NIGERIAN JOURNAL OF COUNSELLING AND HUMAN DEVELOPMENT
ISSN: 2645-2774
Advancing Research. Improving Lives.

Investigating Bias in Standardized Testing: A Case Study of 2024 WAEC SSCE Economics Objective Test Items

By

Jackson Joseph NKEKI (Ph.D)
Department of Educational Evaluation and Counselling Psychology
Faculty of Education, University of Benin, Benin City
jacksonnkeki2@gmail.com; 07031638612, 08158733358
&
Amen Valentine UYIGUE (PhD)
Department of Educational Evaluation and Counselling Psychology
Faculty of Education, University of Benin, Benin City
valentin.uyigue@uniben.edu; 08056742588

Abstract
*The study aimed at investigating bias in standardized testing in the 2024 WAEC SSCE Economics objective test items. To achieve the objectives of this study, two research questions were raised and one hypothesis was tested. The Survey Research Design was adopted. Simple Random Sampling Procedure was used to select the sample for the study. The population of the study comprised 19,661 students in SS3 in Benin metropolis who wrote the 2025 WAEC SSCE Economics. The sample size for the study was one thousand nine hundred and sixty-six (*1,966*) SS3 students. The research instrument used for the study was 2024 WAEC SSCE Economics objective examination question paper. The Raju Area Index Analysis was used to detect items that are bias while the hypothesis was tested using Chi-square Statistical Analysis. The hypothesis was tested at 0.05 level of significance. The findings showed that out of the fifty (50) items in WAEC SSCE 2024 Economics examination question paper, thirteen items exhibited bias by sex. From the results of the hypothesis of the 2024 WAEC SSCE Economics, the null hypothesis of no significant difference in* items *exhibiting bias in 2024 WAEC SSCE Economics test items by sex was accepted. Based on the findings, it was recommended that* test *experts and developers should consider the use of Area Index Measure in investigating bias. Items that discriminate or unfair to examinees should be eliminated from the test before use. An item should be fair to both group of testees before use.*

Keywords: Item Bias, Objective Test Item, Area index Analysis, Standardized Examination Body & Economics

Introduction

Testing of students` ability and performance is a vital aspect of the education process. Testing in education is a systematic method to measure a test taker's knowledge, skills, and abilities, serving purposes like assessing learning, guiding instruction, ensuring accountability, and placing students, with key types including achievement, diagnostic, and placement tests, and benefits like improved memory and identifying learning gaps. Testing is defined as a process that yields a measurement of a student's abilities or characteristics through a series of items, typically in the form of questions or essays, where scores are calculated based on the

NIGERIAN JOURNAL OF COUNSELLING AND HUMAN DEVELOPMENT
ISSN: 2645-2774
Advancing Research. Improving Lives.

number of correct answers. It is considered a subset of measurement and can involve various scoring method including right or wrong assessments or more nuanced evaluations of responses. It is a specific form of measurement that uses a series of items or an essay (or essays) to ascertain where a testee is with regard to some characteristic of interest. Thus, a test typically consists of a set of items that are given to a testee. The testee attempts to get the right answers to the questions. Usually, a total score of the number correct is tallied and then either used directly or modified in some fashion to come up with a score. Sometimes the answers provided by the testees are given scores other than simply right or wrong such as a mark on an essay test.

Testing in the society today has become one of the most important parameters in detecting the standard of the educational system. Testing has consistently been an essential part of the educational system that all those who are not always present in school normally turn up to school during an examination period. The reason for testing is to show the latent ability of test takers which is based on the test takers` likelihood of success on a latent trait. This means that testing helps the tester to know whether test takers have learned what they were expected to learn or the degree to which test takers have learned the materials and to measure learning progress and achievement and to assess the effectiveness of educational programs (Nkeki & Orheruata, 2023).

Testing, as a programme of exercise, consists of many stages which include: Pre-determination of test purposes such as selecting a suitable test; administration of the tests to the testees; assigning scores to the test takers' responses; analzsing, interpretation of the scores and application of the result. Most modern societies have seen testing as the most acceptable method in decision making in schools, government organisations and even industries. Testing is used for teaching, placement, admission, promotion, recruitment, guidance, evaluation and research purpose among others (Emaikwu, 2011). Testing, according to Odili (2010), is the act of using a test to obtain data about a given attribute.

