Asian Journal of Education and Social Studies

16(2): 42-55, 2021; Article no.AJESS.67114
ISSN: 2581-6268

Analysis of Socio-Economic Status and Gender Related Differential Item Functioning Using Item Response Theory Approach

Stella Eteng-Uket[1*]

[1]Department of Educational Psychology, Guidance & Counseling, Faculty of Education, University of Port Harcourt, Rivers State, Nigeria.

Author's contribution

The sole author designed, analyzed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/AJESS/2021/v16i230399
Editor(s):
(1) Dr. E. Seda Koc, Namık Kemal University, Turkey.
Reviewers:
(1) Jardel Coutinho dos Santos, Israel University, Ecuador.
(2) Ana Claudia Tenor, Brazil.
Complete Peer review History: http://www.sdiarticle4.com/review-history/67114

Original Research Article

Received 19 January 2021
Accepted 27 March 2021
Published 08 April 2021

ABSTRACT

The study investigated detecting differential item functioning using item response theory in West African Senior School Certificate English language test in south-south Nigeria. 2 research questions were formulated to guide the study. Using descriptive research survey design for the study, study population was 117845 Senior Secondary 3 students in Edo, Delta, Rivers and Bayelsa state. A sample of 1309 (604 males, 705 females) drawn through multi stage sampling technique was used for the study. Two valid instruments titled: Socio-economic status questionnaire (SSQ) and WASSCE/SSCE English language objective test (ELOT) were used to collect data for the study. The reliability indices of the instruments were estimated using the Cronbach Alpha method of internal consistency and Richard Kuderson 20 with coefficient values of .84 for the English Language objective test and .71 for the socio-economic status questionnaire respectively. Chi-square and Lord Wald test statistics statistical technique employed by Item Response Theory for Patient Reported Outcome (IRTPRO) was the technique used in data analysis which provided answers to the research questions at.05 level of significance. On analysis, the result revealed that 13 items functioned differently significant between the male and female group and significantly 23 items differentially functioned between High and low socio-economic status group. Thus, this shows 18% DIF based on gender and 32% based on socio-economic

**Corresponding author: Email: stella.eteng-uket@uniport.edu.ng;*

status indicating large DIF and items that are potentially biased. Based on the findings, recommendation were made and one among others was that Item Response theory should be used as DIF detection method by large scale public examination and test developers.

Keywords: Item response theory; differential item functioning; gender; socio-economic statu; English language; examination.

1. INTRODUCTION

Examination and testing are processes that are inextricably linked to education, and education is a powerful tool for man's and society's growth. According to Enamiroro [1], education aims to improve manpower and is oriented toward national growth and development. Most governments around the world have developed priorities and machinery to direct growth and development equally through education in the pursuit of national growth and development. The African government, and especially the Nigerian government, has not exempted itself from using this medium. In order to achieve this, the national curriculum offers a forum on which the interests of people from various socioeconomic backgrounds, genders, and ethnic groups can be catered for, resulting in equitable educational opportunities. There are also terminal exams administered by various examination bodies such as the West African Examination Council (WAEC), the National Examination Council, (NECO) and other bodies to which different examinees with the same skill or trait from various languages, cultures, genders, races, ethnic groups, geographical locations, cultural, and socioeconomic backgrounds are subjected.

These tests are supposed to assess what they are supposed to measure, so no things that are unequal to examinees from different classes should be included. As a consequence, it is important that the fairness principle is apparent in a test that is open to all and not biased against any group. Brown [2] described test bias as a situation in which examinees from one group are more likely to answer a question correctly (or support an item) than examinees from another group due to certain characteristics of the test item that are not relevant to the construct being measured. The word (DIF) was coined in the psychometric literature to describe questions about item bias in the sense of test bias [3]. The word (DIF) refers to the theoretical or empirical proof that is used to either refute or justify the presence of bias. Differential Item Functioning (DIF) is a breach of the invariance assumption in Item Response Theory (IRT) models that arise when the likelihood of endorsing an item for test takers of similar skill levels varies throughout classes [4]. After groups have been balanced for ability, DIF refers to differences in item functioning. After conditioning on the latent capacity that the item is built to calculate, there is an unexplained difference between two classes Wiberg [5]. As a result, such disparities can be sources of prejudice in the form of unfair advantages for certain groups and unfair disadvantages for others, and where there is DIF, conclusion derived from evaluation may be inaccurate, and an instrument is credible if only the data obtained do not disadvantage one group of examinees compared to another (Carroll, 2015); [6]. The scientific evidence used to support racism is referred to as DIF. DIF is “prima facie” proof that the test may be skewed [7]. As a consequence, DIF assists in the detection of potentially biased test products [8]. The fact that an item has been flagged or has proof of showing DIF does not imply that it is biased. For prejudice, DIF is necessary but insufficient. Bias is guaranteed if and only if the source of DIF is not included in the construct of test item focus [7]. The focal and reference groups are at least two groups in DIF; in general, DIF is analyzed by comparing item responses for two groups of examinees, commonly referred to as the reference and focal groups. By the measure, the focal group is the potentially disadvantaged group, while the referenced group is the potentially advantaged group; however, group designation is subjective [9,7].

