IJEAPR

International Journal of Educational Administration, Planning, & Research
University of Calabar Press
eISSN: 2787-026X; www.ujs.unical.edu.ng/index.php/ijeapr
Vol. 12, No. 1&2., 2020

Differential Item Functioning of 2018 Basic Education Certificate Examination (BECE) in Mathematics: A Comparative Study of Male and Female Candidates

Ememobong Mfon Ekong[1], Isaac Ofem Ubi[1*], Eni Iferi Eni[1]
[1]*Department of Educational Foundations, University of Calabar, Nigeria*
**Corresponding author*
Email: *doctorubi08@yahoo.com*

Doi: 10.5281/zenodo.5892613

Abstract

The study examined the differential item functioning (DIF) of 2018 Basic Education Certificate examination (BECE) in Mathematics tests of National Examination Council (NECO) and BECE of Akwa Ibom State government in Nigeria. The invariance in the tests with regards to sex was considered using Item Response Theory (IRT) approach. The study area was Akwa Ibom state of Nigeria having a student population of 58,281 for the examination. The sample was made of up 3810 students drawn through a multi-stage sampling approach. The multidimensional IRT (MIRT) package implemented in R-programming language software was applied in analyzing the data. The findings reveal that BECE of NECO displayed 23(38.3%) DIF items while BECE of Akwa Ibom State had 37 (61.7%) DIF items in terms of sex. The findings also revealed that, in the two examinations, more items favoured the male candidates more than the female candidates in terms of performance. It was recommended that IRT model should be adopted by test developers to determine item parameters for selection of good items to ensure quality of items before administration. Test equating of students who write equivalent form examinations conducted by different examining bodies was also recommended for admission and placement of candidates to determine actual group differences in performance. The study posits that research on Differential Item Functioning is inconclusive, so should be encouraged.

Keywords: Differential item functioning, invariance, multidimensional, item response theory, R-programming.

Cite as: Ekong, E. M., Ubi, I. O., & Eni, E. I. C. (2020). Differential item functioning of 2018 Basic Education Certificate Examination (BECE) in Mathematics: A comparative study of male and female candidates. *International Journal of Educational Administration, Planning & Research*, *12*(1&2), 57-65. https://doi.org/10.5281/zenodo.5892613

Introduction

The importance of education as a vehicle for national development has been widely acknowledged. Education aims at bringing an ultimate change in the individual, through systematic and sustained effort at acquisition of knowledge and inculcation of values and skills for social integration (Bassey, Bisong, Isangedighi & Ubi, 2011). In every field of endeavour, there are different instruments designed to detect or measure a given quality or phenomenon of

interest. The educational sector uses test as an instrument in measuring the performance of students on particular skills. At the secondary school level, tests are conducted at every phase of learning process (weekly, monthly, per term or annually, as the case may be) to determine the extent to which the students have been able to learn a specific task. Many researchers have expressed concern on the concept of quality and quantity of education. They observed that the indicators of quality could be categorized as material, physical and financial resources, implying that these resources are components of instruments which people use in evaluating the effectiveness of schools (Ukpor, Ubi & Okon, 2012).

Test as an instrument in the hands of any teacher is very important. This is because it is a very relevant tool in educational measurement and has received an endless interest in education. It is the tool that enables teachers to place judgment, make decision, check performance, get response style or picture and determine students' ability (Ubi & Udemba, 2021). It has become a common practice in the education sector that when two or more examinees obtained the same scores from different examinations or tests drawn from the same syllabus which are conducted by different examination bodies, the examinees are expected to have the same ability. The certificates obtained from these examinations are used interchangeably. The decision on examinees performances who take different forms of examination is taken without considering if the items in the examinations that have generated the scores are of the same quality. This practice is not different in the case of Basic Education Certificate Examination (BECE) that is conducted by National Examinations Council (NECO) and that conducted by state governments in Nigeria. Decision made by test users on the performance of examinees who take different forms of tests measuring similar skills can be misleading when used for admissions, promotions, or placements. The reason is that, the psychometric properties of items in the different forms of examinations may not be equal. It is important for users of test scores to compare the quality of items of different tests to ensure that fair decisions are taken. This forms the basis of the present study.

