004). Capan Melsner, Steiner-Hofbauer, Lilaj, Agis, Knaus, & Holzinger (2020) described variables that are inherent in badly designed tests and which, if managed correctly, would result in high-quality classroom-based examinations. They include (i) a lack of information about the test's targeted audience, the skill or area of ability it was designed to evaluate, the amount of time allotted for each test question, and the number of points test-takers would receive for each accurate answer. (ii) The distinct portions are not defined clearly. Test questions that have more than one viable response because they were poorly formulated. (iv) Failure to indicate on the papers how much time was allotted for each activity. Just the overall amount of time available to complete all the tasks was provided. (v) The test's structure did not take pupils' grade levels into account. (vi) Confusing directives. (vii) Assignments given to pupils that conflict with the assignments they are given during class instruction. (viii) The items are not sufficiently reflective of the subject matter the instructor wishes to assess. (ix) The creation of test item collages. The Contexts, Input, Process, and Product (CIPP) model proposed by Stufflebeam in 1983 served as the foundation for this investigation (Stufflebeam, 2004). The primary purpose of the CIPP model was to ascertain how educational goals may be evaluated in order to reach suitable judgments that could have a

significant impact on the expansion and development of the educational system (Aziz, Mahmood & Rehman, 2018). The supporters of this model based their arguments on the system approach to education, which included making choices about inputs, processes, and outcomes. The creator of this approach believed that as reliable information is essential to making wise decisions, researchers and planners may best serve education by supplying it to those in the field of education who need it, such as policymakers, teachers, administrators, and other stakeholders (Agustina & Mukhtaruddin, 2019). This strategy made it apparent who would utilize the outcomes, how they would be used, and which component of the system they are making choices for by emphasizing distinct levels of decisions and decision makers. The CIPP methodology is all-inclusive and offers four interconnected evaluation levels.

Context assessment: this informs choices on how to develop programs. It entails identifying issues that the educational program has to solve (Aziz, Mahmood & Rehman, 2018). In order for the aims of teaching and learning to be completely realized, a teacher's job in the classroom is to define and analyze the goals and objectives of a specific topic. As a result, the teacher takes into account the needs and gaps between what is and what ought to be.

Input Appraisal: This is done to aid in formulating choices. What resources are offered? Analyzing the current resources—human and material—as well as the financial, physical, and temporal requirements to accomplish program and organizational goals might provide an answer. What alternative tactics are required for the program's success would be addressed when resource availability is established (Stufflebeam, 2004). Teachers may then prioritize the tactics in their classrooms to determine which ones seem to have the highest chance of meeting the system's performance requirements. This will make it easier to create multiple choice questions that are beneficial for the assessment process.

Process assessment: Process assessment is intended to assist in putting choices into action. How effectively is the plan being carried out? What obstacles stand in its way of success? What changes are required? Data gathering for tracking the program's actual execution is the focus of process evaluation (Stufflebeam, 2004). In this sense, it also refers to the teacher's effective utilization of the inputs in achieving the objectives. While making decisions, instructors might use the process evaluation stage to determine if certain resources were used too much or too little.

Product Assessment: This step enables individual instructors to assess the degree to which the program's (subject's) goals and objectives have been met. At this point, queries such "what is the product's quality" are addressed (Agustina & Mukhtaruddin, 2019). This makes it easier to decide afterwards if the instruction should be continued, stopped, revised, reevaluated, or refocused. The four levels mentioned above all aim to assess the degree to which a subject's predetermined goals are met with the constrained availability of high-quality resources. The CIPP model is pertinent to this research since it aims to illustrate the link between input factors and their effects on the targeted learning outcomes. It is also pertinent to this research since it focuses on input factors and students' total learning result or accomplishment to provide reliable data that will aid senior secondary schools in making decisions. This would undoubtedly help educators in their decision-making about the supply of resources in the proper quality and quantity as well as the efficient delivery of programs that are governed by the established educational system objectives. Input evaluation is specifically concerned with the examination of information-providing for assessing the influence of teachers' personal attributes, such sex, on teachers' skills (Sopha & Nanni, 2019). In order to provide a feedback mechanism for the detection and prediction of shortcomings in the procedural design and implementation of policies aimed at improving learners' achievement, the process assessment component of the model assists educators in understanding the impact of utilization status of variables of this study in improving learners' achievement using a well-constructed multiple choice item test. After a time of teaching and learning, the product part of the model measures and assesses the success of the results of the use of the study's input variables in learners' achievement. Moreover, program development, operation, and