A test is a systematic tool or method, like a quiz or examinations, used to measure a test taker's knowledge, skills, aptitude, or characteristics by having them perform tasks or answer questions, providing data to evaluate learning, guide instruction, and make decisions about progress. Tests according to Nkeki and Ohiozua (2025) serve to determine achievement levels, identify strengths and weaknesses, and inform teaching strategies, often resulting in grades or

scores. Nkeki & Orheruata, (2023) stated that an important step in the construction of assessment tests is to ensure that no test taker or group answering the question or instrument is disadvantaged in anyway. Tests have various benefits.Taking a test permits students to assess what they know and what they do not know, so that they can concentrate study efforts on areas in which their knowledge is deficient. Tests can enhance learning during future study sessions. That is, when test takers take a test and then restudy materials, they learn more from the presentation than they would if they restudied without taking a test. Testing can facilitate retrieval of material that was not tested. Testing provides feedback to instructors and also, frequent testing encourages students to study harder.

At the school level, teachers administer tests to measure their students' understanding of specific content or the effective application of critical thinking skills. Such tests are used to evaluate student learning and academic achievements at the end of an instructional period. According to the Glossary for Educational Reform for Journalists, Parents and Community Members (2015) tests are used to determine whether students have learned what they were expected to learn or the degrees to which students have learned the material. This test conducted by the classroom teacher at the school level is refer to as teacher-made test.

Standardized testing refers to a method of evaluating student performance through uniform assessments that allow for comparison among individuals with similar characteristics, such as grade levels or courses taken. A standardized test is an examinations given and scored in a consistent, uniform way, using the same questions and conditions for all test-takers to allow for objective comparison of performance. These tests help evaluate individual student achievement, school effectiveness, and broader educational trends by comparing students against a common benchmark, often featuring multiple-choice formats. Standardized tests are designed in such a way that the questions and interpretations are consistent and are administered and scored in a predetermined, standard manner (Saul & Stephanie, 2024). Most standardized tests are summative assessments, that is, assessments that measure the learning of the participants at the end of an instructional unit). Because everyone gets the same test and the same grading system, standardized tests are often perceived as being fairer than non-standardized tests. Such tests are often thought of as more objective than a system in which some test takers get an easier test and others get a more difficult test. Standardized tests are

designed to permit reliable comparison of outcomes across all test takers because everyone is taking the same test and being graded the same way (Schaeffer et al, 2022).

Examination bodies such as West African Examinations Council (WAEC), National Examinations Council (NECO), National Business and Technical Examinations Board (NABTEB), Joint Admission and Matriculation Board (JAMB) and many others conduct standardized testing to assess the test takers. WAEC plays a pivotal role in evaluating the academic performance and preparedness of secondary school students for higher education and future endeavours (Mohammed & Pitan, 2022; Oguguo & Uboh, 2020). WAEC assess students with the use of Essay and Objective test questions. An essay test is an assessment technique that requires test takers to respond to a question by developing, organizing, and writing an original composition. The purpose of an essay test is to assess test takers' abilities to construct a logical, cohesive and persuasive writing piece. Essay tests provide a better indication of examinee's achievement in learning.

The response to the test items provides a clue to the nature and quality of the examinee's thought process. That is, assessing how the examinees present their ideas (whether their manner of presentation is coherent, logical and systematic) and how they conclude. Essay tests have various advantages such as how test takers can elaborate and provide detailed answers to a question; test takers are not able to guess and select an answer; test takers can review individualized responses from each examinee, and it can be used for all types of subjects and take less time to create questions. The problem associated with essay type of test is that it is difficult to score objectively and accurately; and, it is also difficult to evaluate objectively and partially because the answers are not fixed like the answers of objective items because of the variability in the tester`s judgement regarding the contents of the answers. These notwithstanding, Essay tests have proved to be incredible in testing. However, the Objective test items prove to be more objective in scoring.