Differential Item Functioning (DIF) is a condition in which an item functions differently for respondents from one group to another. The presence of DIF is due to some characteristics in an item that result in differential performance for individuals of equal ability but from different group (Omorogiuwa & Iro-Aghedo, 2016). In testing, Items that show DIF are serious threats to the validity of the instruments that measure the trait levels of members from different populations or groups. Instruments containing such items may reduce the validity of comparison for between-group because their scores may indicate variety of attributes that are different from those the scale intended to measure (Matthew, 2003), thus DIF is applied to determine the fairness of test to different groups of examinee such as sex with the same trait level.

Peterson (2008) recommended that studies for population invariance, equating for gender and major racial/ ethnic subgroups should be conducted on all testing programmes with high-stakes outcomes because the results are likely to be comparable across the subgroups. Testing programmes also need to conduct studies for major subgroups that could differ in ways that are related to the ability being measured and/or that comprise of a varying proportion of the testing population at different administrations.

Adedoyin, Nenty & Chilisa (2008) investigated the invariance of item difficulty parameter estimate based on CTT and IRT. One hundred and fifty five (155) different independent samples were drawn from the population of 35,262 students who sat for the 2004 Paper 1 Botswana Junior Secondary School Certificate in Mathematics. The samples were selected based on gender, gender by educational regions, ability groups and educational regions. The findings were that, the item difficulty parameter estimates based on CTT theoretical framework varied across the different independent samples. The item difficulty parameter estimates based on IRT theoretical framework were invariant across the different independent groups. Hence, there was invariance of item difficulty parameter in terms of sex, based on IRT but there was variance in CTT framework.

Other studies like Akanime (2017), Essen (2015), Beer (2004), Wen-ling and Rui (2008), Le (2006), Lui and Holland (2008), Obinne (2007), Omorogiuda and Iro-Aghedo (2016), Agah (2013) also point to the fact that differences significantly occur in test item functioning based on sex and other heterogeneous groups of any given population. The main problem of this study is the non-application of IRT principles, especially the Differential Item Functioning (DIF) procedure by examining bodies in calibrating test items meant for qualifying examinations. Most examination bodies which attempt to carry out item analysis do so using the Classical Test Theory (CTT) approach. In modern day test appraisal, CTT is considered archaic and unacceptable. It is on this backdrop the present study examined the Differential Test Item Functioning of 2018 Basic Education Certificate Examination (BECE) in Mathematics of two examination bodies in Akwa Ibom State, Nigeria. The examination bodie examined were NECO and the Akwa Ibom State examination unit.

Research question

The following one, all-encompassing research question, was framed to guide the study:

1. Which items functioned differentially among male and female students' groups in Mathematics test in BECE of NECO and BECE of Akwa Ibom State government?

Methodology

The study adopted equivalent group design in which two equivalent samples of examinees from a common population were drawn and each tested with one of the two different Mathematics examination questions. The underlying assumptions were that: (a) scores from BECE of NECO Mathematics from schools are scores from test Form A, while Mathematics examination scores from BECE of Akwa Ibom State government are score from test Form B. (b) test Forms A and B are different tests from the same Mathematics curriculum. The study area was Akwa Ibom State in the South-South geopolitical zone of Nigeria. The population of the study is 58,281 JSS 3 students in the 234 public secondary schools owned by the state and six federal government secondary schools located in Akwa Ibom State. The population of the state schools is made up of 26,503 (45.47%) males and 30,824 (52.89%) females, while that of the federal schools is made up of 380 (39.83%) males and 574 (60.17%) females. A multi-stage sampling approach comprising of purposive, stratified and simple random techniques was used for data collection. This resulted in the selection of samples of 954 for BECE administered by NECO,

and 2,866 for BECE administered by Akwa Ibom State government. This gave a total sample of 3,820 used for the study. The main instrument for data collection was the Mathematics results of the candidates in the 2018 BECE examinations obtained from their scripts. The data used for the study were secondary and were collected as presented by the examining bodies, so no validation of the instrument was carried out by the researchers. The underlying assumption was that those examinations were standardized; hence validation of the instrument had already been done by the examining bodies before the instruments were administered. Data analyses were done using Differential Item Functioning (DIF) of Multidimensional Item Response Theory (MIRT).

