**Assessment of Item Bias in BECE 2024 Mathematics Multiple Choice Using Differential Item Functioning in Public Schools, Zamfara State**

Abstract

The study assess the item bias on students using Differential Item Functioning (DIF) in BECE 2024 mathematics examination in Zamfara State. The research adopted descriptive research of ex-post facto type. The sample size of the study was 400 students drawn through a multi-stage sampling approach from all the Fourteen (14) Local Government Area of the state. A proportionate stratified random sampling technique was used to select 280 male and 120 students from the selected schools in each of the Local Government Area in Zamfara State. The instrument used for data collection was 30 items multiple-choice Mathematics items administered by BECE 2024. Both face and content validity were ascertained to support the validity earlier established by the examination board. The data collected were analyzed using logistic regression. The finding revealed that 17items function differently between male and female examinees, while the second finding shows that 11 items functions differently between mixed and single-sex schools examinees. Based on the finding the study recommended among others, that test experts and developers should explore the use of DIF approach to detect items that are bias in mathematics examination items. It was also recommended that (BECE) should carry out differential item functioning analysis for all the test items as part of test development process.

**Key Words:** Differential Item Functioning, Item bias, Examination, BECE, Mathematics

**Introduction**

Examination has being the foremost instrument for the evaluation of students in Nigeria for either promotion or transition of students from one class to another or certification, in other to provide formal evidence of educational achievement. Examination is conducted at different levels of child’s education. According to Aggarwal, (2008), examination is a process of evaluating the pupils behavioural changes based on several learning experiences systematically presented to a group of pupils over a specific period of time from a given course of study. For any JSS3 students to continue his education at senior secondary school, or admitted to any other senior secondary school, he must have passed the major standardized test for junior secondary school, known as Basic Education Certificate Examination (BECE), organized by National Examination Council (NECO). The Basic Education Certificate Examination (BECE) is the main examination to qualify JSS3 students for admission into senior secondary school (SS1) and vocational schools. The BECE is conducted mostly by each 36 states and FCT for the students at this level. It is conducted, monitored, and marked by this body, for the students of different genders, socioeconomic status, tribes, geographical areas and, private and public schools. Therefore, the items are not rule out of biasness.

It is expected that two individuals at the same level of ability, regardless of what group they belong to, will have the same probability of correctly or affirmatively responding to an items, if this is not true, the item is said to be functioning differentially (Solomon, Momoh & Ikeanumba 2024). Differential item functioning (DIF) according to them is an analysis of performance across groups on specific test items. Bandele & Aborisade, (2018); Adediwura & Asowo (2022) explain that item bias affects the vital psychometric properties of measurements results in terms of validity and reliability. Odili, (2014); Obiebi-Uyoyou, (2023), defines differential item functioning (DIF) as the phenomenon that occurs when test takers from different groups who have been matched on ability level perform in test items in a different way. As a result, subgroups of the students perform differently from other students with the same characteristics, despite taken the same items. Some of them perform outstandingly well, leaving others with below average performance. DIF is said to be present in test when examinees from two subpopulations with the same trait level have different expected scores on the same item (Akhihito & Brandon, 2004).

According to Nigeria Policy on Education (2013), education is compulsory and right of every Nigerian irrespective of gender, social status, religion, colour, ethnic background and any peculiar individual challenges. In order to find the balance among these set of people, it is expedient to design sets of items that is free of biasness.

DIF refers to differences in the functioning in the items across groups, often demographic, which are matched on the latent trait, or more generally, the attribute being measured by the items or test (Agi, Hager & Amuche, 2024). DIF is a statistical property of a test item that indicates how likely it is for individuals from distinct groups, possessing similar abilities, to respond differently to the item. DIF occurs when an item’s properties in one group are different from the item’s properties in another group (Oladele,et al. 2020). The properties of an item may functions differently between examinees of different socio-economic status, gender, ethnic background, religion, or school type despite their uniform learning experiences. David & Curt (2012) opined that no DIF on an item implies that, for the same value on the variable, the expected value of a response to the item from persons from different identifiable group is the same. Differential Item Functioning (DIF), according to Adebule, 2013 also referred to as measurement bias, occurs when people from different groups (commonly gender or ethnicity) with the same latent traits (ability/skill) have a different probability of giving a certain response on a test or questionnaire. An item does not display DIF if people from different groups have a different probability to give a certain response; it displays DIF if and only if people from different groups with the same underlying true ability have a different probability of giving a certain response (Wikipedia.org/wiki).

According to Mohammad, et al. (2020), Differential item Functioning (DIF) is a measurement term or statistical technique used to detect multiple choice items or questions that function differently based on the same ability scale or person’s variable. Karami, (2012) opined that DIF occurs when one group does not have an equal chance of getting a multiple choice item correctly, though its members have a comparable latent trait to the other group.

Obiebi-Uyoyou, (2023) posited that students and pupils are not given equal opportunities because of inequalities in educational opportunities, increased class differentiation, and differential drop out, which results in distinct crimes and vices by students who are not in school. As a result, the items in the examinations required items that function differently which tolerable the students of different status and background equal opportunities, to minimize if not eradicate completely the challenges that are mutually with the inequality in the items. As a result of inequality in the test items, some of these examinees were wounded eternally. In other to stamp out this, all the items in an examination should not benefit a segment of students leaving others detriment.