Objectives tests are a type of assessment that measure a test taker's knowledge, skills, or abilities using predetermined items with specific correct answers. These tests are designed to be scored easily and reliably, typically through automated means, and are often used in educational settings, recruitment processes, and psychological assessments. Objective tests present respondents with a set of items, each with a fixed set of response options. The items are typically structured as multiple-choice, true or false, or matching formats, where there is a

NIGERIAN JOURNAL OF COUNSELLING AND HUMAN DEVELOPMENT
ISSN: 2645-2774
Advancing Research. Improving Lives.

clear correct answer for each item. Hope et al (2018) asserted that objective tests are assessments where there is one correct answer for each question, allowing for standardized scoring and interpretation. The scoring will also be objective because when the answers are fixed there will obviously be complete interpersonal agreement among the test takers. Test scores obtained from the objective tests are used to evaluate the competence of the examinees. Nkeki and Ohiozua (2025) asserted that objective tests can increase the testees opportunities of guessing the right answer to a question by eliminating unlikely choices. Objective tests generally are much more objective, because they are uniformly scored irrespective of the number of scorers. Objective questions usually have only one potential correct answer. Objective test includes matching, true/false, and multiple choices among others. In objective test, the answers are either right or wrong and require no interpretation or judgement on the part of the scorer as it is the case with subjective tests like essays.

Objective test items no matter how well they are constructed permit and encourage guessing by the examinee and the probability of guessing cannot be fully eliminated. The effect of the guessing is the increase of the actual score obtained on the test taker. Guessing is most obvious when the length of the test is short and the two alternative objectives type is used or when difficult alternative responses are included in multiple choice items or matching items and the length of the test is short. Among the important subjects evaluated in WAEC, Economics retains remarkable relevance as it equips test takers with vital knowledge and skills to comprehend economic principles and make informed decisions in various aspects of life (Fourcade, Ollion &Algan, 2015). It was observed that during the period between 2018 and 2022, there was an innovative shift in test takers` level of participation and their achievement of credit passes in WAEC Economics, raising questions and concerns among educators, policymakers, and stakeholders about the quality and effectiveness of the country`s educational system.

To ensure that economics achievement test is fair for all examinees, the instrument should be fair. Fairness of a test is one which ensures that all test takers with equal ability level should have the same opportunity to display the knowledge and skills that they have obtained and which are important to the test purpose. Roever (2005) opined that a fair test is one that allows every test taker the same chance to exhibit the knowledge and skills which they have gained and which are useful to the purpose of the test. A standardized test is said to be

NIGERIAN JOURNAL OF COUNSELLING AND HUMAN DEVELOPMENT
ISSN: 2645-2774
Advancing Research. Improving Lives.

unbiased when two groups of equal ability level test takers with respect to the construct measured by the test earn the same score on each item of the test. If there is difference between results of test takers with equal ability level, it gives indication of items that are bias for distinct groups of test takers. If the test is not fair or yield different scores from subgroups, the item is considered to be bias.

Bias in standardized testing refers to the presence of systematic errors or flaws in the testing process that can result in unfair or inaccurate results for certain groups of test takers. This can occur in various forms, including cultural bias, socio-economic bias, racial bias, gender bias and disability bias. Bias in testing can be intentional or unintentional, but its impact is always significant, leading to inequitable treatment and outcomes for marginalized groups. Bias in standardized testing refers to systematic unfairness where a test favours one group over another, often disadvantaging minority or low-income students due to cultural references, language, socioeconomic factors, or item design, rather than reflecting true ability, leading to unequal opportunities in education and career paths (Neill & Lisa, 2018). For example, a test may use language or references that are more familiar to one demographic group, leading to unequal opportunities for success. Standardized testing bias refers to the systematic unfairness in the design, implementation, and interpretation of standardized tests that can disadvantage certain groups of students, particularly those from marginalized backgrounds. This bias can perpetuate educational disparities and contribute to achievement gaps, as it fails to accurately measure the knowledge and skills of all test-takers equally, often reflecting cultural and socioeconomic differences rather than true academic ability.

A standardized test item is said to be bias when a dimension on the test is deemed to be irrelevant to the construct that is being measured, placing one group of the test taker at a disadvantage in taking the test (Nkeki & Orheruata, 2023). Consequently, if there is a distinction in the performance of groups of test takers with distinct ability level on a particular item then, there is no evidence that the item is bias; rather it is considered item impact. But, if a test item differs in parameters among distinct subgroups of equal ability level, this might be considered to be bias.