Results

Research question one

Which items functioned differentially among male and female students' groups in Mathematics test in BECE of NECO and BECE of Akwa Ibom State government? To answer the question, Multidimensional Differential Item Functioning across examinees' sex was conducted. Examinees' responses to BECE of NECO and BECE of Akwa Ibom State Mathematics test items were subjected to analysis using Multidimensional Item Response Theory (MIRT) package implemented in R programming language. Consequently, when the p-value is less than 0.05, there is evidence of DIF, but if greater than 0.05, there is no evidence of DIF. Tables 1 and 2 indicate the results of the multidimensional IRT approach to analysis of DIF of the BECE of NECO and BECE of Akwa Ibom State Mathematics test items with respect to sex.

Table 1 depicts multidimensional IRT statistics for detecting Differential Item Functioning in BECE of NECO across sub-group. The technique was assessed using the information criteria such as Akaike Information Criterion (AIC), Akaike Information Criterion corrected (AICc), Bayesian Information Criterion (BIC) and the sample size adjusted BIC (SABIC). The nested log-likelihoods test of IRT was evaluated by comparing log-likelihoods to detect DIF due to a single two-level grouping variable.

Results in Table 1show that in BECE of NECO, 23 items representing 38.3% of the items in the test (items 5, 6, 11, 12, 13, 17, 20, 21, 22, 24, 26, 28, 30, 32, 36, 40, 42, 43, 51, 52, 56, 59 & 60) function end differentially across the male and female examinees while 37 items representing 61.7% of the total items did not flag DIF. Results also show that 13 items (items 6, 11,12, 13, 20, 28,30, 36, 40, 52, 56, 59, & 60) favoured male group, while 10 items (items 5, 17, 21, 22, 24, 26, 32, 42, 43 & 51) favoured female group of examinees. This reveals that male group of examinees had advantage over the female group of examinees in BECE of NECO. Put more straight forwardly, it means that the males performed better than the females.

IJEAPR

International Journal of Educational Administration, Planning, & Research
University of Calabar Press
eISSN: 2787-026X; www.ujs.unical.edu.ng/index.php/ijeapr
Vol. 12, No. 1&2., 2020

Table 1: Multidimensional Item Response Theory statistics of BECE of NECO DIF with respect to sex