Different methods for detecting DIF have been used over time, and all of these methods fall within the realm of the item response theory (IRT) approach [3,10,11,12]. IRT procedures have the best outcomes for identifying objects that are biased due to their greater complexity, as IRT provides a rigorous method for identifying DIF using object characteristics curve (ICC) [13]. According to Wiberg [4], if the item does not display DIF, the ICC will be the same, while when DIF is present, the ICC(s) for the two classes will vary. IRT detection methods include estimating or comparing object parameters (Lord Wald test), probability function, and Area

methods for detecting potentially biased items, as well as p-values, ICC curves, and trace lines, depending on the parameter model. According to empirical reports, the proportion of DIF products varies from a small 1.5 percent to a large percentage of the total (64 percent). Studies classify it as a small amount of DIF when a test contains less than 10% DIF, a medium amount of DIF when a test contains 10 to 30% DIF, and a significant amount of DIF when it reaches 30% when the parentage of DIF exceeds 10%. Xioting (2010), Hambleton & Roger [14]; Raju [15]; Hambleton & Roger [14]. Furthermore, DIF is typically measured in magnitude and varies in degree, which can be calculated by looking at parameters or statistics specific to the process used to detect DIF. DIF magnitude levels range from 0.25, 0.50, 0.75, and 1.00, with the highest DIF magnitude reflecting small to high DIF magnitude [16,17].

The English language is a subject matter in which any object can work in a variety of ways, just like any other subject. The importance of English language as one of the core subjects that will allow a person to gain admission into any higher institution is stipulated by the Government of Nigeria [18] in her National Policy on Education, making it a course that holds the key to further academic advancement. A credit pass in English language in the West African Senior School Certificate Examination (WASSCE) or its equivalent is necessary to study any discipline at the University, Polytechnic, or College of Education. It is important that English Language tests, regardless of who administers them, contain element that are equal, impartial, and do not discriminate against any group based on socioeconomic status, ethnic group, ethnicity, geographic position, gender, or other factors, on any subject matter.

The definition of socioeconomic status can be characterized as an individual's ranking based on educational attainment, wealth, and occupation. It is a construct that can be important or irrelevant to an individual's likelihood of endorsing an object in any achievement test in general, or in the English language in particular. Relevant when it pertains to the construct or trait under consideration, and irrelevant when it does not pertain to the trait under consideration. The socio-economic status of a learner is normally measured by the mean of a composite measure that takes into account the learner's income level as well as the parent's educational occupation; both of these variables are referred to as socio-economic status [19]. To categorize families, it is usually divided into three categories: high, middle, and low socioeconomic class. According to Ariani and Ghafourhia [20], socioeconomic status is a significant factor that can influence first and second language learning. Students in Nigeria are tested in English, which is a second language (L2). This is due to the fact that they already have a first language or mother tongue (L1) with which they begin their education.

Thus, what has been learned in (L1), which students come into contact with first from their families, whether they come from a family of high, poor, or middle socioeconomic status, can interfere with competence in (L2). According to studies, people with a high socioeconomic status are more likely to have a high level of English language competence than people with a low socioeconomic status, and therefore have a higher chance of correctly answering items on an English language test and other subjects [19,20,21]. Most Nigerian families, especially those from low socioeconomic status and those from rural areas, easily mix English (L2) with their mother tongue (L1) and pidgin English. As a result, most Nigerians, especially teenagers, are unable to fully express themselves without the use of words such as "abi," "shea," "nko," "na," and other gestures. All of these may pose more of a problem for peers from lower socioeconomic groups than for those from higher socioeconomic groups. [19,20,21].

Gender is a concept that can be important or irrelevant to the likelihood of correctly endorsing an object. It refers to the generally held beliefs and norms within a culture regarding acceptable male and female behavior, attributes, and roles.

Various studies have shown that the majority of tests used in certain public exams around the world, including those in Nigeria, include things that exhibit DIF. For instance, Madu [22] conducted an analysis on gender related differential item functioning in mathematics multiple choice items administered by WAEC, the study consisted of a sample of 1671 secondary students and Scheuneumen Modified Chi-square Statistics (SSX2) was used to detect 39 items that functioned differentially. Reuben & Akorede [23] also carried out a study titled differential item functioning technique for detecting item bias in economics among secondary school students in Abuja metropolis, using a sample of 750 through multi-stage sampling. They were able to identify in the NECO (SSCE) 2013 economics objective

test items, 3 items that functioned differentially based on socio-economic status, 5 items based on gender and 21 items exhibited DIF based on school location. Also Engelhard, Wind, Kobrin, & Chajewski [24] conducted a study on differential Item and person functioning in large-scale writing assessments within the context of the SAT. The findings of the research suggest that the SAT-10 did not appear to have any item subsets functioning in an unexpected way across subgroups of individuals (gender, race, ethnicity and best language subgroup). In a similar vein, Ogbebor & Onuka [25] used a sample size of 447 SS3 students to investigate differential item functioning system as an item bias indicator in Delta state, finding that the National Examination Council Economic questions for 2010 had 18 items that functioned differently for examines based on school form and school location. In addition, Umoinyang [26] discovered that there were items on the November/December 1990 WAEC mathematics objective test that functioned differently for test takers based on gender, educational development level, and area (northern and southern Nigeria). In the First School Leaving Certificate Examination performed by the Cross Rivers State Ministry of Education in 1992, Abiam [27] stated that some things worked differently in favor of some examinees. According to Uwhekadom [22], the chemistry multiple choice question used by WAEC in the 2009, 2010 and 2011 SSCE includes test items that work substantially differently for students of various genders, socioeconomic statuses, urban and rural geographical locations. DIF was also stated by Nworgu & Odili [28] in a WAEC biology multiple choice query. These inequalities led to candidates' poor performance on the Senior School Certificate Examination. Item functioning differs for individuals with similar skill from various subgroups of test takers, which has significant implications for policy, management, and classroom decision-making where test results serve as the foundation. A test with different functioning items can result in low achievement for a minority group in a subject matter, mutilating the meaning of the test result and the decision that is based on it for some groups, especially in core subjects such as English, which is a requirement for further educational advancement. This is a problem since the announcement of the Senior School Certificate Test is always greeted with public dismay (SSCE). This is due to the fact that there has been a persistent low performance in English Language exams over time, with 29.99 percent (2009), 23.36 percent (2010), 30.9 percent (2011), 38.81 percent (2012), 36.57 percent (2013), 31.28 percent (2014), and 38.68 percent (2015) obtaining five credit and above including English language in WASSCE in recent years from a total of 29.99 percent (2009), 23.36 percent (2010), 30.9 percent (2011), 38.81 percent (2012), 36.57 percent. One of the effects of this is that many students are excluded or unable to gain admission to a higher institution for further study, leaving some people unprepared to cope with the challenges that changes in globalization and technological development bring in today's world. In recent years, numerous criticisms have been leveled at examination bodies such as WAEC, alleging that items in their tests which unfairly benefit examines of certain ethnic groups, place, gender, language, school form, or even socioeconomic status because they are taken from examines from these groups, meaning that DIF and prejudice exist for test takers. It's likely that WAEC and other exam bodies didn't recognize the pedagogical implications of DIF as a possible contributor to low English language results, as shown by studies. The aim of the study was to investigate whether there are elements in the multiple choice test of English used by WEAC in the Senior School Certificate Examination that act differently for candidates of equal ability from different socioeconomic backgrounds and of different sexes, so that they are likely to be able to lead to poor performance in English Language.