assessment may be guided by the CIPP model (Sopha & Nanni, 2019). Due to the fact that it offers a structure for record keeping that makes it easier for the public to assess educational needs, goals, plans, actions, and results, it is also helpful for program development and even accountability reasons. Also, it satisfies public demand for program information. An Input, Process, Product Paradigm of assessing learners' readiness based on the CIPP model is presented diagrammatically in Figure I below:

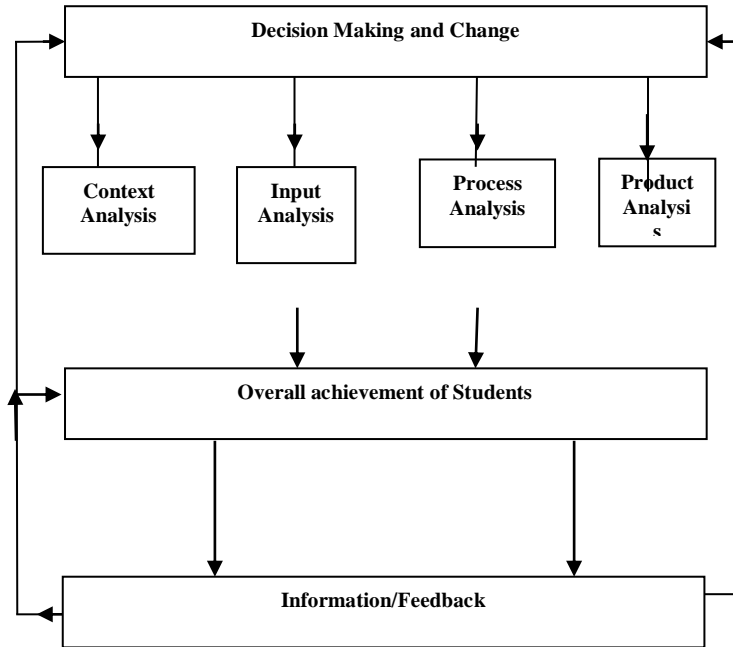


Fig. 1: Paradigm of Assessing Learners' readiness based on the CIPP Model

Source: Adapted from Stufflebeam (1983)

Figure 1 shows how information arising from the analysis of students' achievement in an input, process and product assessment passes through the information loop to the decision making loop for making the decision arrived at. This ensures that set criteria are met or even modified to meet the goals of education. This explains why the arrow moves from the decision making alternatives to achievement of students and back to information loop, giving room for recycling of decisions. The findings and recommendation and conclusion arrived at in any of the four types of assessment in the CIPP model can provide necessary information for decision makers to enhance the decision process in any of the four items in the model. Noviani (2016) looked into the procedures teachers used to create multiple-choice test questions, challenges Indonesian instructors encountered, and revealed the quality of the test items they produced. This descriptive research included 211 students' response sheets, 275 multiple-choice exam questions, and 6 English instructors. The results demonstrated that the way instructors still create multiple-choice exam items is not ideal. Several problems were discovered in the test products' quality, indicating the need for more modification and improvement. At Nigeria's Nasarawa State senior secondary schools, Salihu (2019) analyzed the exam construction and economics content validity of teachers. The study used a co-relational research approach and content analysis. 95 Economics instructors were chosen at random from Nasarawa North's senior secondary public and private schools. The study's conclusions showed that secondary school instructors have found it difficult to create tests with economics content validity. This is due to a variety of factors, including inaccurate evaluations of students' accomplishments, bad language and sentence structure, and a lack of adequate monitoring by the evaluators, among others, caused by instructors' lack of test building abilities. Opie, Oko-Ngaji, Eduwem, and Nsor (2021) investigated the degree to which science instructors in secondary schools in Yala Local Government Area of Cross River State used their understanding of the test building technique in establishing Multiple Choice Objective Tests. The descriptive survey study method was utilized to do this, and a deliberate sample of 87 Chemistry