Items	AIC	AICc	SABIC	HQ	BIC	X2	Df	P	Male	Female
V1	2.54	8.85	5.91	6.24	12.26	1.46	2.00	0.48	0.49	0.56
V2	0.64	6.94	4.00	4.34	10.36	3.37	2.00	0.19	0.83	1.30
V3	-1.72	4.59	1.65	1.98	8.00	5.72	2.00	0.06	-5.93	-7.67
V4	-2.36	3.95	1.01	1.35	7.37	6.36	2.00	0.05	0.51	0.44
V5	-4.18	2.13	-0.81	-0.47	5.55	8.18	2.00	0.02*	0.65	0.50
V6	-15.75	-9.45	-12.38	-12.05	-6.03	19.75	2.00	0.00*	-0.68	-0.73
V7	3.19	9.49	6.56	6.89	12.91	0.82	2.00	0.67	0.65	1.08
V8	-1.69	4.61	1.68	2.01	8.03	5.69	2.00	0.06	0.48	0.46
V9	0.78	7.09	4.15	4.48	10.50	3.22	2.00	0.20	2.68	3.38
V10	-0.47	5.84	2.90	3.23	9.25	4.47	2.00	0.11	0.95	1.30
V11	-4.14	2.17	-0.77	-0.43	5.59	8.14	2.00	0.02*	0.77	0.90
V12	-9.35	-3.04	-5.98	-5.65	0.37	13.35	2.00	0.00*	-1.10	-1.20
V13	-3.75	2.56	-0.38	-0.04	5.98	7.75	2.00	0.02*	0.92	1.96
V14	3.69	10.00	7.06	7.39	13.41	0.31	2.00	0.86	0.68	0.58
V15	-1.09	5.22	2.28	2.62	8.63	5.09	2.00	0.08	-3.77	-2.81
V16	3.72	10.03	7.09	7.42	13.44	0.28	2.00	0.87	0.81	1.41
V17	-6.41	-0.10	-3.04	-2.70	3.32	10.41	2.00	0.01*	0.75	0.57
V18	2.53	8.84	5.90	6.23	12.25	1.47	2.00	0.48	-0.36	-0.43
V19	-3.13	3.17	0.24	0.57	6.59	7.13	2.00	0.05	0.85	1.11
V20	-7.68	-1.38	-4.32	-3.98	2.04	11.68	2.00	0.00*	0.56	0.57
V21	-16.59	-10.29	-13.22	-12.89	-6.87	20.59	2.00	0.00*	-10.54	-7.62
V22	-39.89	-33.58	-36.52	-36.18	-30.16	43.89	2.00	0.00*	0.70	0.53
V23	2.43	8.74	5.80	6.13	12.15	1.57	2.00	0.46	0.59	1.60
V24	-14.19	-7.88	-10.82	-10.48	-4.47	18.19	2.00	0.00*	0.48	1.41
V25	2.06	8.36	5.43	5.76	11.78	1.95	2.00	0.38	0.60	1.09
V26	-20.85	-14.55	-17.49	-17.15	-11.13	24.85	2.00	0.00*	0.76	0.60
V27	-0.09	6.21	3.28	3.61	9.63	4.09	2.00	0.13	-0.85	-0.88
V28	-34.73	-28.42	-31.36	-31.03	-25.01	38.73	2.00	0.00*	0.65	0.70
V29	-1.99	4.32	1.38	1.72	7.73	5.99	2.00	0.05	0.79	0.71
V30	-25.73	-19.42	-22.36	-22.03	-16.01	29.73	2.00	0.00*	-0.60	-0.70
V31	3.35	9.66	6.72	7.06	13.08	0.65	2.00	0.72	0.81	1.31
V32	-40.14	-33.83	-36.77	-36.43	-30.42	44.14	2.00	0.00*	0.83	0.77
V33	1.43	7.74	4.80	5.13	11.15	2.57	2.00	0.28	4.10	7.79
V34	0.72	7.03	4.09	4.43	10.44	3.28	2.00	0.19	0.74	0.91
V35	2.97	9.27	6.33	6.67	12.69	1.04	2.00	0.60	0.79	1.30
V36	-3.62	2.69	-0.25	0.08	6.10	7.62	2.00	0.02*	1.27	2.77
V37	-2.63	3.67	0.74	1.07	7.09	6.63	2.00	0.05	0.72	0.78
V38	3.31	9.61	6.68	7.01	13.03	0.70	2.00	0.71	0.73	1.01
V39	-2.14	4.17	1.23	1.57	7.59	6.14	2.00	0.05	-0.94	-1.05
V40	-10.21	-3.91	-6.84	-6.51	-0.49	14.21	2.00	0.00*	0.85	24.70
V41	-0.88	5.42	2.49	2.82	8.84	4.88	2.00	0.09	0.74	0.74
V42	-6.86	-0.56	-3.49	-3.16	2.86	10.86	2.00	0.00*	-0.20	-0.10
V43	-3.97	2.34	-0.60	-0.27	5.75	7.97	2.00	0.02*	-1.62	-0.78
V44	2.11	8.42	5.48	5.82	11.84	1.89	2.00	0.39	6.01	1.76
V45	1.47	7.78	4.84	5.18	11.20	2.53	2.00	0.28	-0.09	-0.10
V46	2.63	8.93	6.00	6.33	12.35	1.37	2.00	0.50	0.92	0.92
V47	-1.91	4.39	1.46	1.79	7.81	5.91	2.00	0.05	0.71	0.94
V48	6.38	12.69	9.75	10.08	16.10	-2.38	2.00	1.00	0.48	0.36
V49	1.26	7.56	4.63	4.96	10.98	2.74	2.00	0.25	0.79	1.06
V50	0.62	6.93	3.99	4.33	10.35	3.38	2.00	0.19	3.53	4.91
V51	-4.04	2.26	-0.67	-0.34	5.68	8.04	2.00	0.02*	-2.27	-1.73
V52	-30.99	-24.68	-27.62	-27.29	-21.27	34.99	2.00	0.00*	0.97	1.04
V53	3.73	10.03	7.10	7.43	13.45	0.27	2.00	0.87	-2.03	-2.16
V54	1.76	8.07	5.13	5.47	11.48	2.24	2.00	0.33	0.40	0.92
V55	2.76	9.06	6.13	6.46	12.48	1.24	2.00	0.54	0.94	1.30
V56	-19.03	-12.73	-15.66	-15.33	-9.31	23.03	2.00	0.00*	0.83	0.96
V57	-0.11	6.20	3.26	3.59	9.61	4.11	2.00	0.13	-0.65	-0.66
V58	2.78	9.09	6.15	6.49	12.51	1.22	2.00	0.54	0.83	1.23
V59	-12.50	-6.19	-9.13	-8.79	-2.77	16.50	2.00	0.00*	-7.31	-9.25
V60	-31.22	-24.91	-27.85	-27.51	-21.49	35.22	2.00	0.00*	-0.26	-0.06