The following research questions guided the study:

1. Which are the items in the English language multiple choice test that functioned significantly differently between the focal group (male) and reference group (female)?
2. Which are the items in the English language multiple choice test that functioned significantly differently between the focal group (low socioeconomic status) and reference group (high socioeconomic)?

2. METHODOLOGY

The descriptive survey research design was used for this study. The population of the study consisted of one hundred and seventeen thousand, eight hundred and forty-five (117,845) Senior Secondary three (3) students in 1190 public Secondary Schools who are studying English language as a certificate subject in the

2016/2017 academic session in Delta, Bayelsa, Rivers and Edo State. The sample size of the study was 1309 students of the population who are studying English language as a certificate subject. A multistage sampling technique was employed for the study at different stages and several sampling techniques like simple random, cluster and stratified, were employed.

The instrument for the study was a 15 item Socio-economic status questionnaire and the English Language Objective Test (ELOT) which was based on the WASSCE/SSCE English language paper one used in the 2016 examination. This contains 70 multiple choice type questions constructed by subject experts and developed by WAEC into test form. This instrument was employed to detect items that differentially function. The face and content validity of English language WAEC / SSCE used in 2016 has been established based on the fact that the questions were owned by an examining body WAEC/SSCE and has been validated by experts of the examining body through statistical techniques. The reliability of the coefficient of the instrument English Language Objective Test (ELOT) was established using the Kuder-Richardson (KR20) with internal consistency coefficient of 0.84 been the coefficient obtained. The instruments for the study were directly administered to the respondents on individual basis in their classes and were retrieved on the spot after they have been properly answered by the students.

Lord Wald Chi test which is used by IRTPRO software [29] was employed to detect items that function differentially between the focal and reference groups as seen by the P-value (tested at 0.05 sig level) for both the reference groups and focal groups and as well as their Chi Square ($\chi 2$) value.

3. RESULTS AND DISCUSSION

Research Question 1: Which are the items in the English language multiple choice test that functions differently between the focal group (male) and reference group (female)?

To answer this research question, the 70 items were subjected to DIF analysis using IRTPRO. Items that showed DIF were items that differ significantly between group membership and this is represented in Table 1.

Table 1 shows that items 3, 14, 15, 17, 22, 28, 29, 33, 51,60, 63, 66, and 67 shows DIF as seen from their Lord Wald $\chi 2$ value. Item 3, Wald $\chi 2$ (2) = 23.5, item 14, $\chi 2$ (2) = 6.2, item 15, $\chi 2$ (2) =14.1, item 17, $\chi 2$ (2) = 10.2, item 22, $\chi 2$ (2) = 6.7, item 28, $\chi 2$ (2) = 9.0, item 29, $\chi 2$ (2) =7.3, item 33, $\chi 2$ (2) = 6.5, item 51, $\chi 2$ (2) =8.9, item 60, $\chi 2$ (2) =7.2, item 63, $\chi 2$ (2) = 6.8, item 66, $\chi 2$ (2) =11.4 and item 67, $\chi 2$ (2) = 7.0 has the above as their $\chi 2$ value.

The result in Table 1 further reveals the p-values for the Wald $\chi 2$ statistics that tests the difference between reference (female) and focal (male) group item parameters (a* & b) revealing items that significantly functions differentially at 0.05 level of significance. From the table, it can be seen that Item 3 has p =.0001, p< 0.05, item 14 has p = .0400, p, < 0.05, item 15 has p = .0009, p< 0.05, Item 17 has p =.0006, p< 0.05 Item 22 has p =.0344, Item 28 has p =.0112, p< 0.05 Item 29 has p =.0263, p< 0.05, Item 33 has p =.0388, p< 0.05, Item 51 has p =.0117, Item 60 has p =.0277, p< 0.05, Item 63 has p =.0334, p< 0.05, Item 66 has p =.0028, p< 0.05 and Item 67 has p =.0296, p< 0.05. All showing P< 0.05 revealing further that these items significantly differentially functions between the focal group male and the reference group female in the English Language achievement test.

Research Question 2: Which are the items in the English language multiple choice test that functions differently between the focal group (low socio-economic status) and reference group (high socio-economic status) of students?

