



(RESEARCH ARTICLE)



A STUDY OF DIFFERENTIAL ITEM FUNCTIONING IN SOCIAL STUDIES 2025 MULTIPLE-CHOICE TEST ITEMS CONSTRUCTED BY MINISTRY OF EDUCATION EKITI STATE

Ademola Amos ADEGBOYE ^{1,*} and Olawale OMOYA ²

¹ Department of Social Science Education Bamidele Olumilua University of Education, Science and Technology Ikere-Ekiti, Nigeria.

² Department of Counseling Psychology, Bamidele Olumilua University of Education, Science and Technology, Ikere-Ekiti, Nigeria.

* Corresponding Author

World Journal of Advanced Research and Reviews, 2026, 32(01), 127–140

Publication history: Received on 25 August 2026; revised on 30 September 2026; accepted on 02 October 2026

Article DOI: <https://doi.org/10.30574/wjarr.2026.32.1.2451>

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

The study investigated Differential Item Functioning (DIF) Differential Item Functioning in the 2025 Ekiti State Ministry of Education Social Studies multiple-choice test. The study was guided by three research questions and three hypotheses. An ex-post facto research design was adopted for the study. The population comprised all Junior Secondary School III (JS III) students in the sixteen Local Government Areas of Ekiti State. A sample of 500 JS III students was selected using proportionate random sampling. The instruments used for data collection were the 2025 Junior Secondary School certificate Social Studies multiple-choice test developed by Ekiti State Ministry of Education and a Socio-Economic Status (SES) questionnaire. developed by the Researcher. The validity and reliability of the instruments were established using appropriate procedures Descriptive statistics were used to answer the research questions, while the independence chi-square test was used to test the hypotheses at the 0.05 level of significance. The findings revealed the presence of Differential Item Functioning based on gender, geographical location, socio-economic status and school type on Social Studies multiple-choice test developed by Ekiti State Ministry of Education. The study concluded that some test items did not function uniformly across the identified subgroups of examinees. It was therefore recommended that examination bodies should carefully examine test items for possible sources of bias associated with gender, location and socio-economic status, before such items are used for educational decisions. .

Keywords: Assessment, Differential Item Functioning, Gender, Location, Socio-Economic Status, Social Studies

1 INTRODUCTION

Assessment is an essential of process obtaining information that is used in making decision sit provides information about learners' knowledge, skills, abilities, and achievement. Educational test are commonly used to make important decisions concerning students' promotion, placement, selection, certification, and admission. For these decisions to be meaningful, test scores must accurately represent the abilities or characteristics that the test is intended to measure. An important requirement of a quality assessment is therefore fairness to all examinees, regardless of characteristics such as gender, location, socioeconomic background, or school type

Nigeria, several examination bodies conduct standardized examinations for different educational purposes. These examinations are administered to candidates who differ in terms of gender, geographical location and socio-economic background. Such diversity makes fairness in educational testing particularly important. When an examination item is more difficult for one group of examinees than another group, even when the two groups have comparable levels of the underlying ability being measured, the item may exhibit Differential Item Functioning (DIF). Emaikwa (2012) noted that candidates who participate in educational examinations come from diverse personal and environmental backgrounds, which may influence their experiences with assessment.

Differential Item Functioning refers to a situation in which examinees from different groups, but who possess comparable levels of the ability measured by a test, have different probabilities of responding correctly to a particular test item. In other words, DIF occurs when an item performs differently for two or more groups after their overall ability has been controlled. The concern with DIF is not simply that different groups obtain different scores; rather, the concern is that an item may advantage or disadvantage members of a particular group for reasons unrelated to the ability the test is intended to measure.

The issue of fairness is particularly important in a heterogeneous society such as Nigeria, where students participating in public examinations come from diverse social, cultural, economic, and educational backgrounds. The National Policy on Education emphasizes the provision of equal educational opportunities to Nigerian citizens irrespective of individual differences (Federal Government of Nigeria, 2004). Consequently, assessment practices should also promote fairness by ensuring that test items do not systematically favour or disadvantage examinees because of characteristics unrelated to the construct being measured.

The presence of DIF can constitute a serious threat to the validity and fairness of test scores. When an item favours one subgroup over another, differences in performance may reflect characteristics of the item rather than genuine differences in the ability of the examinees. This may result in inaccurate decisions concerning students' promotion, placement, selection, certification, or educational opportunities. Thus, identifying and eliminating problematic items is essential, particularly when test results are used to make important educational decisions.

Odili (2010) observed that interest in DIF analysis is closely related to the view of education as an instrument for promoting equality and social development. For educational assessment to contribute to this objective, test items should measure knowledge and skills that have been taught and should not require examinees to possess experiences or characteristics that are unrelated to the intended learning outcomes. A test item should therefore discriminate on the basis of the construct being measured rather than on the basis of gender, location, socio-economic status, or other irrelevant characteristics.

DIF can also have implications for the interpretation of students' academic performance. Where students are disadvantaged by test items because of characteristics unrelated to their actual ability, their test scores may underestimate their true competence. In high-stakes examinations, such outcomes may affect students' opportunities for further education. This makes DIF analysis an important quality-control procedure in the development, validation, and use of standardized educational tests.