*Items p<0.05 considered DIF

IJEAPR

International Journal of Educational Administration, Planning, & Research
University of Calabar Press
eISSN: 2787-026X; www.ujs.unical.edu.ng/index.php/ijeapr
Vol. 12, No. 1&2., 2020

TABLE 2: Multidimensional Item Response Theory statistics of BECE of Akwa Ibom State government DIF with respect to sex

Items	AIC	AICc	SABIC	HQ	BIC	X2	Df	p	Male	Female
V1	-121.49	-120.26	-115.93	-117.19	-109.57	125.49	2.00	0.00*	-12.01	6.36
V2	-1.80	-0.57	3.77	2.50	10.12	5.80	2.00	0.06	0.76	0.60
V3	-61.15	-59.92	-55.58	-56.85	-49.23	65.15	2.00	0.00*	-9.62	-1.37
V4	0.56	1.79	6.13	4.86	12.48	3.44	2.00	0.18	-0.95	-0.90
V5	-11.14	-9.91	-5.58	-6.85	0.78	15.14	2.00	0.00*	1.87	-0.35
V6	-26.05	-24.82	-20.49	-21.76	-14.13	30.05	2.00	0.00*	-0.03	1.80
V7	13.99	15.22	19.56	18.29	25.92	-9.99	2.00	1.00	-0.06	-0.07
V8	-34.72	-33.49	-29.15	-30.42	-22.80	38.72	2.00	0.00*	0.01	4.25
V9	-35.58	-34.35	-30.01	-31.28	-23.66	39.58	2.00	0.00*	0.15	2.12
V10	-45.74	-44.51	-40.17	-41.44	-33.82	49.74	2.00	0.00*	0.14	7.26
V11	-37.23	-36.00	-31.66	-32.93	-25.31	41.23	2.00	0.00*	-0.13	5.13
V12	-170.44	-169.21	-164.88	-166.14	-158.52	174.44	2.00	0.00*	4.48	9.20
V13	-67.91	-66.68	-62.35	-63.61	-55.99	71.91	2.00	0.00*	0.13	3.79
V14	-7.19	-5.96	-1.62	-2.89	4.73	11.19	2.00	0.00*	-0.28	7.43
V15	4.93	6.16	10.50	9.23	16.85	-0.93	2.00	1.00	-0.06	0.14
V16	-6.01	-4.78	-0.44	-1.71	5.91	10.01	2.00	0.01*	-0.02	0.08
V17	-93.70	-92.47	-88.13	-89.40	-81.77	97.70	2.00	0.00*	-0.42	-0.07
V18	2.50	3.73	8.07	6.80	14.42	1.50	2.00	0.47	-0.61	-2.56
V19	-91.57	-90.34	-86.01	-87.27	-79.65	95.57	2.00	0.00*	-0.64	-1.27
V20	-114.26	-113.03	-108.69	-109.96	-102.34	118.26	2.00	0.00*	0.48	-3.54
V21	-131.16	-129.93	-125.60	-126.86	-119.24	135.16	2.00	0.00*	-0.47	0.89
V22	-112.46	-111.23	-106.89	-108.16	-100.54	116.46	2.00	0.00*	-3.44	-0.04
V23	1.92	3.15	7.49	6.22	13.85	2.08	2.00	0.35	-1.19	-0.97
V24	-62.49	-61.26	-56.92	-58.19	-50.57	66.49	2.00	0.00*	0.30	-0.81
V25	31.00	32.23	36.57	35.30	42.92	-27.00	2.00	1.00	0.70	0.87
V26	-70.80	-69.57	-65.23	-66.50	-58.87	74.80	2.00	0.00*	0.19	0.76
V27	27.20	28.43	32.77	31.50	39.13	-23.20	2.00	1.00	0.20	0.10