There are various ways of investigating or detecting bias in standardized test. Firstly, use statistical tools like Differential Item Functioning (DIF) to see if groups with the same ability answer a question differently, and Item Response Theory (IRT) to model item difficulty

across groups. Secondly, look for questions where specific student groups consistently underperform, even when overall achievement is similar, suggesting bias rather than just knowledge gaps. Thirdly, have test developers from varied backgrounds review questions for potential blind spots or culturally loaded content. Fourthly, ask students to explain their thinking on confusing questions to reveal interpretation issues. Lastly, examine "real-world" scenarios for hidden assumptions or unfamiliarity that might disadvantage some students.

Several demographic variables or attributes were assessed for standardized test bias. These variables are sex, ethnicity, socio-economic background, age, school location, race, school ownership and religion. The researcher`s focus variables for the study is sex. Sex according to researchers is an essential source of item bias. Adedoyin (2010) concentrated of the distinction between male sex and female sex in tests. According to him, male sex has greater numeric intuition than that of their female counterpart and that makes the males to perform better in geometrical items. According to the funding of Nkeki and Orheruata (2023) there is an existence of bias in NECO 2017 and 2018 Economics multiple test items with respect to sex.

There are several procedures of detecting test item bias. Some of these procedures are Logistic Regression Procedure, Mantel Haenszel, Item Characteristic Curve, Lord`s Chi-square Method, Transformed Item Difficulty Method, Likelihood Ratio Test, Purification Method, Simultaneous Item Bias (SIBTEST), Raju's Area Measures, Statistical Bias Test, Z-Test, One-Way Anova and Two-way Anova. Most of these procedures provide similar but not identical fact about item bias in standardized testing.This study focuses on Raju's Area Measures.

Raju's Area Index is a method in Item Response Theory (IRT) to quantify item bias by measuring the area between the Item Characteristic Curves (ICCs) of a focal group (e.g., females) and a reference group (e.g., males) across ability levels (θ). The formula involves integrating the absolute difference in item probabilities (or expected scores) over ability, using item parameters (discrimination 'a', difficulty 'b', guessing 'c' for 3PL) from IRT models, to get a single effect size value (e.g., > 0.22 indicates significant bias). Area Index **(A)**: Represents the total difference between the two groups' performance on an item, essentially the area between their ICCs.A larger area indicates greater bias. A common threshold for significance (e.g., for bias classification) is around 0.22. In essence, Raju's Area Index provides a quantitative measure (an effect size) of how much an item behaves differently for two groups,

NIGERIAN JOURNAL OF COUNSELLING AND HUMAN DEVELOPMENT
ISSN: 2645-2774
Advancing Research. Improving Lives.

independent of a specific ability level, by summing up those differences across the entire ability scale.

Formula (Simplified for 2PL Model):

$$A = \int_{-\infty}^{\infty} ((P_F(\Theta) - P_R(\Theta))d\Theta$$

- $P_F(\Theta)$ is the probability of a correct response for the focal group at ability θ
- $P_R(\Theta)$ is the probability of a correct response for the reference group at ability θ
- This integral is often calculated numerically or using specific formulas for IRT models (like 1PL, 2PL, 3PL).

Parameters (for 2PL Model) are;

a_1, b_1: Discrimination and difficulty for the **reference** group.

a_2, b_2: Discrimination and difficulty for the **focal** group.

$P(\Theta)$ (Logistic Function): $P(\Theta) = \frac{1}{1+e-a(\theta-b)}$ (for 2PL)

Statement of the Problem

Standardized testing plays a significant role in education systems worldwide, used by examination bodies like NECO, WAEC and NABTEB to evaluate test taker`s learning, inform instruction and make high-stakes decisions. However, concerns persist about potential biases embedded in these tests, which may unfairly impact certain student groups. Bias in standardized testing refers to systematic errors or distortions that affect test outcomes for specific groups, often based on demographics like race, ethnicity, socio-economic status, gender, or language background. Such biases can lead to inaccurate measures of students abilities, perpetuating achievement gaps and limiting opportunities. The presence of bias can stem from various factors such as questions may reflect dominant cultural norms or experiences, disadvantaging students from diverse backgrounds, test format, language or context may favour certain groups and testing environments or procedures might impact students differently. These may make certain groups to perform poorly due to biases, not lack of ability, students may be incorrectly labelled or placed in programs, test results may not accurately reflect students abilities, biased tests can increase anxiety and affect performance.