instructors was chosen from the 213 scientific teachers in the region. The findings demonstrated that science teachers designed objective exams with the help of their understanding of test creation techniques, but they did not administer a pre-test to the students prior to the main test. The assessment of multiple-choice test creation proficiency among Nigerian public junior secondary school teachers has been the subject of several researches. The test-building abilities of instructors at senior high schools (SHS) in the Cape Coast Metropolis were examined by Quansah, Amoako, & Ankomah (2019). Three selected SHS in the Cape Coast Metropolis were chosen at random (lottery technique) to provide samples of the End-of-Term Examination papers in Integrated Science, Core Mathematics, and Social Studies. The findings showed that the instructors' abilities to design end-of-term exams were restricted. This was clear when problems with the test's content representativeness and relevance, as well as the validity and fairness of the assessment tasks that were assessed, were discovered. The link between senior high school teachers in the Kwahu-South District's multiple-choice test creation skills and the quality of the multiple-choice test questions they create, as well as the impact of the teachers' number of years of experience in the classroom, was examined by Kissi (2020). Correlational research design was used in the quantitative method. The study's population consisted of 157 different instructors in total. Findings showed that, there is no correlation between instructors' test construction skills and the quality of the multiple-choice test questions and that there are high levels of multiple-choice test construction competence. Using a descriptive study methodology, Ibrahim, Ibrahim&Amina (2022) assessed teachers' test building knowledge in the Ungogo Local Government Area (LGA) of Kano State, Nigeria. The 260 high school teachers in the local government employed for the research made up the total population, and the results showed no discernible gender or degree of experience variations in teachers' knowledge of test creation in Ungogo. In a setting of Open and Distance Learning (ODL) higher education, Opatye (2016) investigated the amount of difficulty lecturers' encounter when creating test items for different kinds of examinations depending on gender and institutional method of delivery. The study population consisted of lecturers from one single-mode and one dual-mode ODL institution in South West Nigeria. The research paradigm used was descriptive cross-sectional survey research. 240 professors were chosen as the sample using stratified simple random sampling procedures. The findings revealed that the test item creation was somewhat tough for ODL professors. Case study, multiple choice, matching, essay, and completion tasks were more challenging for female ODL instructors to create than for male colleagues. The difficulties that male and female ODL teachers had in creating exam items varied significantly. Opatye (2016) investigated the use of Differential Item Functioning (DIF) in senior school certificate exams to identify gender-biased items in Economics multiple-choice questions. The study used a causal comparative or ex-post factor research design. 2,985 Senior Secondary School Three (SS3) Economics students in Enugu State's Nsukka school zone made up the study's population. For the research, a sample size of 339 SS3 Economics students was employed. The West African Examination Council's 50-item multiple-choice questions in Economics from the 2018 SSCE were the study's instrument (WAEC). Using the Kuder-Richardson formula, a reliability coefficient of 0.87 was derived. The study's data were examined utilizing the Logistic Regression technique. Out of 50 items in the 2018 WAEC Economics questions, 14 items—or 28% of the test items—displayed significant gender DIF at the 0.05 level of significance, according to the research. Only one item—representing 2% of the 14 items—was found to demonstrate substantial gender DIF in favor of male students, while 13 items—representing 26% of the items—differentially functioned in favor of female students.

METHODS

The descriptive research design based on survey methods was used for this investigation. Six hundred sixteen (616) secondary school teachers from all the public secondary schools in the Edo Central Senatorial District make up the study's population. For the purpose of determining the competency of instructors, 125 subject heads were chosen at random and served as the representative sample. These topic heads include