V28	-122.34	-121.11	-116.78	-118.05	-110.42	126.34	2.00	0.00*	2.26	0.95
V29	-8.01	-6.78	-2.45	-3.72	3.91	12.01	2.00	0.00*	-1.10	3.06
V30	-68.70	-67.47	-63.13	-64.40	-56.78	72.70	2.00	0.00*	-0.45	1.46
V31	-9.07	-7.84	-3.50	-4.77	2.85	13.07	2.00	0.00*	-0.02	4.11
V32	-97.46	-96.23	-91.89	-93.16	-85.54	101.46	2.00	0.00*	1.52	6.96
V33	-126.33	-125.11	-120.77	-122.04	-114.41	130.33	2.00	0.00*	0.18	2.26
V34	-32.87	-31.64	-27.30	-28.57	-20.95	36.87	2.00	0.00*	-0.52	2.96
V35	-0.39	0.84	5.18	3.91	11.54	4.39	2.00	0.11	-2.15	-1.37
V36	3.17	4.40	8.73	7.46	15.09	0.83	2.00	0.66	-0.59	-1.18
V37	-41.42	-40.19	-35.86	-37.13	-29.50	45.42	2.00	0.00*	0.63	2.91
V38	-6.54	-5.31	-0.97	-2.24	5.38	10.54	2.00	0.01*	0.09	7.01
V39	-7.89	-6.66	-2.32	-3.59	4.04	11.89	2.00	0.00*	-0.86	4.10
V40	4.27	5.50	9.83	8.57	16.19	-0.27	2.00	1.00	-0.11	-0.57
V41	-5.15	-3.92	0.41	-0.85	6.77	9.15	2.00	0.07	-0.32	0.14
V42	-30.30	-29.07	-24.73	-26.00	-18.37	34.30	2.00	0.00*	-0.93	11.04
V43	-43.55	-42.32	-37.98	-39.25	-31.63	47.55	2.00	0.00*	-1.03	7.94
V44	-27.14	-25.91	-21.57	-22.84	-15.22	31.14	2.00	0.00*	-1.03	5.90
V45	32.92	34.15	38.49	37.22	44.84	-28.92	2.00	1.00	-1.01	-0.50
V46	25.44	26.67	31.01	29.74	37.36	-21.44	2.00	1.00	-0.68	-0.34
V47	-0.91	0.32	4.66	3.39	11.01	4.91	2.00	0.09	-1.10	-0.58
V48	-46.95	-45.72	-41.39	-42.66	-35.03	50.95	2.00	0.00*	0.88	0.36
V49	7.24	8.47	12.81	11.54	19.16	-3.24	2.00	1.00	-0.95	-1.13
V50	4.88	6.11	10.44	9.17	16.80	-0.88	2.00	1.00	-0.21	-0.50
V51	-274.31	-273.08	-268.75	-270.02	-262.39	278.31	2.00	0.00*	-0.22	1.13
V52	8.93	10.16	14.49	13.23	20.85	-4.93	2.00	1.00	-0.69	-0.25
V53	6.28	7.51	11.85	10.58	18.20	-2.28	2.00	1.00	-1.08	-1.03
V54	17.64	18.87	23.21	21.94	29.57	-13.64	2.00	1.00	-1.05	-0.88
V55	11.92	13.15	17.49	16.22	23.84	-7.92	2.00	1.00	-1.08	-0.99
V56	-5.25	-4.02	0.32	-0.95	6.68	9.25	2.00	0.06	-1.05	-1.09
V57	-54.60	-53.37	-49.04	-50.30	-42.68	58.60	2.00	0.00*	-1.04	-0.91
V58	2.90	4.13	8.47	7.20	14.83	1.10	2.00	0.58	-0.57	-0.63
V59	-6.74	-5.51	-1.18	-2.44	5.18	10.74	2.00	0.01*	-1.35	3.23
V60	-22.96	-21.73	-17.40	-18.67	-11.04	26.96	2.00	0.00*	-1.24	5.78