Atar (2006) emphasized that test items should distinguish examinees according to their relevant abilities rather than according to characteristics such as gender, race, or ethnic background. A fair test should provide comparable opportunities for members of different groups to demonstrate the knowledge and skills relevant to the purpose of the assessment. The presence of a substantial number of DIF items may therefore threaten the validity of a test and the interpretations made from its scores.

Two major forms of DIF are commonly identified: uniform and non-uniform DIF. Uniform DIF occurs when the difference in the probability of a correct response between two groups remains relatively consistent across levels of ability. Non-uniform DIF occurs when the magnitude or direction of the difference varies across ability levels. In non-uniform DIF, the interaction between ability and group membership means that an item may favour one group at some ability levels and another group at different ability levels.

Several factors may contribute to DIF. Odili (2003) identified factors such as test-wiseness, culture, and language ability as possible sources of systematic error that may result in DIF. These factors are particularly relevant in a multicultural and multilingual educational environment such as Nigeria. Therefore, examination developers need to consider whether the language, context, content, or presentation of an item introduces an advantage or disadvantage for a particular subgroup.

1.1 Gender and Differential Item Functioning

Gender is one of the variables frequently examined in DIF research. Gender refers to the socially and culturally constructed roles, behaviours, expectations, and characteristics associated with males and females (Okoro, 2011; Ija, 2009). Differences in educational experiences and social expectations may influence how male and female students respond to particular test items. Previous studies have reported the presence of DIF associated with gender in educational tests. Abiam and Odok (2006), for example, reported that some test items exhibited differential functioning across gender groups. Such findings demonstrate the need to examine examination items to ensure that differences in performance are attributable to the intended construct rather than to gender-related characteristics.

1.2 Socio-Economic Status and Differential Item Functioning

Socio-economic status (SES) represents the social and economic position of individuals or families within a society. It is commonly associated with factors such as family income, parental educational attainment, parental occupation, and social status (Alordiah, 2015; Okafor, 2007). Students from different socio-economic backgrounds may have unequal access to educational resources, learning materials, private tutoring, technology, and other learning opportunities. Consequently, test items that rely unintentionally on experiences associated with a particular socio-economic background may function differently for students from high- and low-SES groups.

1.3 Geographical Location and Differential Item Functioning

The geographical location of a school may also influence students' educational experiences. Schools located in urban areas often have greater access to educational facilities, qualified teachers, transportation, communication networks, and other instructional resources than schools in some rural communities. Ndifon, Umoinyang, and Idiku (2013) noted that differences in educational environments may influence students' academic achievement. Such differences make location an important variable when investigating DIF. Previous research, including the work of Iyang (1991), has reported evidence of location-related DIF in standardized examinations.

1.4 Research Questions

The following research questions guided the study:

- What is the incidence of Differential Item Functioning between male and female students in 2025 Social Studies Multiple-Choice Test Items constructed by Ministry of Education Ekiti State?
- What is the incidence of Differential Item Functioning between urban and rural students in 2025 Social Studies Multiple-Choice Test Items constructed by Ministry of Education Ekiti State?
- What is the incidence of Differential Item Functioning between students with high and low socio-economic status in 2025 Social Studies Multiple-Choice Test Items constructed by Ministry of Education Ekiti State?

1.5 Research Hypotheses

The following null hypotheses were formulated and tested at the 0.05 level of significance:

- There is no significant difference between the number of items that function differentially for male and female students and the number of items that do not function differentially.
- There is no significant difference between the number of items that function differentially for urban and rural students and the number of items that do not function differentially.
- There is no significant difference between the number of items that function differentially for students with high and low socio-economic status and the number of items that do not function differentially.

2 METHOD

2.1 Research Design

The study adopted an ex-post facto research design. The design was considered appropriate because the variables of interest had already occurred and could not be manipulated by the researcher. The design enabled the researcher to examine the incidence of Differential Item Functioning in the 2025 Ekiti State Ministry of Education Social Studies multiple-choice test across different groups of examinees.

2.2 Population of the Study

The population of the study comprised all Junior Secondary School III (JSS III) students in the Sixteen Local Government Areas of Ekiti State. The population was considered appropriate because the students had completed or were expected to have covered the Social Studies curriculum relevant Junior Secondary School Certificate Examinations

There were 179 Junior Secondary Schools in Ekiti State with a total population of 22,467 JS III students at the time of the study, according to records obtained from the Ministry of Education, Ado Ekiti. Of this population, 13,481 students were from urban areas, while 8,986 were from rural areas.

2.3 Sample and Sampling Technique

A sample of 500 JSS III students was selected from the study population. Proportional sampling was used to ensure adequate representation of students from the relevant subgroups. The population was stratified according to geographical location, while students were randomly selected from public and private schools within the Local Government Ekiti State.

2.4 Instruments for Data Collection

Two instruments were used for data collection: the Junior Secondary School (JSS III) Social Studies multiple-choice test constructed by Ekiti State Ministry of Education and the Socio-Economic Status Questionnaire.

The 2025 Junior Secondary Schools Social Studies multiple-choice test consisted of 60 multiple-choice items covering the prescribed Social Studies syllabus for Junior Secondary Schools in Ekiti State. Each item had one correct answer and distractors.