NIGERIAN JOURNAL OF COUNSELLING AND HUMAN DEVELOPMENT
ISSN: 2645-2774
Advancing Research. Improving Lives.

One of the important considerations in the selection and use of any test is that test must not be biased, that is test must be fair to all test takers. If the results of tests are to be used for decisions, then, the quality of test items has to be investigated so as to ensure that the test do not have traces of bias. That is, the items should be fair to all sub-groups in terms of sex, race and many others. Roever (2005) as cited in Perrone (2006) pointed out that a fair test is one being valid for all groups, individuals and society providing each test taker with an equal opportunity of demonstrating his or her skills and knowledge relevant to the purpose of the test. In other words, test takers with similar knowledge of material on a test (based on their total scores) must perform similarly on individual examination items irrespective of their sex or race, otherwise it is said to be bias. In other words, test items are considered to be bias if they contain sources of difficulty which is irrelevant to the construct measured. Thus, items containing sources of difficulty beyond those of interest which results in a discrimination against particular groups are regarded as bias. Therefore, this study seeks to empirically investigate bias in standardized testing of 2024 WAEC SSCE.

Purpose of the Study

The following is the purpose of the study:

1. To ascertain how many items in Economics objective test item administered by WAEC SSCE in 2024 exhibit bias between male and female.

Research Questions

The following research questions were raised to guide the study:

1. How many items in Economics objective test item administered by WAEC SSCE in 2024 exhibit bias between male and female?
2. Will there be a difference in the number of items exhibiting bias in 2024 WAEC SSCE Economics test items between male and female?

Research Hypothesis

The following research hypothesis was formulated in the study:

1. There is no significant difference in the number of items exhibiting bias in 2024 WAEC SSCE Economics test items between male and female.

Methodology

Survey Research Design was adopted in the study. The population comprised 19,661candidates that enrolled and sat for the WAEC Senior Secondary School Certificate

NIGERIAN JOURNAL OF COUNSELLING AND HUMAN DEVELOPMENT
ISSN: 2645-2774
Advancing Research. Improving Lives.

Examination in 2025 in Benin Metropolis. This population is the entire students who offered Economics in Benin Metropolis. The statistical population of the study is 50 Economics objective test items of WAEC SSCE. The sample size for this study comprised 1,966 SS3 students. Simple random sampling technique was used to select 983 male examinees and 983 female examinees from public and private schools representing 10% of the total population.

The instrument for this study was 2024 WAEC SSCE Economics objective question paper. The Economics objective question paper is made up of 50 items. The instrument is made up of four options (A-D). The instrument was standardized test conducted by West African Examination Council (WAEC). So, it is presumed to have undergone the process of validation. The reliability of the instruments was determined using test retest method and the reliability co-efficient of 0.73 was obtained from the 2024 WAEC Economics objective test item. The instrument was administered to the students by the senior Economics teachers of the selected schools visited under the supervision of the researcher. The instrument was administered under similar conditions as given by West African Examination Council (WAEC)

Raju Area Measure was used to analyze the data collected. An item is said to exhibit bias when the area index is greater than a critical value of 0.22, while an item does not exhibit bias when the area index is zero or close to zero (De Beer, 2004). According to Ling and Lau (2004), when the parameter *b* (item difficulty) for a particular group (for instance, Male) is greater than the other group (for instance, Female), it shows that such item is more difficult for the male group and then the item is considered to favour the female group vice versa, while the hypothesis was tested using the chi-square statistical analysis. The hypothesis was tested at 0.05 level of significance

Results

Research Question One: How many items in Economics objective test item administered by WAEC SSCE in 2024 exhibit bias between male and female?