language subject heads, science subject heads, and art subject heads at educational institutions. To eliminate biases and sentiments that may be connected with gathering a self-report on instructors' competency, it was decided to use head teachers as proxy participants to evaluate their teachers' multiple-choice item creation ability. A survey questionnaire titled: Teachers' Multiple-Choice Construction Competencies and Item Quality was adopted from Kissi's work (2020). A total of 20-items on the test were scored on a five-point Likert type scale, with Strongly Agree being rated at 5, Agree being rated at 1, and neither agreeing nor disagreeing being rated at 3. By deleting the neutral answer of 3, the response ratings were altered during adaptation from a 5-point scale to a 4-point scale. As a result, the scale was adjusted from Highly Competent -4 to Highly Incompetent -1. Two specialists from Ambrose Alli University in Ekpoma's Department of Guidance and Counselling evaluated the instrument's face and content validity. The instrument's dependability was assessed using the test-retest reliability approach. In order to carry out this method, 30 senior secondary school students from outside the research sample and inside the study region were given copies of the instrument. The same responders were given the same instrument a second time a few weeks later. The Pearson's Product Moment Correlation method was used to correlate their replies from the first and second tests. The coefficient resulted in an r-value of 0.74, demonstrating the instrument's dependability. Before completing the study activity, the copies of the questionnaire were distributed by contacting the principals of the chosen schools to request their consent. Standard deviation (SD) and the mean (X) were used to analyze research question 1. (S.D). The degree of multiple-choice construction skill was assessed using a criteria mean of 2.50. This result was arrived at by summing the four (4) likert scales (Highly Competent – 4, Competent – 3, Incompetent – 2, and Highly Incompetent – 1), and then dividing the total (10), which equals 2.50, by the total scales (4). So, a score of 2.50 or more denotes strong competence, whereas a score of 2.40 to 2.49 denotes poor competence. The 0.05 threshold of significance independent t-test for two sample means was employed to test the hypothesis. In order to evaluate and test the study topic and hypothesis, the Statistical Package for Social Science (SPSS) was used (IBM Version 20).

FINDINGS

The result of the analysis are presented as follows

Research Question 1: What is the level of multiple-choice construction competence among public junior secondary school teachers in Edo Central Senatorial District, Nigeria?

Table 1: Analysis of Multiple-Choice Construction Competence among Public Junior Secondary School Teachers in Edo Central Senatorial District, Nigeria

	Item statements	N = 125		
			S.D	Remark
1.	properly space the test items for easy reading	2.59*	0.919	HC
2.	review test items for construction errors	2.55*	0.817	HC
3.	keep all parts of an item (stem and its options) on the same page	2.49	0.858	LC
4.	use appropriate number of test items	2.31	0.854	LC
5.	give specific instructions on the test	2.36	0.925	LC
6.	appropriately assign page numbers to the test	2.30	1.050	LC
7.	number the test items one after the other	2.34	0.816	LC
8.	make sure each item deals with an important aspect of content area	2.45	0.954	LC
9.	match test items to instructional objectives (intended outcomes of the appropriate difficulty level)	2.52*	0.970	HC

10.	prepare marking scheme while constructing the items	2.44	1.031	LC
11.	give appropriate time for completion of test	2.30	0.933	LC
12.	pose clear and unambiguous items	2.42	0.982	LC
13.	include questions of varying difficulty	2.26	0.960	LC
14.	match items to vocabulary level of the students	2.49	0.920	LC
15.	make alternatives approximately equal in length	2.56*	0.946	
16.	include in the stem any word(s) that might otherwise be repeated in each alternative	2.40	0.915	LC
17.	present alternatives in some logical order (e.g., chronological, most to least, alphabetical) when possible	2.71*	0.927	HC
18.	make alternatives independent of each other	2.56*	0.946	HC
19.	avoid the use of “none of the above” as an option when an item is of the best answer type	2.40	0.915	LC
20.	make the alternatives grammatically consistent with the stem	2.66*	0.976	HC
	Overall mean = 2.381			

HC – High competence LC–Low Competence

The result in Table 1 shows that majority of the respondents had low response score mean on items 1, 2, 9, 15, 17, 18, 20, 21, 22, 23 and 30 at an overall mean score of 2.38 which is less than the criterion mean of 2.50 (i.e = $2.38 < 2.50$). Hence, this implies that the level of multiple-choice construction competence among public junior secondary school teachers in Edo Central Senatorial District, Nigeria is low.

Hypothesis 1: There is no significant sex difference in the level of multiple-choice construction competence among public junior secondary school teachers in Edo Central Senatorial District, Nigeria.