To answer this research question, the 70 items were subjected to DIF analysis using IRTPRO. Items that exhibited DIF were items that differentiate significantly between group membership and this was represented in Table 2.

From Table 2, it can be seen that items item 3, Wald $\chi 2$ (2) = 9.4, item 4, $\chi 2$ =15.1, item 7, $\chi 2$ = 21.7, item 9, $\chi 2$ =14.2, item 10, $\chi 2$ = 9.0, item 11, $\chi 2$ =6.9, item 12, $\chi 2$ = 7.5, item 16 $\chi 2$ =13.4, item 19, $\chi 2$ =7.5, item 23, $\chi 2$ = 10.5, item 30, $\chi 2$ = 15.3, item 35, $\chi 2$ = 13.6, item 37, $\chi 2$ =6.2, item 40, $\chi 2$ = 7.2, item 42, $\chi 2$ =7.9, item 47, $\chi 2$ =9.2, item 50, $\chi 2$ = 12.6, item 51, $\chi 2$ = 7.7, item 52, $\chi 2$ = 6.0, item 54, $\chi 2$ = 15.2, item 58, $\chi 2$ =6.6, item 62, $\chi 2$ = 8.1, and item 64, $\chi 2$ = 12.3 shows DIF as seen from their Lord Wald $\chi 2$ value.

The result in Table 2 further reveals the p-values for the Wald $\chi 2$ statistic revealing `items that significantly functions differentially at 0.05 level

Table 1. DIF statistics for gender

Item Numbers in:										
Group 1	**Group 2**	**Total X^2**	***d.f.***	***P***	X^2_a	***d.f.***	***P***	$X^2_{c\|a}$	***d.f.***	***P***
1	1	0.6	2	0.7447	0.0	1	0.9705	0.6	1	0.4434
2	2	4.5	2	0.1043	3.1	1	0.0761	1.4	1	0.2420
3	3	23.5	2	0.0001*	0.1	1	0.7268	23.3	1	0.0001
4	4	0.5	2	0.7919	0.3	1	0.5818	0.2	1	0.6865
5	5	3.0	2	0.2208	0.3	1	0.6132	2.8	1	0.0967
6	6	2.0	2	0.3698	1.0	1	0.3237	1.0	1	0.3139
7	7	5.6	2	0.0593	2.3	1	0.1277	3.3	1	0.0685
8	8	1.3	2	0.5299	0.2	1	0.6930	1.1	1	0.2915
9	9	1.6	2	0.4465	1.6	1	0.2055	0.0	1	0.9329
10	10	0.0	2	0.9885	0.0	1	0.9945	0.0	1	0.8793
11	11	5.4	2	0.0670	1.3	1	0.2515	4.1	1	0.0433
12	12	0.0	2	0.9801	0.0	1	0.8798	0.0	1	0.8955
13	13	1.4	2	0.4929	1.4	1	0.2384	0.0	1	0.8805
14	14	6.2	2	0.0450*	2.4	1	0.1221	3.8	1	0.0512
15	15	14.1	2	0.0009*	9.3	1	0.0023	4.8	1	0.0283
16	16	0.0	2	0.9839	0.0	1	0.9137	0.0	1	0.8856
17	17	10.2	2	0.0062*	4.7	1	0.0309	5.5	1	0.0190
18	18	1.5	2	0.4716	0.9	1	0.3332	0.6	1	0.4520
19	19	5.6	2	0.0607	5.6	1	0.0181	0.0	1	0.8995
20	20	3.2	2	0.1989	3.0	1	0.0864	0.3	1	0.5959
21	21	3.1	2	0.2157	3.1	1	0.0799	0.0	1	0.9112
22	22	6.7	2	0.0344*	6.3	1	0.0122	0.5	1	0.5012
23	23	4.8	2	0.0900	0.0	1	0.8301	4.8	1	0.0290
24	24	0.5	2	0.7673	0.3	1	0.5637	0.2	1	0.6580
25	25	3.9	2	0.1466	0.5	1	0.4742	3.3	1	0.0674
26	26	3.1	2	0.2159	2.3	1	0.1299	0.8	1	0.3807
27	27	1.8	2	0.4165	0.8	1	0.3677	0.9	1	0.3325
28	28	9.0	2	0.0112*	4.0	1	0.0449	5.0	1	0.0258
29	29	7.3	2	0.0263*	6.5	1	0.0110	0.8	1	0.3705
30	30	2.8	2	0.2448	2.5	1	0.1155	0.3	1	0.5648
31	31	1.7	2	0.4277	0.2	1	0.6465	1.5	1	0.2228