The Socio-Economic Status Questionnaire consisted of 20 items structured on a four-point Likert scale of Strongly Agree (SA), Agree (A), Disagree (D), and Strongly Disagree (SD). The minimum possible score was 20, while the maximum possible score was 80. Students who obtained scores of 41 and above were classified as having high socio-economic status, whereas those who obtained scores of 40 and below were classified as having low socio-economic status.

2.5 Validity of the Instruments

The validity of the instruments was established through appropriate procedures. The 2025 Junior Secondary School (JSS) Social Studies multiple-choice test was examined to determine its coverage of the prescribed JS III Social Studies syllabus. The content validity of the Socio-Economic Status Questionnaire was established by ensuring that the items adequately represented the indicators used to classify students into high and low socio-economic status groups.

The questionnaire was further subjected to expert review. Their comments, observations, and recommendations were used to improve the clarity, relevance, and appropriateness of the items.

2.6 Reliability of the Instruments

The reliability of the Social Studies multiple-choice test was established by administering the test to 40 JS III students who were not included in the main study. Their responses were scored and analysed using the Kuder-Richardson Formula 20 (KR-20). A reliability coefficient of 0.82 was obtained, indicating satisfactory internal consistency.

The reliability of the Socio-Economic Status Questionnaire was determined using Cronbach's alpha coefficient. A reliability coefficient of 0.68 was obtained. This coefficient was considered adequate for the purpose of the study.

2.7 Data Collection Procedure

The Social Studies multiple-choice test was administered to the sampled students in their respective schools. Social Studies teachers in the selected schools assisted the researcher with the administration of the test. The examination was administered under conditions that were as similar as possible to those specified by the examination body.

Each correct response was scored one (1), while an incorrect response was scored zero (0). The SES questionnaire was administered alongside the Social Studies test to obtain information required for classifying students according to their socio-economic status.

2.8 Method of Data Analysis

The data collected were analysed using the Rasch measurement model with WINSTEPS statistical software and the Statistical Package for the Social Sciences (SPSS). Descriptive statistics were used to answer the five research questions by determining the incidence of Differential Item Functioning across the identified subgroups. The independence chi-square test was used to test the five null hypotheses at the 0.05 level of significance.

The Rasch model was employed to examine whether individual test items functioned differently for examinees belonging to different groups after taking their underlying ability levels into consideration. This approach provided a

basis for identifying items that demonstrated differential functioning according to gender, geographical location, socio-economic status,

3 RESULTS

3.1 Question One: What is the incidence of DIF for male and female students in Social Studies multiple-choice test

Table 1: Rasch Model DIF Statistics for Gender

Items	Bm	Bf	Δb	Probability	DIF Incidence	Favoured group
1	-2.78	-2.55	-0.22	0.57	1	F
2	-2.63	-2.01	-0.62	0.69	1	F
3	0.56	0.27	0.29	0.14	1	M
4	1.02	1.28	-0.26	0.20	1	F
5	0.78	1.26	-0.47	0.02	2	F
6	-1.92	-1.47	-0.45	0.11	1	F
7	-0.97	-0.52	-0.46	0.04	2	F
8	1.00	1.26	-0.26	0.20	1	F
9	0.88	0.32	0.55	0.01	2	M
10	-1.46	-1.87	0.41	0.14	1	M
11	-0.61	-0.38	-0.22	0.29	1	F
12	1.14	1.45	-0.32	0.13	1	F
13	-1.10	-1.37	0.27	0.26	1	M
14	0.81	0.81	0.00	1.00	1	M
15	0.41	-0.03	0.44	0.03	2	M
16	0.01	0.25	-0.23	0.24	1	F
17	0.82	1.15	-0.33	0.10	1	F
18	-0.72	-0.65	-0.06	0.78	1	F
19	0.05	-0.01	0.06	0.76	1	M
20	0.25	-0.21	0.47	0.02	2	M
21	-2.05	-1.92	-0.13	0.66	1	F
22	-1.84	-1.74	-0.10	0.72	1	F
23	-1.20	-1.58	0.38	0.14	1	M
24	1.12	0.67	0.44	0.03	2	M
25	2.24	1.92	0.32	0.19	1	M
26	0.81	0.81	0.00	1.00	1	M
27	-1.20	-1.74	0.54	0.04	2	M
28	1.30	1.41	-0.11	0.61	1	F
29	-0.16	-0.36	0.20	0.32	1	M
30	-1.43	-1.82	0.39	0.15	1	M
31	0.84	0.95	-0.11	0.59	1	F
32	1.15	1.15	0.00	1.00	1	M

33	-0.32	-0.07	-0.12	0.22	1	F
34	0.63	0.63	0.00	1.00	1	M
35	1.64	1.76	-0.12	0.58	1	F
36	0.13	0.21	-0.08	0.68	1	F
37	-1.02	-1.15	0.13	0.57	1	M
38	0.18	-0.01	0.19	0.33	1	M
39	1.52	1.52	0.00	1.00	1	M
40	-0.83	-0.78	-0.05	0.81	1	F
41	1.81	1.52	0.29	0.19	1	M
42	0.09	0.23	-0.14	0.48	1	F
43	-1.34	-0.83	-0.52	0.03	2	F
44	-1.15	-1.09	-0.06	0.81	1	F
45	-0.02	-0.02	0.00	1.00	1	M
46	0.94	0.94	0.00	1.00	1	M
47	0.09	0.21	-0.12	0.55	1	F
48	-0.74	-0.74	0.00	1.00	1	M
49	2.37	2.11	0.25	0.31	1	M
50	1.61	1.52	0.09	0.67	1	M
51	0.03	0.67	-0.64	0.00	2	F
52	1.66	2.05	-0.39	0.09	1	F
53	-0.47	-0.49	0.02	0.92	1	M
54	0.32	0.32	0.00	1.00	1	M
55	-1.02	-1.12	0.10	0.66	1	M
56	-1.18	-1.62	0.44	0.08	1	M
57	-0.16	-0.30	0.14	0.49	1	M
58	-0.44	-0.21	-0.22	0.27	1	F
59	0.54	0.36	0.17	0.37	1	M
60	-0.24	-0.47	0.23	0.25	1	M