Table 1: Summary of Area Index in investigating Bias between Male and Female of the 2024 WAEC Economics

Items	a1(Male)	b1(Male)	a2(Female)	b2(Female)	Area Index	Dec	Fav.Grp
1.	1.440	-2.503	0.662	-2.524	0.024	NB	
2.	0.636	-2.831	0.844	-2.205	0.016	NB	
3.	0.324	-0.882	0.430	-0.562	0.229	B	Male
4.	0.549	-2.901	1.634	-0.324	0.281	B	Male
5.	1.699	-1.482	1.725	-1.462	0.161	NB	
6.	1.222	-0.649	0.763	-0.921	0.040	NB	
7.	0.666	-1.524	0.981	-1.546	0.251	B	Female

NIGERIAN JOURNAL OF COUNSELLING AND HUMAN DEVELOPMENT
ISSN: 2645-2774
Advancing Research. Improving Lives.

8.	0.321	-0.486	0.630	-0831	0.061	NB	
9.	0.685	-1.201	1.606	-0.542	0.114	NB	
10.	0.330	-0.840	0.992	-0.489	0.295	B	Male
11.	0.634	-1.922	1.691	-1.911	1.121	B	Male
12.	0.650	-0.286	0.421	-0.418	0.029	NB	
13.	1.332	-0.486	0.231	-0.524	0.116	NB	
14.	0.63	-0.595	1.012	-0.811	0.201	NB	
15.	0.689	-0.444	0.321	-0.580	0.066	NB	
16.	0.246	-0.105	0.620	-0.820	0.342	B	Female
17.	1.265	-0.839	0.299	-0.469	0.110	NB	
18.	0.724	-0.633	1.211	-0.428	0.006	NB	
19.	1.010	-0.110	0.624	-0.812	0.049	NB	
20.	0.666	-1.483	0.727	-0.481	0.295	B	Male
21.	1.486	-0.612	0.851	-1.424	0.011	NB	
22.	0.633	-1.489	0.229	-0.848	0.201	NB	
23.	0.444	-0.380	1.224	-0.216	0.210	NB	
24.	0.393	-0.141	1.464	-0.244	0.092	NB	
25.	1.481	-0.632	0.096	-0.685	0.190	NB	
26.	1.301	-0.482	0.614	-1.888	0.062	NB	
27.	0.890	-0.962	0.889	-0.249	0.211	NB	
28.	1.48	-0.269	0.871	-0.352	0.094	NB	
29.	0.669	-0.831	0.244	-0.929	0.200	NB	
30.	1.632	-0.800	0.266	-0.789	0.299	B	Male
31.	1.096	-0.616	0.321	-0.240	0.091	NB	
32.	0.818	-0.046	1.329	-0.582	0.063	NB	
33.	0.990	-0.682	0.332	-0.643	0.124	NB	
34.	1.632	-0.240	1.690	-0.129	0.223	B	Male
35.	1.301	-1.244	0.639	-1.820	0.321	B	Female
36.	0.668	-0.342	0.320	-0.398	0.090	NB	
37.	0.221	-0.296	0.624	-0.432	0.061	NB	
38.	0.636	-0.491	0.532	-0.822	0.124	NB	
39.	0.631	-0.404	1.632	-0.821	0.200	NB	
40.	0.839	-1.422	1.362	-0.941	0.062	NB	
41.	0.636	-0.982	0.931	-0.824	0.532	B	Male
42.	0.221	-0.411	1.827	-0.493	0.298	B	Female
43.	1.999	-0.828	0.639	-0.419	0.022	NB	
44.	0.626	-0.330	1.414	-0.321	0.219	NB	
45.	0.888	-0.341	0.366	-0.848	0.019	NB	
46.	0.636	-0.244	0.119	-1.221	0.412	B	
47.	0222	-1.810	1.632	-0.441	0.098	NB	
48.	0.610	-0.204	0.633	-0.912	0.016	NB	
49.	1.932	-1.331	0.888	-1.236	0.151	NB	
50.	0.640	-0.389	0.424	-0.715	0.210	NB	

Table 1 shows the Summary of Area Index of 2024 WAEC Economics objective test. It exhibits the items that show bias and the group it favoured. The finding shows that thirteen (13) items such as items 3, 4, 7, 10, 11, 16, 20, 30, 34, 35, 41, 42, 46 out of fifty economics objective test items exhibit bias by sex.