*Items p<0.05 considered DIF

Results in Table 2 reveal that in BECE of Akwa Ibom State 37 items representing 61.7% of items in the test (items 1, 3, 5, 6, 8, 9, 10, 11, 12, 13, 14, 16, 17, 19, 20, 21, 22, 24, 26, 28, 29, 30, 31, 32, 33, 34, 37, 38, 39, 42, 43, 44, 48, 51, 57, 59 & 60) functioned differentially across the male and female examinees, while 23 representing 38.3% of items in the test did not flag DIF. Results also show that 29 items (items 6, 8, 9,10,11,12,13,14, 16,19, 20,21, 24, 26, 29, 30, 31, 32, 33, 34, 37, 38, 39, 42, 43, 44, 51,59 & 60) favoured the male group, while only 8 items (items 1, 3, 5, 17, 22, 28, 48, 57) favoured female group of examinees. This reveals that male group of examinees had advantage over the female group of examinees in BECE of Akwa Ibom state government. Put more straight forward, it means that the males performed better than the females.

Discussion

Findings of the study reveal that some items in the test functioned differentially between the sexes with the male group of examinees having advantage over the female group of examinees in BECE of NECO and of Akwa Ibom State government. Put more straight forward, it means that the males performed better than the females in those two examinations. The finding disagrees with the that of Agah (2013), which revealed that all the items showed some level of DIF above 0.50 absolute value criterions hence were insignificant. This implies that, the items of the tests were consistent across sex.

The finding, however agrees with the findings of Essen (2015) on Differential Item Functioning (DIF) of 2014 Unified Tertiary Matriculation Examination (UTME) Mathematics of candidates in Akwa Ibom State, Nigeria. The study applied Bilog-MG technique with 2-parameter IRT model and determined IRT-based difficulty DIF and IRT-based discrimination DIF in terms of gender. The result showed that 17 items displayed IRT-based difficulty DIF where 13 favoured males and 4 favoured female candidates but on the contrary none of the items displayed IRT-based discrimination DIF between the groups. The similarity of the results may be that, the data for the studies were obtained from Mathematics test. The result of the study also agrees with the result of the study by Beer (2004) who assessed the use of differential item functioning analysis for bias analysis in test construction. The data obtained were analyzed using the ITEMAN programmme for the CTT while the ASCAL programme was employed for IRT analysis where 3-PLM was applied in the analysis. The result revealed that for the sex comparison, 18 items favoured the female while 7 favoured the male group. The disparity of the results may be that, the data for the study were obtained from other subject other than Mathematics test.

In the same vein, the result of the study is in consonance with the result of Le (2006) who conducted an analysis of Differential Item Functioning (DIF) for Science items. The mean of the gender DIF across countries and items was -0.05 with a standard deviation (SD) of 0.25, and the mean of the unsigned DIF (absolute value of the DIF) is 0.33 with a SD of 0.12. On the average, about 10% of the 210 items (or about 21 items) were flagged by each country, (6.7% favoured males, 3.3% favoured females), in particular 12% flagged by OECD countries (7.9% favoured males, 4.1% favoured females). The correlation between item difficulty estimates and

gender DIF in this study suggested a trend with a similar ability levels, males seemed to perform better than females on more difficult items. The similarity of the results may be that, the data for the studies were obtained from Mathematics test.

Also, the result agrees with that of the analysis of Differential Item Functioning of 2014 Junior Secondary Certificate Examination in Mathematics in Akwa Ibom State, Nigeria by Akpaime (2017) who carried out the analysis using 2-parameter IRT model on IRT-based difficulty DIF and IRT-based discrimination DIF in terms of gender. The result revealed that 28 items displayed IRT-based difficulty DIF in which 27 items favoured male while one (1) item favoured female students. However, there was no IRT-based discrimination DIF in the items. The findings must have agreed with each other in the sense that, data for the studies were obtained from Mathematics test.