Note: (b_m) represents the item measure for male examinees, (b_f) represents the item measure for female examinees, and (Δb) denotes the DIF contrast between the two groups.

Table 1 presents the Differential Item Functioning (DIF) statistics obtained using the Rasch model for each of the 60 Social Studies multiple-choice test items based on gender. An item was considered to exhibit DIF when its probability value was less than 0.05. The results indicate that 9 items exhibited DIF at the 0.05 level of significance. This represents 15% of the 2014 BECE Social Studies multiple-choice test items, indicating that these items functioned differently for male and female examinees. The items identified as exhibiting DIF were Items 5, 7, 9, 15, 20, 24, 27, 43, and 51.

Research Question Two: What is the incidence of Differential Item Functioning (DIF) between urban and rural students in the 2025 Social Studies Multiple-Choice Test Items constructed by Ministry of Education Ekiti State

Table 2 presents the Differential Item Functioning (DIF) statistics obtained using the Rasch model for each of the 60 Social Studies multiple-choice test items based on students' location. An item was considered to exhibit DIF when its probability value was less than 0.05. The results show that 24 items exhibited DIF at the 0.05 level of significance. This represents 40% of the 2014 BECE Social Studies multiple-choice test items, indicating that these items functioned differently for examinees from urban and rural locations. The items identified as exhibiting DIF were Items 3, 7, 16, 17, 20, 23, 24, 25, 26, 30, 31, 34, 36, 39, 41, 42, 45, 48, 52, 54, 56, 58, and 59.

Table 2: Rasch Model DIF statistics for Location

Item	Bu	Br	Δb	Probability	DIFincidence	Favouredgroup
1	-2.34	-2.84	0.50	0.20	1	U
2	-2.55	-2.22	-0.33	0.40	1	R
3	1.05	-0.04	1.09	0.00	2	U
4	1.07	1.21	-0.13	0.51	1	R
5	0.88	1.13	-0.25	0.22	1	R
6	-1.46	-1.82	0.36	0.21	1	U
7	-0.29	-1.02	0.72	0.00	2	U
8	1.12	1.12	0.00	1.00	1	U
9	0.79	0.47	0.32	0.10	1	U
10	-1.51	-1.70	0.19	0.19	1	U
11	-0.62	-0.43	-0.20	0.38	1	R
12	1.40	1.19	0.22	0.29	1	U
13	-1.36	-1.16	-0.20	0.44	1	R
14	0.79	0.81	-0.02	0.17	1	R
15	0.03	0.30	-0.28	0.17	1	R
16	0.66	-0.23	0.89	0.00	2	U
17	1.52	0.56	0.95	0.00	2	U
18	-0.59	0.74	0.15	0.52	1	U
19	-0.21	0.16	-0.37	0.07	1	U
20	-0.21	0.18	-0.39	0.06	2	R
21	-1.80	-2.07	0.27	0.38	1	U
22	-2.16	-1.65	-0.51	0.12	1	R
23	-1.80	-1.20	-0.60	0.04	2	R
24	1.12	0.71	0.40	0.04	2	U
25	1.84	2.39	-0.54	0.03	2	R
26	0.55	1.03	-0.49	0.02	2	R
27	-1.68	-1.33	-0.35	0.22	1	R
28	1.25	1.46	-0.21	0.32	1	R
29	-0.38	-0.19	-0.19	0.36	1	R
30	-1.22	-1.82	0.60	0.03	2	U
31	1.16	0.68	0.48	0.01	2	U
32	1.03	1.27	-0.24	0.25	1	R
33	-0.24	-0.17	-0.07	0.74	1	R
34	-0.03	1.13	-1.15	0.00	2	R
35	1.58	1.83	-0.25	0.27	1	R
36	0.46	-0.03	0.48	0.01	2	U
37	-1.27	-1.00	-0.27	0.29	1	R

38	-0.08	0.19	-0.27	0.19	1	R
39	1.82	1.25	0.57	0.01	2	U
40	-0.47	-1.00	0.53	0.18	1	U
41	1.94	1.39	0.55	0.01	2	U
42	-0.21	0.38	-0.59	0.00	2	R
43	-1.01	-1.12	0.10	0.67	1	U
44	-0.90	-1.24	0.34	0.16	1	U
45	-0.62	0.32	-0.94	0.00	2	R
46	0.88	1.00	-0.12	0.56	1	R
47	0.32	0.04	0.28	0.16	1	U
48	-1.05	-0.60	-0.46	0.06	2	R
49	2.32	2.13	0.19	0.46	1	U
50	1.61	1.53	0.08	0.72	1	U
51	0.55	0.21	0.34	0.86	1	U
52	2.66	1.19	1.47	0.00	2	U
53	-0.98	-2.23	-0.74	0.00	2	R
54	0.05	0.50	-0.45	0.03	2	R
55	-1.27	-0.98	-0.29	0.26	1	R
56	-2.01	-1.14	-0.87	0.01	2	R
57	-0.08	-0.31	0.24	0.25	1	U
58	-0.79	-0.09	-0.70	0.00	2	R
59	0.10	0.70	-0.60	0.00	2	R
60	-0.32	-0.35	0.02	0.91	1	U

Note:bu=measureof urban,br=measureofrural, Δb =DIFcontrast.