NIGERIAN JOURNAL OF COUNSELLING AND HUMAN DEVELOPMENT
ISSN: 2645-2774
Advancing Research. Improving Lives.

Research Question One revealed that thirteen items representing 26% exhibit bias by sex with area indices greater than the critical value of 0.22 and thirty-seven items representing 74% do not exhibit bias with area indices less than 0.22. The findings displayed that out of the thirteen items that exhibit bias by sex, eight items representing 61.5% were in favour of male test takers while five items representing 38.5% were in favour of the female test takers. There is no significant difference in the number of items functioning differentially by sex in favour of males and those in favour of females in the 2024 WAEC Economics objective test items.

Hypothesis One: There is no significant difference in the number of items exhibiting bias in 2024 WAEC SSCE Economics test items between male and female.

Table 2: Chi-square (χ^2) Summary of test bias in favour of Male and Female students

Sex	N	df	Chi-square	Sig 2-tailed (P-value)
Male	8			
Female	5	1	1.61	.079
Total	13			
$\alpha = 0.05$				

Table 2 shows a chi-square value of 1.61 and a p-value of 0.079, testing the hypothesis at analpha level of 0.05, the p-value is greater than the alpha level, so the null hypothesis which states that 'there is no significant difference in the number of items exhibiting bias by sex in favour of males and those in favour of females in the 2024 WAEC Economics objective test is accepted.

Discussion of Results

The findings of research question one revealed that Out of 50 objective questions by sex in 2024 Economics, thirteen items (26%) exhibit bias by sex with alpha value of 0.05 greater than the P-value and thirty-seven (74%) do not exhibit bias with alpha value less than 0.05. Out of the thirteen items that exhibited bias in Economics 2024 objective test items by sex, eight items representing 61.5% were in favour of male test takers while five items representing 38.5% were in favour of the female test takers.The finding of this study agrees with the findings of Nkeki and Orheruata (2023) who in their finding revealed that Out of 60 multiple choice questions by sex in 2017 Economics, eight items (13%) functioned differentially by sex with alpha value of 0.05 greater than the P-value and fifty-two (87%) do

not function differentially with alpha value less than 0.05. Out of the eight items that functioned differentially in Economics 2017 multiple choice items by sex, six items (75%) were in favour of the male examinees while two items (15%) were in favour of the female examinees. The finding of this study also agreed with the findings of Adebule (2013) that out of the 40 items examined for the first factor program structure in Computer Science, only seven items representing 17.5% flagged DIF, in favour of male testees.

The finding of this study disagreed with the findings of Omorogiuwa and Iro-Aghedo (2016) that out of 50 items in NABTEB Mathematics multiple choice items functioned differentially for male and female testees, 17 items flagged DIF. 6 items favoured male examinees while 11 items favoured female examinees.The finding is also in disagreement with the report of Birjandi and Mohadeseh (2007) that in the general reading comprehension, seven out of the 13 DIF flagged items favoured females and 6 proved much easier for males. The finding of this study also disagreed with the finding of Nkeki and Ohiozua (2025) who in their finding one revealed that Out of 60 multiple choice questions by sex in 2022 Government, eleven items representing 18.3% shows bias by sex with cut-off value of >+1.0 or <-1.0 greater than alpha value of 0.05 and forty-nine items representing 81.7% do not display bias with alpha value less than cut-off value. Out of the eleven items that shows bias in Government 2022 multiple choice items by sex, four items (6.67%) were in favour of the male examinees while seven items (11.7%) were in favour of the female examinees.

Conclusion

The study concluded that there is an existence of bias in WAEC 2024 Economics objective test items with respect to sex.

Recommendations

Based on the findings of the study, the researcher made the following recommendations:

- Test experts and developers should consider the use of Area index measure in investigating bias.
- Examination bodies should ensure that examinations administered to examinees are bias free by ensuring that there is no evidence of item bias.
- Items that discriminate or unfair to examinees should be eliminated from the test before use.