Moreover, the result of the study agrees with the claim of existence of DIF in Biology examination of WAEC and NECO of 2000 – 2002 by Obinne (2007) who researched on Differential Item Functioning (DIF) effect of Biology examination items. DIF was analyzed in terms of gender and location. It was discovered that some items favoured girls while some others favoured boys. By this result, the author concluded the existence of DIF effects in the Biology test constructed by these two examination bodies in Nigeria. The similarity of the findings must have been as a result of using standardized test in the two examinations.

Conclusion

The findings of this study have revealed that gender was a significant source of DIF in Mathematics tests. The items that displayed sex related DIF favoured mostly the male students than the female students. The results also indicate that more items in BECE of Akwa Ibom state government differentiated the male students from females than they did in BECE of NECO. This probably indicates that BECE of NECO for that year had more difficult items than BECE of Akwa Ibom State government. It was recommended that IRT model should be adopted by test developers to determine item parameters for selection of good items to ensure quality of items before administration. Test equating of students who write equivalent form examinations conducted by different examining bodies was also recommended for admission and placement of candidates in order to determine actual group differences in performance. The study posits that research on Differential Item Functioning is inconclusive, so should be encouraged.

References

Adedoyin, O. O., Nenty H. J., & Chilisa, B. (2008). Investigating the invariance of item difficulty parameter estimates based on CTT and IRT. *Educational Research and Review, 3*(2), 083-093. https://www.academicjournals.org/ERR.

Agah, J. J. (2013). *Relative efficiency of test scores equating methods in the comparison of students' continuous assessment measures.* (Ph.D. Dissertation), University of Nigeria, Nsukka.http://www.unn.edu.ngpdf

Bassey, P. U., Bisong, N., Isangedighi, A. J., & Ubi, I. O. (2011). Job performance effectiveness: A comparison of Teachers in public and private Secondary Schools in Cross River State, Nigeria. *Lwati: A Journal of Contemporary Research, 8*(2), 9-15.

Beer, M. D. (2004). Use of differential item functioning analysis for bias analysis in test construction, *SA Journal of Industrial Psychology, 30*(4), 52-58. https://sajip.co.za/index.php/sajip/article/download/175/172.

Essen, C. B. (2015). *Differential item functioning of 2014 Unified Tertiary Matriculation Examination (UTME) Mathematics of candidates in Akwa Ibom state, Nigeria.* (Unpublished Ph.D. Dissertation), University of Calabar, Nigeria.

Le, L. (2006). Analysis of differential item functioning. *Australian Council for Educational Research*, Melbourne/Australia. https://www.acer.org>filespdf

Liu, M., & Holland, P. W. (2008). Exploring population sensitivity of linking functions across three school admission test administrations. *Applied Psychological measurement, 32*, 27– 44. https://doi.org/10.1177/0146621607311576

Matthew, J. (2003). The bias question. *The Atlantic monthly, 292*(4), 130 – 140. https://wwwtheatlantic.com>2003/11

Obinne, A. D. (2007). Differential item functioning (DIF) effects of biology examination items of WAEC and NECO for 2000 – 2002. http://docplayer.net>5219278

Omorogiuwa, O. K., & Iro-Aghedo, E. P. (2016). Determination of differential item functioning by gender in the National Business and Technical Examinations Board (NABTEB) 2015 Mathematics multiple-choice examination. *International Journal of Education, Learning and Development* 4(10) 25-35. www.eajournals.org.

Petersen, S. N. (2008). A dissension of population invariance of equating. *Applied Psychological Measurement, 32*(1), 98-101. http://apm.sagepub.com/cgi/content/abstract/32/1/98.

Ubi, I. O., Udemba, E. C. (2021). Age differentials in calibrated items of WAEC English Language objective test taken by students in Nigeria. *Global Journal of Educational Research, 20*(1), 45-54. https://doi.org/10.4314/gjedr.v20i1.5

Ukpor, C. O., Ubi, I. O., & Okon, A. E. (2012). Assessment of factors determining parents' preference for private secondary schools in rural communities of Cross River State. *Global Journal of Educational Research, 11*(2), 99-106. https://doi.org/10.4314/gjedr.v11i2.4

Wen-Ling, Y., & Rui, G. F. (2008). Invariant of score linking across gender groups for forms of a testlet-based college level examination program. *Examination Measurement, 32*(1), 45-81. https://doi.org/10.1177/0146621607311577