Item Numbers in:										
Group 1	Group 2	Total X^2	*d.f.*	*P*	X^2_a	*d.f.*	*P*	$X^2_{c\|a}$	*d.f.*	*P*
32	32	0.7	2	0.7083	0.1	1	0.7375	0.6	1	0.4476
33	33	6.5	2	0.0388*	5.3	1	0.0212	1.2	1	0.2761
34	34	4.1	2	0.1298	0.1	1	0.7696	4.0	1	0.0456
35	35	1.9	2	0.3825	1.9	1	0.1671	0.0	1	0.9220
36	36	2.1	2	0.3556	1.2	1	0.2666	0.8	1	0.3615
37	37	1.2	2	0.5395	1.0	1	0.3114	0.2	1	0.6482
38	38	0.9	2	0.6542	0.0	1	0.9772	0.8	1	0.3574
39	39	2.1	2	0.3489	2.1	1	0.1502	0.0	1	0.8592
40	40	0.7	2	0.7194	0.6	1	0.4225	0.0	1	0.9039
41	41	1.8	2	0.3991	0.1	1	0.7507	1.7	1	0.1880
42	42	1.4	2	0.5026	0.0	1	0.9670	1.4	1	0.2414
43	43	0.2	2	0.8924	0.2	1	0.6379	0.0	1	0.9388
44	44	1.0	2	0.6219	0.1	1	0.7489	0.8	1	0.3576
45	45	3.2	2	0.2064	3.1	1	0.0769	0.0	1	0.8452
46	46	3.5	2	0.1730	2.9	1	0.0882	0.6	1	0.4403
47	47	0.7	2	0.7085	0.1	1	0.7766	0.6	1	0.4355
48	48	1.3	2	0.5105	0.1	1	0.7685	1.3	1	0.2623
49	49	1.0	2	0.5947	0.8	1	0.3659	0.2	1	0.6382
50	50	5.6	2	0.0612	1.3	1	0.2555	4.3	1	0.0384
51	51	8.9	2	0.0117*	6.7	1	0.0096	2.2	1	0.1398
52	52	4.6	2	0.0981	4.1	1	0.0433	0.6	1	0.4555
53	53	0.7	2	0.7019	0.0	1	0.9545	0.7	1	0.4015
54	54	0.9	2	0.6313	0.1	1	0.7909	0.9	1	0.3569
55	55	4.1	2	0.1312	1.9	1	0.1700	2.2	1	0.1418
56	56	1.4	2	0.5018	0.4	1	0.5127	1.0	1	0.3299
57	57	1.9	2	0.3843	0.5	1	0.4867	1.4	1	0.2322
58	58	0.8	2	0.6695	0.7	1	0.4165	0.1	1	0.7067
59	59	3.8	2	0.1519	2.1	1	0.1471	1.7	1	0.1967
60	60	7.2	2	0.0277*	3.1	1	0.0775	4.1	1	0.0441
61	61	3.8	2	0.1526	3.4	1	0.0657	0.4	1	0.5332
62	62	0.0	2	0.9802	0.0	1	0.8441	0.0	1	0.9709
63	63	6.8	2	0.0334*	2.7	1	0.1016	4.1	1	0.0428
64	64	3.8	2	0.1510	3.7	1	0.0532	0.1	1	0.8041

Item Numbers in:										
Group 1	**Group 2**	**Total X^2**	***d.f.***	***P***	X^2_a	***d.f.***	***P***	$X^2_{c\|a}$	***d.f.***	***P***
65	65	4.8	2	0.0909	1.2	1	0.2758	3.6	1	0.0576
66	66	11.7	2	0.0028*	4.3	1	0.0376	7.4	1	0.0064
67	67	7.0	2	0.0296*	0.8	1	0.3723	6.2	1	0.0125
68	68	2.2	2	0.3257	2.2	1	0.1383	0.0	1	0.8362
69	69	0.3	2	0.8570	0.1	1	0.7524	0.2	1	0.6477
70	70	1.3	2	0.5213	0.1	1	0.7737	1.2	1	0.2696

*Asterisks * shows DIF items. Critical χ2 value = 5.99 at df (2) at 0.05 sig level*

Table 2. Showing DIF Statistics for Items for HSES and LSES

Item numbers in:										
Group 1(HSES)	**Group 2(LSES)**	**Total X^2**	***d.f.***	***P***	X^2_a	***d.f.***	***P***	$X^2_{c\|a}$	***d.f.***	***P***
1	1	4.1	2	0.1317	3.6	1	0.0586	0.5	1	0.4898
2	2	3.2	2	0.1999	0.4	1	0.5070	2.8	1	0.0958
3	3	9.4	2	0.0089*	0.1	1	0.7615	9.3	1	0.0022
4	4	15.1	2	0.0005	14.7	1	0.0001	0.3	1	0.5694
5	5	1.3	2	0.5184	0.0	1	0.9311	1.3	1	0.2533
6	6	9.4	2	0.0089*	0.1	1	0.7615	9.3	1	0.0022
7	7	21.7	2	0.0001*	13.5	1	0.0002	8.1	1	0.0044
8	8	0.9	2	0.6481	0.7	1	0.4178	0.2	1	0.6467
9	9	14.2	2	0.0008*	8.9	1	0.0028	5.3	1	0.0212
10	10	9.0	2	0.0111*	8.9	1	0.0029	0.1	1	0.7382
11	11	6.9	2	0.0315*	3.8	1	0.0510	3.1	1	0.0779
12	12	7.5	2	0.0234*	0.9	1	0.3378	6.6	1	0.0103
13	13	11.9	2	0.0026	11.3	1	0.0008	0.6	1	0.4245
14	14	5.2	2	0.0750	2.3	1	0.1318	2.9	1	0.0898
15	15	5.7	2	0.0580	3.6	1	0.0570	2.1	1	0.1508
16	16	13.4	2	0.0012*	13.4	1	0.0003	0.0	1	0.8352
17	17	2.3	2	0.3114	2.3	1	0.1271	0.0	1	0.9999
18	18	0.2	2	0.9107	0.1	1	0.7976	0.1	1	0.7277
19	19	7.5	2	0.0239*	4.2	1	0.0415	3.3	1	0.0689
20	20	3.1	2	0.2141	2.9	1	0.0891	0.2	1	0.6685
21	21	4.0	2	0.1388	3.9	1	0.0480	0.1	1	0.8068