NIGERIAN JOURNAL OF COUNSELLING AND HUMAN DEVELOPMENT
ISSN: 2645-2774
Advancing Research. Improving Lives.

References

Adebule, S.O. (2013). A study of differential item functioning in Ekiti State unified mathematics examination for senior secondary schools. *Journal of Education and Practice, 4(17), 43- 46*

Adedoyin, O. O. (2010). Investigating the invariance of person parameter estimates based on classical test and item response theories. *International Journal of Education Science, 2(2), 107-113.*

Birjandi, P., & Mohadeseh, M. (2007). Differential item functioning (test bias) analysis paradigm across manifest and latent examinee groups (on the construct of (IELTS). *Journal of Human Sciences,55,1-20.*

Emaikwu, S. O. (2011). Evaluation of student's ability in schools. *Being a paper presented at a workshop on teaching practice on Friday, 29th July in the College of Agricultural & Science Education, Federal University of Agriculture Makurdi, Benue State.*

Fourcade, M., Ollion, F & Algan, Y (2015) The Superiority of Economist. *Journal of Economic Perspectives29 (1)89–114*

Glossary of Educational Reform for Journalists, Parents and Community Members (2015). Passion in Education at the Heart of Teaching, *Great Schools Partnership*, Portland, ME 04101 207.773.0505

Hope, D., Adamson, K., Mcmanus, I.C, Chris, L. & Elder, A. (2018). Using Differential Item Functioning to Evaluate Potential Bias in a High-stakes Postgraduate Knowledge Based Assessment. *Journal of Education and Evaluation,18(1) 64-78*

Mohammed, A.O & Pitan, O.S. (2022), Innovation and Renewal of Economics Curriculum in Nigerian Secondary Schools: Approaches, Problems and Prospects. *Journal of Economics and Economic Education Research, (23) 1-14*

Neill, M. & Lisa, G. (2018)*Test Reform Victories Surge in 2017:* What's behind the Winning Strategies? Fair Test, 2017, www.Fairtest.Org/sites/default/files/Fair TestTestReformVictoriesReport2017.pdf.

Nkeki, J.J & Orheruata, M.U (2023):Determination of Differential Item Functioning by sex in the 2017 and 2018 NationalExamination Council SSCE Economics Multiple Choice Items using Logistic Regression Analysis: *UniAfrica Journal of Education. 2 (1). 44-62*

Nkeki, J.J . & Ohiozua E.N (2025) Assessment of Potential Item Bias in the 2022 NECO SSCE Government Objective Test Items Using Transformed Item Difficulty (TID) Procedure. *International Journal of Educational Foundations and Management. 19 (3) 155-168*

Odili, J.N. (2010) Reducing Differential Item Functioning through Language of Test Items: An Avenue for Promoting Peace Through Psychological Testing, *Readings in Conflict Management and Peace Building in Africa1, 341-357*

Oguguo, O.S.U. & Uboh, U.V. (2020), A State-Based Analysis of Candidates` WASSCE Participation and Achievement of five credit passes, including Mathematics and English Language, in Nigeria. *International Journal of Advanced Academic Research (Science, Technology and Engineering), 6(6) 30-44*

Omorogiuwa, K.O. & Iro-Aghedo, E.P. (2016). Determination of Differential Item Functioning by gender in the National Business and Technical Examination Board 2015 Mathematics multiple choice Examinations*;International journal of Education, Learning&Development,4(10).25-35.*

Perrone, M. (2006). Differential item functioning and item bias: Critical considerations in test fairness. Columbia University *Working Papers in TESOL & Applied Linguistics. 6 (2), 1-3.*

Roever, C (2005). "*That's not fair!" Fairness, bias, and differential item functioning in language testing.* Retrieved from http://www2.hawaii.edu/~roever/brownbag.pdf

Saul, L & Stephanie, N. (2024) "Yale to Require Standardized Test Scores for Admissions." ***The*** New York Times, 22 Feb. 2024, www.nytimes.Com /2024/02/22/us/yale standardized-testing-sat-act.html?

Schaeffer, B., Monty, N., and Lisa, G. (2022) "Resistance to High-Stakes Testing To Spreads." *District Administration, 48 (8) 34-42.*