Table 2: T-test summary Analysis on Sex difference and Level of Multiple-Choice Construction Competence among Public Junior Secondary School Teachers

Variables	Gender	(n=125)	S.D	t-cal.	p-value	Remarks
Multiple-Choice Construction Competence	Male	58	2.15	0.53	3.551	Reject null hypothesis
	Female	67	2.43	0.30		

The result in Table 2 shows that the mean competence score of 2.43 for female teachers was higher than that of their male counterpart at a mean score of 2.15. Hence, this answered the research question that female teachers had a higher multiple-choice construction competence score than male teachers at a mean score difference of 0.28. However, to determine whether the observed mean score difference is significant, the test result from the test of hypotheses shows that the calculated t-value of 3.551 is statistically significant ($p < 0.05$). Therefore, the null hypothesis which states that there is no significant sex difference in the level of multiple-choice construction competence among public junior secondary school teachers in Edo Central Senatorial District, Nigeria is rejected. This implies that sex difference exist in the level of multiple-choice construction competence among public junior secondary school teachers in Edo Central Senatorial District, Nigeria.

DISCUSSION

The findings indicate that Nigeria’s Edo Central Senatorial District’s public junior secondary school teachers have a poor degree of proficiency in creating multiple-choice tests. The result shows that many of

them had low competence in keeping all parts of an item (stem and its options) on the same page, using appropriate number of test items, giving specific instructions on the test, appropriately assigning page numbers to the test, numbering the test items one after the other, and making sure each item deals with an important aspect of content area. Furthermore, many of the teachers demonstrated low competence in preparing marking scheme while constructing the items, giving appropriate time for completion of test, posing clear and unambiguous items, including questions of varying difficulty, matching items to vocabulary level of the students, including in the stem any word(s) that might otherwise be repeated in each alternative and avoiding the use of “none of the above” as an option when an item is of the best answer type. Possible reasons for this could be either low knowledge on constructing multiple-choice tests or low readiness or morale to undertake the procedure. The outcome is consistent with the results of Quansah, Amoako and Ankomah (2019), who discovered that instructors had inadequate expertise in creating end-of-term exams. This was clear when problems with the test’s content representativeness and applicability, as well as the validity and fairness of the assessment tasks, were discovered after they had been tested among school teachers in the Cape Coast Metropolis. The outcome conflicts with Kissi’s (2020) research, which revealed that senior high school instructors in the Kwahu-South District had high levels of proficiency in creating multiple-choice tests. Findings also agree with Noviani (2016) whose result demonstrated that the way teachers still create multiple-choice exam items is not ideal in Indonesia as several problems were discovered in the test products’ quality, indicating the need for more modification and improvement. Furthermore, findings are consistent with the result of Salihu (2019) who found that secondary school teachers in Nasarawa State found it difficult to create tests with Economics content validity. This is due to a variety of factors, including inaccurate evaluations of students’ accomplishments, bad language and sentence structure, and a lack of adequate monitoring by the evaluators, among others, caused by instructors’ lack of test building abilities. As female teachers in schools had higher competency skills in multiple-choice construction than male teachers in schools, the results also demonstrated that there are gender differences in the level of multiple-choice construction competence among public junior secondary school teachers in Edo Central Senatorial District, Nigeria. This contrasts with the findings of Ibrahim, Ibrahim, and Amina (2022), who found no significant variations in teachers’ understanding of test building by gender or degree of experience in Ungogo Local Government Area (LGA), Kano State, Nigeria. The results, on the other hand, are consistent with those of Opataye (2016), who discovered that in a setting of Open and Distance Learning (ODL) higher education, the amount of difficulty lecturers have in creating test items for different kinds of exams varies dependent on gender.

CONCLUSION

Based on research results, it was found that female teachers in the district’s secondary schools had higher levels of multiple-choice construction competencies than male teachers, and that there is a low level of multiple-choice construction competence among public junior secondary school teachers in Edo Central Senatorial District, Nigeria.