Research Question Three: What is the incidence of Differential Item Functioning (DIF) between students from high and low socio-economic status groups in 2025 Social Studies Multiple-Choice Test Items constructed by Ministry of Education Ekiti State

Table 3: Rasch Model DIF Statistics for Socio-economic Status

Items	Bh	Bl	Δb	Probability	DIF incidence	Favoured group
1	-2.52	-2.79	0.27	0.48	1	H
2	-2.16	-0.43	0.26	0.44	1	H
3	0.60	0.26	0.34	0.08	1	H
4	0.93	1.35	-0.42	0.04	2	L
5	0.85	1.16	-0.32	0.12	1	L
6	-1.65	-1.73	0.08	0.76	1	H
7	-0.77	-0.74	-0.02	0.92	1	L
8	1.16	1.09	0.08	0.70	1	H
9	0.66	0.56	0.10	0.59	1	H
10	-1.52	-1.73	0.21	0.44	1	H

11	-0.49	-0.49	0.00	1.00	1	H
12	1.06	1.52	-0.46	0.03	2	L
13	-1.26	-1.20	-0.06	0.80	1	L
14	0.81	0.81	0.00	1.00	1	H
15	0.27	0.14	0.13	0.52	1	H
16	0.41	-0.11	0.52	0.01	2	H
17	1.01	0.95	0.06	0.76	1	H
18	-0.71	-0.68	-0.03	0.90	1	L
19	0.05	0.00	0.05	0.79	1	H
20	-0.01	0.07	-0.08	0.68	1	L
21	-1.78	-2.14	0.36	0.23	1	H
22	-1.94	-1.70	-0.24	0.42	1	L
23	-1.40	-1.37	-0.03	0.89	1	L
24	1.30	0.56	0.74	0.00	2	H
25	2.00	2.14	-0.13	0.58	1	L
26	0.68	0.93	-0.25	0.20	1	L
27	-1.29	-1.55	0.25	0.32	1	H
28	1.25	1.45	-0.20	0.33	1	L
29	-0.31	-0.21	-0.09	0.64	1	L
30	-1.29	-1.87	0.58	0.03	2	H
31	1.04	0.77	0.27	0.18	1	H
32	1.23	1.09	0.14	0.48	1	H
33	-0.08	-0.29	0.21	0.31	1	H
34	0.27	0.95	-0.69	0.00	2	L
35	1.43	1.99	-0.56	0.01	2	L
36	0.35	0.02	0.33	0.09	1	H
37	-1.03	-1.12	0.10	0.68	1	H
38	0.14	0.05	0.09	0.66	1	H
39	1.52	1.52	0.00	1.00	1	H
40	-0.45	-1.10	0.65	0.00	2	H
41	1.62	1.71	-0.09	0.69	1	L
42	-0.04	0.31	-0.35	0.08	1	L
43	-1.00	-1.15	0.15	0.52	1	H
44	-0.94	-1.27	0.33	0.16	1	H
45	-0.19	0.12	-0.31	0.12	1	L
46	0.85	1.03	-0.18	0.36	1	L
47	0.39	-0.05	0.44	0.02	2	H
48	-0.97	-0.59	-0.38	0.09	1	L
49	2.18	2.30	-0.12	0.64	1	L
50	1.62	1.52	0.10	0.65	1	H

51	0.27	0.42	-0.15	0.44	1	L
52	1.98	1.73	0.25	0.28	1	H
53	-0.74	-0.29	-0.45	0.03	2	L
54	0.20	0.42	-0.21	0.28	1	L
55	-1.29	-0.92	-0.38	0.12	1	L
56	-1.44	-1.32	-0.12	0.64	1	L
57	-0.26	-0.20	-0.06	0.75	1	L
58	-0.33	-0.33	0.00	1.00	1	H
59	0.31	0.57	-0.27	0.18	1	L
60	-0.19	-0.47	0.28	0.17	1	H

Note: bh=measure of high SES, bl=measure of low SES, Δb=DIF contrast

Table 3 presents the Differential Item Functioning (DIF) statistics obtained using the Rasch model for each of the 60 Social Studies multiple-choice test items based on socio-economic status. An item was considered to exhibit DIF when its probability value was less than 0.05. The results indicate that 10 items demonstrated significant DIF at the 0.05 level of significance. This represents approximately 17% of the 2014 BECE Social Studies multiple-choice test items, indicating that these items functioned differently for examinees from high and low socio-economic status groups. The items identified as exhibiting DIF were Items 4, 12, 16, 24, 30, 35, 40, 47, and 53.

Hypothesis One: There is no significant difference between the number of test items that function differentially for male and female students and the number of items that do not exhibit Differential Item Functioning (DIF).