Item numbers in: Group 1(HSES)	Group 2(LSES)	Total X^2	*d.f.*	*P*	X^2_a	*d.f.*	*P*	$X^2_{c|a}$	*d.f.*	*P*
22	22	2.4	2	0.2979	2.3	1	0.1283	0.1	1	0.7471
23	23	10.5	2	0.0052*	5.5	1	0.0190	5.0	1	0.0251
24	24	0.0	2	0.9858	0.0	1	0.8677	0.0	1	0.9778
25	25	1.9	2	0.3947	0.0	1	0.9111	1.9	1	0.1746
26	26	3.1	2	0.2121	0.5	1	0.4971	2.6	1	0.1045
27	27	0.0	2	0.9859	0.0	1	0.8680	0.0	1	0.9780
28	28	4.3	2	0.1144	3.4	1	0.0663	1.0	1	0.3276
29	29	0.4	2	0.8070	0.1	1	0.7837	0.4	1	0.5523
30	30	15.3	2	0.0005*	15.3	1	0.0001	0.0	1	0.9146
31	31	5.7	2	0.0580	0.0	1	0.8464	5.7	1	0.0174
32	32	0.5	2	0.7827	0.3	1	0.5795	0.2	1	0.6692
33	33	0.6	2	0.7281	0.5	1	0.4853	0.1	1	0.7011
34	34	1.0	2	0.6132	0.5	1	0.4778	0.5	1	0.4913
35	35	13.6	2	0.0011*	13.1	1	0.0003	0.5	1	0.4996
36	36	2.3	2	0.3185	2.3	1	0.1331	0.0	1	0.8683
37	37	6.2	2	0.0441*	3.4	1	0.0646	2.8	1	0.0938
38	38	1.2	2	0.5386	0.4	1	0.5086	0.8	1	0.3712
39	39	1.1	2	0.5701	0.1	1	0.7326	1.0	1	0.3159
40	40	7.2	2	0.0269*	2.8	1	0.0959	4.4	1	0.0351
41	41	2.1	2	0.3502	1.8	1	0.1791	0.3	1	0.5894
42	42	7.9	2	0.0190*	7.4	1	0.0065	0.5	1	0.4755
43	43	1.4	2	0.5035	1.4	1	0.2427	0.0	1	0.9420
44	44	3.4	2	0.1841	0.5	1	0.4849	2.9	1	0.0891
45	45	0.3	2	0.8772	0.0	1	0.8484	0.2	1	0.6350
46	46	0.5	2	0.7919	0.3	1	0.5671	0.1	1	0.7097
47	47	9.2	2	0.0101*	7.1	1	0.0077	2.1	1	0.1488
48	48	5.3	2	0.0700	0.0	1	0.8807	5.3	1	0.0214
49	49	1.6	2	0.4531	1.6	1	0.2103	0.0	1	0.9150
50	50	12.6	2	0.0018*	7.1	1	0.0076	5.5	1	0.0190
51	51	7.7	2	0.0216*	2.7	1	0.1011	5.0	1	0.0258
52	52	6.0	2	0.0508*	0.1	1	0.7294	5.8	1	0.0157
53	53	4.4	2	0.1106	4.3	1	0.0387	0.1	1	0.7209
54	54	15.2	2	0.0005*	13.7	1	0.0002	1.5	1	0.2199

Item numbers in: Group 1(HSES)	Group 2(LSES)	Total X^2	*d.f.*	*P*	X^2_a	*d.f.*	*P*	$X^2_{c\|a}$	*d.f.*	*P*
55	55	4.1	2	0.1289	0.0	1	0.8284	4.0	1	0.0442
56	56	1.1	2	0.5908	0.2	1	0.6684	0.9	1	0.3515
57	57	1.1	2	0.5832	1.0	1	0.3126	0.1	1	0.8102
58	58	6.6	2	0.0373*	6.1	1	0.0134	0.5	1	0.4991
59	59	0.0	2	0.9915	0.0	1	0.9057	0.0	1	0.9562
60	60	3.2	2	0.2038	3.1	1	0.0789	0.1	1	0.7448
61	61	1.1	2	0.5850	0.3	1	0.6173	0.8	1	0.3647
62	62	8.1	2	0.0170*	7.0	1	0.0080	1.1	1	0.2904
63	63	0.1	2	0.9356	0.1	1	0.7373	0.0	1	0.8858
64	64	12.3	2	0.0021*	12.0	1	0.0005	0.3	1	0.6147
65	65	1.6	2	0.4509	0.2	1	0.6913	1.4	1	0.2312
66	66	3.6	2	0.1706	2.9	1	0.0884	0.6	1	0.4283
67	67	4.3	2	0.1157	0.1	1	0.7092	4.2	1	0.0411
68	68	0.5	2	0.7802	0.4	1	0.5167	0.1	1	0.7835
69	69	0.9	2	0.6300	0.5	1	0.4903	0.4	1	0.5035
70	70	2.3	2	0.3161	0.1	1	0.7132	2.2	1	0.1413

*Asterisks * shows DIF items. Critical χ2 value = 5.99 at df (2) at 0.05 sig level*

of significance. From the table, it can be observed that Item 3 has p =.0089, p< 0.05, item 4 has p = .0005, p, < 0.05, item 7 has p = .0001, p< 0.05, Item 9 has p =.0008, p< 0.05, Item 10 has p =.0111, p< 0.05, Item 11 has p =.0315, p < 0.05, Item 12 has p =.0234, p< 0.05, Item 16 has p =.0012, p < 0.05, Item 19 has p =.0239, p< 0.05, item 23 has p = .0052, p< 0.05, Item 30 has p =.0005, p< 0.05, Item 35 has p =.0011, p< 0.05, Item 37 has p =.0441, p,< 0.05, Item 40 has p =.0269, p < 0.05, Item 42 has p =.0190, p< 0.05, Item 47 has p =.0101, p< 0.05, Item 50 has p =.0018, p< 0.05, Item 51 has p =.0216, p< 0.05, Item 52 has p =.0508, item 54 has p= 0005, p < 0.05, Item 58 has p =.0373, p< 0.05, Item 62 has p =.0170, p< 0.05, and Item 64 has p =.0021, p< 0.05 revealing that these items significantly differentially functions. Thus, between the focal group Low socio-economic status (LSES) and the reference group High socio-economic status (HSES), these items

4. DISCUSSION OF FINDINGS

The findings of this study are discussed below.