RECOMMENDATIONS

Based on results, the following recommendations are made:

1. The district education administrators from the Local Government Education Authority Council has to put more focus on exposing teachers to more realistic methods of creating multiple-choice questions, particularly on how to manage item alternatives. This can be done by organizing a periodic workshops and specialized seminars on test construction skill acquisition for teachers in the district.
2. School authorities should strive to ensure and encourage instructors to create significantly more items when using multiple-choice questions to gauge student learning results. To cope with format and constructional faults, however, effective moderation and qualitative analysis of the test items should be used to guarantee that the items are properly built.

REFERENCES

1. Agustina, N. Q., & Mukhtaruddin, F. (2019). The CIPP model-based evaluation on Integrated English Learning (IEL) program at language center. *English Language Teaching Educational Journal*, 2(1), 22-31.
2. Aziz, S., Mahmood, M., & Rehman, Z. (2018). Implementation of CIPP Model for Quality Evaluation at School Level: A Case Study. *Journal of Education and Educational Development*, 5(1), 189-206.
3. CapanMelser, M., Steiner-Hofbauer, V., Lilaj, B., Agis, H., Knaus, A., & Holzinger, A. (2020). Knowledge, application and how about competence? Qualitative assessment of multiple-choice questions for dental students. *Medical education online*, 25(1), 1714199.
4. CapanMelser, M., Steiner-Hofbauer, V., Lilaj, B., Agis, H., Knaus, A., & Holzinger, A. (2020). Knowledge, application and how about competence? Qualitative assessment of multiple-choice questions for dental students. *Medical education online*, 25(1), 1714199.
5. Ibrahim, A., Ibrahim, A., & Amina, B. M. (2022). Assessment of test construction knowledge of senior secondary school teachers in Ungogo Local Government Area of Kano State, Nigeria. *Asian Basic and Applied Research Journal*, 4(5) 26-30.
6. Karimi, L. (2014). The effect of constructed-responses and multiple-choice tests on students' course content mastery. *Southern African Linguistics and Applied Language Studies*, 32(3), 365-372.
7. Kissi, P (2020). *Multiple-choice construction competencies and items? quality: evidence from selected senior high school subject teachers in Kwahu-South District* (Doctoral dissertation, University of Cape Coast). URI: <http://hdl.handle.net/123456789/4641>
8. Mc Coubrie, P. (2004). Improving the fairness of multiple-choice questions: a literature review. *Medical teacher*, 26(8), 709-712.
9. Noviani, Y. T., (2016). *Teacher-made multiple-choice test construction and the quality of the test items* (Doctoral dissertation, Universitas Pendidikan Indonesia). <https://www.semanticscholar.org/paper/TEACHER-MADE-MULTIPLE-CHOICE-TEST-CONSTRUCTION-AND-Noviani/e2b38976bf033034f0b7dc71d3e6dd963ae6421f> [Last assessed 20th March, 2023].
10. Opatye, J. A. (2016). Nigerian open and distance learning lecturers' difficulty in constructing students' test items. *Progressio*, 38 (1), 1-13.
11. Opie, O. N., Oko-Ngaji, V. A., Eduwem, J. D., & Nsor, J. A. (2021). An Assessment of Science teachers' utilization of the knowledge of test construction procedure in Multiple Choice Objective Tests in secondary schools in Yala LGA, Cross River State. *British Journal of Education*, 9(11), 54-62.
12. Opie, O. N., Oko-Ngaji, V. A., Eduwem, J. D., & Nsor, J. A. (2021). An Assessment of Science Teachers' Utilization of the Knowledge of Test Construction Procedure in Multiple Choice Objective Tests in Secondary Schools in Yala LGA, Cross River State. *British Journal of Education*, 9(11), 54-62.
13. 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.
14. Salihu, A. G. (2019). Assessing Teachers Ability on Test Construction and Economics Content Validity in Nasarawa State Senior Secondary Schools, Nigeria. *International Journal of Innovative Research in Education, Technology & Social Strategies*, 6(1).1-16.
15. Sopha, S., & Nanni, A. (2019). The cipp model: Applications in language program evaluation. *Journal of Asia TEFL*, 16(4), 1360.
16. Stufflebeam, D. L. (2004). The 21st century CIPP model. *Evaluation roots*, 245-66.
17. Van der Merwe, H. (2015). Quality assuring multiple-choice question assessment in higher education. *south African Journal of Higher education*, 29 (2), 279-297.