Table 4: Chi-square Statistics for Gender DIF

Gender	DIF-Gender		Total	X ²	Df	sig	Decision
	Non DIF Items	DIF Items					
Male	28	5	33				
Female	23	4	27	0.10	1	0.97	Not
Total	51	9	60				

significant

Table 4 shows a chi-square value of 0.10 and a P-value of 0.97. Testing at an alpha level of 0.05, the P-value is greater than the alpha level. Hence, the null hypothesis is accepted. Consequently, the number of items that function significantly differently for male and female students is not significantly different from the number that did not differentially function.

Hypothesis Two: The number of items that function significantly differently for urban and rural students is not significantly different from the number that did not differentially function.

Table 5: Chi-square Statistics for Location DIF

Location	DIF-Location		Total	X ²	Df	Sig	Decision
	NonDIFItems	DIFItems					
Urban	18	11	29				
Rural	18	13	31	0.19	1	0.75	Not
Total	36	24	60				

significant

Table 5 presents a chi-square value of 0.19 and a corresponding p-value of 0.75. At the 0.05 level of significance, the p-value (0.75) is greater than the alpha level (0.05). Therefore, the null hypothesis is not rejected. This indicates that there

is no statistically significant difference between the number of test items that functioned differentially for urban and rural students and the number of items that did not exhibit Differential Item Functioning (DIF). Thus, the observed difference in the number of DIF and non-DIF items across the two location groups is not statistically significant.

Hypothesis Three: There is no significant difference between the number of test items that function differentially for students from high and low socio-economic status groups and the number of items that do not exhibit differential functioning.

Table 6: Chi-square Statistics for SESDIF

SES	DIF-SES		Total	X^2	Df	Sig	Decision
	Non DIF Items	DIF Items					
High	27	5	32				
Low	23	5	28	0.15	1	0.82	Not
Total	50	10	60				

significant

Table 6 presents a chi-square value of 0.15 and a corresponding p-value of 0.82. At the 0.05 level of significance, the p-value (0.82) is greater than the stipulated alpha level (0.05). Therefore, the null hypothesis is not rejected. This implies that there is no statistically significant difference between the number of test items that functioned differentially for students from high and low socio-economic status groups and the number of items that did not exhibit differential functioning. Thus, the observed difference in the frequency of DIF items across the two socio-economic status groups is not statistically significant.

4 DISCUSSION OF FINDINGS

4.1 Incidence of Differential Item Functioning Based on Gender

The analysis of students' responses to the 2014 Basic Education Certificate Examination (BECE) Social Studies multiple-choice test revealed the presence of Differential Item Functioning (DIF) in nine items based on gender. This finding indicates that some of the test items did not function in the same manner for male and female examinees with comparable levels of ability. In other words, the probability of responding correctly to these items differed according to gender, suggesting that the items may have favoured one gender over the other. The finding is consistent with the study of Abdul-Aziz (2010), who reported the presence of gender-related DIF in Mathematics test items. Similarly, Odili (2003) found evidence of gender DIF in the WAEC/SSCE Biology Paper II examinations conducted in 1999, 2000, and 2001. The consistency of these findings suggests that gender-related DIF can occur across different subject areas and assessment contexts and therefore deserves careful consideration during test construction and validation.

4.2 Incidence of Differential Item Functioning Based on school Location

The analysis further revealed the presence of location-based Differential Item Functioning in 24 items of the 2014 BECE Social Studies multiple-choice test. This represents approximately 40% of the test items, indicating that a substantial proportion of the items functioned differently for examinees from urban and rural areas. The finding suggests that students' geographical or educational location may have influenced their performance on some of the test items, even when they possessed similar levels of the underlying ability being measured. Such differences may be associated with variations in educational facilities, availability of qualified teachers, learning resources, exposure to educational experiences, and other environmental factors between urban and rural schools.

This finding is consistent with the study of Odili (2003), which also reported evidence of location-related DIF in educational assessments. Similarly, Umoinyang (1991), in an analysis of Mathematics multiple-choice items used by the West African Examinations Council (WAEC) in the 1990 General Certificate Examination (GCE), identified 29 items that functioned differentially in favour of candidates from educationally advantaged areas. These findings underscore the importance of considering candidates' educational and geographical backgrounds when developing and evaluating examination items. Test items should be designed in such a way that differences in performance reflect students' knowledge and ability rather than disparities in their educational environments.

4.3 Incidence of Differential Item Functioning Based on Socio-Economic Status

The findings also showed the presence of Socio-Economic Status (SES) Differential Item Functioning in 17 items of the 2014 BECE Social Studies multiple-choice test. This accounts for approximately 17% of the items that functioned differently for examinees from high and low socio-economic backgrounds. The result suggests that students' socio-economic backgrounds may have influenced their chances of responding correctly to some test items, even when students possessed comparable levels of the ability being assessed.

The finding corroborates the results reported by Alordiah (2015), who identified evidence of SES-related DIF in educational assessment. It is also consistent with Odili (2003), who found evidence of socio-economic status differential item functioning in the WAEC/SSCE Biology Paper II examinations for 1999 and 2001. The occurrence of SES-related DIF may be associated with differences in students' access to educational materials, quality of learning environments, parental support, private tutoring, and other educational opportunities. Therefore, examination bodies should pay particular attention to the possibility that some test items may unintentionally favour students from economically advantaged backgrounds. Careful item review and DIF analysis can help ensure that examination items provide a fair and equitable assessment of students' knowledge and abilities.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Disclosure of conflict of interest

The authors declare no conflicts of interest regarding this manuscript.