4.1 Differential Items Functioning between Male and Female Group

This present study reveals that of the 70 items based on gender 13 functioned differentially significantly between the male and female group as seen from their Wald Chi- square values which were greater than the critical value of 5.99 at df 2 as well as their p-values were all significant at .05 (p< .05). This represents as well a percentage of 18.5% of the total percentage for the 70 items. This implies that the English Language achievement test used in WASSCE 2016 contains items that significantly functions differently between male and female examinees. Thus revealing that these 13 items which represents 18.5% of the whole items have bias potentials against the male group. This finding is in line with the findings of Uwhekadom [21] whose research revealed that chemistry multiple choice questions used by WAEC in the 2009, 2010 and 2011 SSCE contains 7 test items that significantly function differently for male and female test takers students. These items functioning differentially between these two groups could be as a result of unfamiliarity to content of test items causing attractions to responses other than the right key. Consonant with this is the findings of Umoinyan [26] who investigation on mathematic multiple-choice test used by West African Examinations Council (WAEC) in the 1990 General Certificate Examination showed evidence of gender differentially functioning items on 5 items.

In the same vein, Odili [30] research findings indicated that biology multiple choice question used by WAEC in the SSCE contains test items that significantly functioned differently for male and female testes. Literature also reveals that this tendency is not specific to questions used by WAEC only both other examination body like NECO contains items with similar test characteristics. Similarly, in a study by Reuben & Akorede [23] Five (5) items functioned differentially for male and female group in the NECO (SSCE) 2013 economics objective test items. Also Metibemu [31] found out that based on gender, 33 items showed DIF. However, this result is not in line with Igbokwe [32] who found out that there was no significant difference between male and female when she developed item bank for mathematics for NECO common entrance examination. This divergent result with that of this present finding may be due to the difference in sample size, the sample size for this study are students in certificate class (SS3) while that for Igbokwe [32] was students in Primary school showing a vast difference in the sample particulars of the two studies.

4.2 Items Differential Functioning between High Socio-economic Status Group and Low Socio-economic Status Group

Result reveals that items functioning differently between the focal group which is the Low socio-economic status group and the reference group which is the high socio-economic status group as seen from the 23 items Wald Chi- square values which were greater than the critical value of 5.99 at df 2 as well as their p-values were all significant at .05 (p< .05). This represents as well a percentage of 32.5% of the total percentage for the 70 items for high and low socio-economic status group. This implies that the English Language achievement test used in WASSCE 2016 contains items that shows DIF significantly between the focal and reference group with the DIF items favoring the LSES focal group. Supporting this result is the result of Reuben & Akorede [23] who identify 3 items in the NECO (SSCE) 2013 Economics objective test items that functioned differentially for groups from two different socio-economic status (SES). Similarly, Uwhekadom [21] research showed that chemistry multiple choice questions used by WAEC in the 2009, 2010 and 2011 SSCE

contained items that significantly function differently examinees from high and low socio-economic status. In the same vein, Odili [30] research findings indicated that biology multiple choice question used by WAEC in the SSCE contains items that significantly functioned differently for examinees from high and low socio-economic status. One of the reasons for this, as put by Odili, is that students from high socio-economics status are exposed to varied reading textbooks. Also efficiency in the use of English language tends to be higher for students from high socio-economic status. Another reason for this DIF between these two groups could be disparities of examinees exposure to vocabularies, concepts or skills reflected in the items and probably inferior lessons received from examines

5. CONCLUSION

Based on findings, it is concluded that significantly there are items that functioned differentially between male and female test takers and also between high and low socio-economic status in the English Objective test used by WAEC in 2016 WASSCE and as such they have bias potentials. as items exhibiting DIF is the empirical evidence used to refute or support bias. Findings also reveal that Item response theory method a method in detecting DIF items which was the method adopted for the study was able to adequately identify items exhibiting DIF using IRTPRO software which employs Lord Wald statistics

Concluded again based on findings is that the percentage of DIF identified for the English language objective test items for the groups based on gender, and socio-economic status was quite large. The issue of DIF is an important tool in helping test developers recognize some questions that may be unfair for test-takers because of their gender, and socio-economic status background and so on also it is an issue of great concern as it can be deduced from findings that public examination bodies like WAEC in some cases contain items that exhibit DIF and as such produces test scores that are do not reflect the true picture of examines pe rformance and thus do not lend itself to accurate and valid inference.

The results, which described items that substantially functioned differentially, suggest that the English language test used by WAEC in the 2016 SSCE includes items with bias potentials. This suggests that, in addition to English language ability, the test could be assessing other factors that are unrelated to English language ability for test takers from various socioeconomic backgrounds, as well as male and female test takers. According to the results of the report, items found as having a high percentage of DIF in large scale or public examinations should be examined further using qualitative analysisand content analysis by subject matter experts. If such DIF items are found to be biased during an investigation, they can be edited or removed from a test or item bank.

ETHICAL APPROVAL

As per international standard or university standard written ethical approval has been collected and preserved by the author(s).

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Enamiroro EP. Education poverty and development in nigeria: The way forward.in the 21st century. Journal of Social Science. 2007;14(1):19-24.
2. Brown A. Measurement invariance. Cambridge: Cambridge University Press; 2012.
3. Lee SH. Lord Wald test for detecting DIF in multidimensional IRT model; A comparism of two approaches. Unpublished doctoral dissertation. State School University of New Jersey. New Jersey; 2015.
4. Battuaz M. On Wald's test on differential item functioning detection method; 2017.
   Retrieved on 6/4/2017
5. Wiberg M. Measuring and detecting differential item functioning in criterion referenced licensed test: A theoretical comparison of methods. Educational Measurement. 2007;60:1-33.
6. Siebert C. Differential item functioning identification strategy for items with dichotomous responses. Using the item information cure: A weighted Area Method (WAM). Unpublished Ph.D. Dissertation. Florida State University; 2013.
   Retrieved: on 11/08/2016
   Available:http://diginole.lib.fsu.edu/etd

7. Karami H, Nodoushan MA. Differential item functioning dif: Current problems and future directions. International Journal ofLanguage Studies. 2011;5(3):133-142.
8. Perrone M. Differential item function item bias: Critical considerations in test fairness. Teachers college, Columbia University Working Papers in FESOL & Applied Linguistics. 2006;16(2):1-3.
9. McNamara T, Roever C. Language testing: Thesocial dimension. Malden, MA & Oxford: Blackwell; 2006.
10. Kim SH, Cohen AS. A comparison of Lord's chi-square, Raju's area measures, and the likelihood ratio test in detection of differential item functioning. Applied Measurement in Education.1995;8:291–312.
11. Oshima TC, Morris SB. An NCME instructional module on Raju's differential functioning of items and tests (DFIT). Educational Measurement: Issues and Practice. 2008;27:43-50.
12. Woods CM, Cai L, Wang M. The Langer-improved Wald test for DIF testing with multiple groups: Evaluation and comparison to two-group IRT. Educational and Psychological Measurement. 2013;73: 532-547.
13. Ertuby C, Russell RJH. Dealing with compatibility problems of cross cultural data. Paper Presented at the International Congress of Psychology. Montreal; 1996.
14. Hambleton R, Rodgers J. Item bias review: Practical Assessment, Research and Evaluation. 1995;4(6).
Retrived: on 10/11/2016
Available:http://PAREonline.net/getvn.asp/v=4andn=6
15. Raju NS. An Empirical assessment of the mantel-haenszel statistic for studying differential item performance. Applied Measurement in Education. 1989;12(1): 1-13.
16. Hidalgo MH, Lopez Pina JA. Differential item functioning detecting and effect size: A comparison between logistic regression and mantel-Haenszel procedures. Educational and Psychological Measurement. 2004;64:903-915.
17. Parshall CG, Miller TR. Exact versus asymptotic mantel-Haenszeldif statistics: A comparison of performance under small sample conditions. Journal of Educational Measurement. 1995;32(3);302-316.
18. Federal Republic of Nigeria National Policy on Education (1st Ed). Lagos: NERDC Press; 2004.
19. Akran M, Ghani M. The relationship of socioeconomic status with language learning motivation. International Journal of Englishand Education. 2013;2(2):406-413.
20. Ariani MG, Ghafournia N. The relationship between socio-economic status and beliefs about language learning: A study of irania post graduate EAP students. English Language Teaching. 2015;8(9):17-25.
21. Uwhekadom EE. Influence of language manipulation on differential item functioning in chemistry multiple choice test items. Unpublished Ph.D thesis, University of Port Harcourt; 2014.
22. Madu BC. Analysis of gender related differential item functioning in mathematics multiple choice items administered by West Africa examination council (WAEC). Journal of Educational and Practice. 2012;3(8):71-78.
23. Reuben E, Akorede S. Differential item functioning technique for detection of item bias in Economic among secondary school students in Abuja metropolis. Nigerian Journal of Educational Research and Evaluation. 2016;15(1):112-125.
24. Engelthard G, Wind SA, Kobrin JL, Chajewski M. Differential item and person functioning in large –scale writing assessments within the context of the SAT. College Board Research Report no. New York: The college Board; 2013.
25. Ogbebor U, Onuka A. Differential item functioning method as an item bias indicator. International Research Journal. 2013;4(4):367-373
26. Umoinyang. Item bias in mathematics achievement test. Unpublished M.ed. thesis, University of Calabar; Calabar; 1991.
27. Abiam PO. An analysis of differential item functioning of 1992. first school leaving certificate examination (FSLSE) in Cross River State. In Gani, A Badmus & Peter, I. Odor (eds) challenges of managing Educational Assessment in Nigeria, JAMB, NABTE, NBEM; 1996.
28. Nworgo BG, Odili JN. Analysis of dIF in waec, ssce biology multiple choice test; 2005.
Retrieved: on 10/09/2016
Available:http//ww.aeafrica.orgPAPERS/effectof20%wlagmanipulation/

29. Cai L, Du Toit SHC,Thissen D. IRTPRO.2.1 [Computer software]. Lincolnwood, IL: Scientific Software International; 2011.
30. Odili JN. Effect of language manipulation on differential item functioning of test item in biology in a multinational setting. Journal of Education Assessment in Africa. 2010;4: 264-286.
31. Metibemu MA. Comparing Classical Test Theory and Item Response Theory in the Development, Scoring and Equating of Secondary School Physics test in Ondo state Nigeria. Ph.D. International Centre for Educatioal Evaluation, Thesis University of Ibadan; 2016.
32. Igbokwe E. Development of differential item functioning for placement of students into federal unity schools in Nigeria. Ph.D. International Centre for Educatioal Evaluation, Thesis University of Ibadan; 2004.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/67114