REFERENCES

- [1] Adebalaziz, N. (2011). Detecting DIF using item characteristic curve approaches. *The International Journal of Education and Psychological Assessment*, 8(2), 1-15.
- [2] Adedoyin, O. O. (2010). Using IRT Approach to Detect Gender Based Items in Public Examination. *Education Research and Reviews Academic Journals*. 5(7), 385-399.
- [3] Alordiah, C. O. (2015). Comparison of index of Differential Item Functioning Under the methods of Item Response theory and classical test theory in Mathematics. An unpublished PhD thesis of Delta State University, Abraka, Delta State, Nigeria.
- [4] Amina, J. (1995). *Quality as Stress in Nigeria Education Kaduna*, Northern Nigeria Publishing Company.
- [5] Anastasia A. & Urbina S. (2006). *Psychological Testing*. New Delhi: Prentice Hall.
- [6] Arnkelsson, G. A. (2000). *Stability of Related Attributes and the Inference of Ability Through Social Comparison*. *Personality and Social Psychology Bulletin*, 26, 937-938.
- [7] Atar, B. (2006). Differential Item Functioning Analysis for mixed response data Using IRT-likelihood ratio test, logistic regression, and GLLAMN Procedures. Electronic thesis, treatises and dissertations (ETDs). Paper 248
- [8] <http://diginole.fsu.edu/etd/248>.
- [9] Augemberg, K. E.; & Morgan, D. L. (2008). Differential performance of test items by geographical regions. Paper presented at the annual meeting of the National Council on measurement in Education in New York, NY. Australian Bureau of Statistics Retrieved: 11-11-2016.
- [10] Baker, F. B. (2001). *The Basics Item Response Theory*. USA. ERIC clearing-house on assessment and evaluation.
- [11] Berk, T. (2007). *Item Bias Detection Methods for Small Samples*. Dissertation Abstract International.
- [12] Blanton, H. (2001). Evaluating the Self in the Context of Another: The Three Selves Model of Social Comparison Assimilation and Contrast. *Journal of Applied Social Psychology*. Bulletin 27 (12) 1628-1629.
- [13] Bolt, D. M. (2002). A Monte Carlo comparison of parametric and non-parametric polytomous DIF detection methods. *Applied psychological measurement*, 15, 133-141.
- [14] Braun, H., Jenkins, K. & Griggs, W. (2006). *Comparing Private Schools and Public Schools using Hierarchical Linear Modelling*. U.S Department of Education, Institute of Education Sciences, National Centre for Education Statistics, NCES 2006 461.
- [15] Camilli, G. & Shepherd, L. A. (1994). *Methods for Identifying Bias Test: Items Vol. 4*.

- [16] ThousandOaks,CA:SagePublication.
- [17] Chong, H. Y. (2010). A simple guide to the item response theory (IRT) and Rasch modelling. <http://www.creative-wisdom.com>.
- [18] Considine, G. & Zappala, G. (2002). Influence of Social Land Economics Disadvantaged inTheAcademicPerformance ofSchool StudentsinAustralia.Journalof Sociology, 38, 129-148. Retrieved 11-11-2016 from <http://www.jos.sage.pub>.
- [19] Cronbach,L.J.(1971).Testvalidation.PsychologicalBulletin52,281-302.
- [20] Egbule, J. F. (2002). Principles of Psychological Testing, Owerri. Whyte and Whyte Publishers.
- [21] Emaikwa, S. O.(2012).IssuesinTestItemBiasin PublicExaminationsinNigeriaand Implications for Testing. International Journal of Academic Research in Progressive Education and Development. 1(1), 175-185.
- [22] Embretson, S. & Reise, S. P. (2000). Item Response Theory for Psychologists. Mahwah Hew Jersey Lawrence Erlbaum Associates Publishers.
- [23] Emanovwe,A.(2009).TheImpactofGender,Socio-EconomicStatusandSchoolLocation on Students' Academic Achievement in Basic Technology. Unpublished M.Ed. Project, Delta State University, Abraka.
- [24] Federal Republic of Nigeria (2004). National policy on Education. Lagos. NERDC press.
- [25] Ferder,R.M.,Mohr,P.H.&Dietz,E.J(1994).ALongitudinalStudyofEngineering Student Performance andRetention.11Differences BetweenStudents from Rural and Urban Backgrounds. Journal of Engineering Education, 83 (3).
- [26] George, A. & Goethais, V. (2001). Peers Effects, Gender and Intellectual Performance amongStudentsataHighlySelectiveCollege.Comparison ofAbilityinAnalysis: Research in Higher Education. 26, 427-446.
- [27] Gieris,M.J.(1999).Testbias,SagePublications, Inc.,California.
- [28] Goux,D.&Maurin,E.(2008).PublicSchoolAvailabilityforTwoYearsoldandMothers' LabourSupply.AmericanEconomicAssociation.104(1)252-276.
- [30] Greenward, A. G. (2009). The Brief Implicit Applications and Data Analysis Methods (2nded) Thousand Oaks, CA. Saga Publication.
- [31] Hambleton, R. K., Swaminathan, H., & Rogers, J. H. (1995). Fundamentals of Item response theory. Newbury park, CA. Sage.
- [32] Ija, C. N. (2009). Towards promoting gender equality and women Empowerment for sustainabledevelopmentinRiverstate.JournalofInternationalGenderstudies,4,225-235.
- [33] Jorgensen, A. R. (2005). Cumulative Deficit in Compensatory Education. Journal of Psychology, 4, 37-47.
- [34] Joshua,M.T.(2005).Test/itemBiasinPsychologicalTesting:EvidencenonNigeriaSystem. A Paper Presented at the Annual Conference of Nigerian Association of Educational Psychologists held at Ahmadu University Zaria from 24th– 28th.
- [35] Kolcic,I.(2006).AcademicPerformanceandScientificInvolvementofFinalYearMedical Students Coming From Urban And Rural Background of Andjija Stamper School of Public Health, Medical School, University of Zagreb, Croatia <http://rrh.Deakia.rdu.au/>
- [36] Lord,F.M.(2003).ApplicationofItemResponseTheorytoPracticalTestingProblems.
- [37] Hilldale,NewJersey:ErbaumPress.
- [38] Lee, J. & Melritire, W. (2001). Interstate Variation in the Mathematics Achievement of Rural and Non-Rural Students. Journal of Research inRuralEducation, 16(3).
- [39] Malouff,J.M.(2008).BiasinGrading.CollegeTeaching,58(3),191-192.
- [40] Mercer, J. R. (1979). In Defence of Racially and Culturally Non-Discriminatory Assessment. School Psychology Digest, 8, 89-115.
- [41] Odili,J.N.(2003).Effectoflanguagemanipulationondifferentialitemfunctioningoftest items in Biology multiple choice test. An unpublished PhD thesis. University of Nigeria Nsukka.
- [42] Odili, J. N. (2010). Effect of language manipulation on differential item Functioning of test items in Biology in a multicultural setting. Journal of Educational assessment in Africa, 4,268-286

- [43] Ogbebor, U. C. (2012). Differential item Functionary Economics Question Paper of National Examination Council in Delta State Nigeria. Unpublished M.Ed. Thesis. University of Ibadan.
- [44] Owoeye, J. S. (2000). The Effect of Integration of Location Facilities and Class Size on Academic Achievement of Secondary School Students in Ekiti State, Nigeria PhD Thesis, Unpublished. University of Ibadan.
- [45] Pahike, E., Hyde, J. S & Allison, C. M. (2014). The Effect of Single Sex Compared with Coeducational Schooling on Students' Performance and Attitudes: A Meta-Analysis. *Psychological Bulletin*, 140 (4), 1042.
- [46] Park, J. H., Barhem, R. J. & Choi, J. (2012). Causal Effect of Single-Sex School on College Entrance Exams and College Attendance: Random Assignment in Seoul High Schools. PSC Working Paper Servers.
- [47] Perrone, M. (2006). Differential Item Functioning and Test Bias: Critical Consideration in Test Fairness. *Applied linguistics*.
- [48] Rauschemberg, S. W. (2012). *Hierarchical Linear Models: Applications and Analysis Methods* (2nd ed) Thousand Oaks, CA. Sage Publication.
- [49] Regner, I., Huguet, P., & Monteil, J. M. (2002). Effects of Socio-Economic Status (SES) Information on Cognitive Ability Inferences: When Low SES Students Make Use of a Self-Threatening Stereotype. *Social psychology of Education*, 5, 259-269.
- [50] Rudner, L. A. (2001). Item response theory. <http://echo.edres.org.8080/irt/baker/chapter/.pdf>.
- [51] Schemo, D. J. (2006). Most Students in Big Cities Lag Badly in Basic Science. *New York Times*, November 16 2006, Section A, 22.
- [52] Skiba, R. J., Knesting, K. & Bush, L. D. (2002). Culturally Competent Assessment: More than Nonbiased Tests. *Journal of Child and Family*, 11, 61-78.
- [53] Swaminathan, H. & Rogers, H. J. (1990). Detecting differential item Functioning using logistic regression procedures. *Journal of Educational measurement*, 27, 311-370.
- [54] Teresi, J. (2004). Differential item functioning and health assessment. Colombia University Stroud centre and faculty of medicine, New York State psychiatric institute, research division, Hebrew home for the aged at Riverdale.
- [55] Trembley, S., Ross, N. A. & Berthelot, J. M. (2001). Individual and Community Factors Affecting Grade Three Academic Performance, a Multi-Level Analysis. *Education Quarterly Review*, 7(4), 25-36.
- [56] Umoinyang, I. E. (2011). Method of detecting item functioning (DIF) as a consistent error in achievement test. *Journal of the Science Teachers Association of Nigeria (JSTAN)*.
- [57] University of Pennsylvania (2012). University Archives and Research Centre. University of Pennsylvania. Retrieved November 25, 2016.
- [58] U.S. Department of Education (2005). *Single Sex Versus Co-Educational Schooling: A System Review* Department of Education, Washington, D.C.
- [59] Vandevijver, F. & Tanzer, N. K. (2004). *Towards and Integrated Analysis of Bias in Cross-Cultural*