There are two forms or types of DIF: uniform DIF and non-uniform DIF. Uniform DIF exists when group consistently has an advantage over the other group. While non-uniform DIF, exists where the advantage varies based on the individual’s ability level. The existence of non-uniform or crossing DIF demonstrates that the difference in probabilities of a correct response is not the same at all levels of proficiency between the two comparisons groups (Solomon, et al., 2024). Either of the two, DIF becomes pronounced when individuals from different groups who possess the same underlying true ability, display different probabilities of giving a certain response. Common procedures for accessing DIF are Mentel-Haenszel procedure, Logistic Regression, the Rasch model, Lord’s Chi-Square, Item Response Theory (IRT) based methods, SIBTEST method and Confirmatory Factor Analysis (CFA) based methods.

The issue of fairness is what critics label as “bias” in testing (Agi, et al., 2024). According to Omirin & Audu, (2023), content bias refers to a situation where knowledge and skills tested are not part of educational background of the examinees. They stress further, that lack of familiarity with content in test items disadvantages individuals in their performance. According to them, the individual’s responses to items are not based on other irrelevant abilities. In their submission, language bias occurs where words in items have different or unfamiliar meanings for different examinee subgroups. Kunnan (2004) viewed fairness in terms of the whole system of a testing practice, not just the test itself. Therefore, multiple facets of fairness that includes multiple tests uses (for intended and unintended purposes), multiple stakeholders in the testing process (test takers, test users, teachers, and employers), and multiple steps in the test development process (test design, development, administration, and use) are implicated.

Item bias refers to the presence of differential performance on test item based on certain distinctiveness, such as gender, race, or school type (Andrew, 2023). A test item is considered biased if it functions differently for a specified subgroup of test-takers. Test bias is described as any aspect of a test item or the testing process that unfairly advantage or disadvantages certain of students such as those from different culture, linguistic, socio economic or demographic backgrounds. If test equity is not achieved, a test or test item is biased toward a particular subpopulation of the test taking population (Akihito and Brandon, 2004). According to Adediwura & Asowo, (2023), item bias occurs when there is a vast difference in the performance between male and female students, private and public schools, schools in urban and rural areas, religious views, race, social values and beliefs or from state to state. Effiom, (2021) opined that fairness in assessment of student’s achievement tests in mathematics in our secondary schools is very fundamental as mathematics is the basis for studying other subjects especially in science related courses.

**Statement of the Problem**

The fear-provoking rate at which the biasness in the item create dilemma for the examinees growing loop hole in the education institution specifically in Nigeria has to be corrected. Many at times, some students do perform below average in an examination which has implication on them everlastingly especially for those that do not make the grade due to one or two marks caused by the biasness in the items. These as affected the students in continuing their education when they cannot meet up with the criteria expected. This has proved the notion of how DIF could be harmful and threatening (Agi, et al., 2024).

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 (Ememobong, Isaac & Eni, 2020). Tests that contained differential functioning items cannot be used in achieving the goals of a stable educational objectives and the mathematics curriculum, rather, they will give rise to school drop-out who will eventually become serious social or media and political problems in the society (Obiebi-Uyoyou, 2023).

The report has also shown that the performance of public schools and private schools in mathematics examinations are quite different despite all being the products of the same curriculum. This and the other related issues has created the gap between the students in public and private schools as a result of the threats being posed by the effects of DIF. Examinees’ failure may not be because of their incapability to answer the items correctly but because of the disparity of the test; it can result in many examinees withdrawing from school (Agi, et al., 2024). In a number of cases, either of female or male examinees may be at the receiving end of the various harms formed by the inability of the items to balance.

Many studies have been carried out over the years on Mathematics multiple choice and the findings have revealed that gender was a significant source of DIF in Mathematics tests. According to Ememobong et al., (2020), the items that displayed sex related DIF favoured mostly the male students than the female students. In some cases, a few of research indicates that some items that displayed sex/gender related DIF favoured female students than the male students.

Persistence in the gender-based DIF prompt the researcher to the study. The BECE, as a standardized test for junior secondary school, may not be ruled out of biasness as the researches has been carried out on final year examinations such as NECO, WASSCE, etc, leaving the Junior Level Certificate Examination (JLCE) unattended to over the years. This study assess the item bias using differential Item functioning technique in BECE 2024 Mathematics Multiple Choicein Zamfara State.

**Objective of the Study**

This study was carried out to assess the item bias using Differential Item functioning technique in BECE 2024 Mathematics Multiple Choicein Zamfara State. In specific terms, the study:

1. examine the items in BECE 2024 Mathematics Examination questions that differentially function between male and female examinees.
2. examine the items in BECE 2024 Mathematics Examination questions that differentially function between mixed and single-sex schools examinees.

**Research Questions**

The research work seeks to provide answers to the following research questions:

1. What items in BECE 2024 Mathematics Examination questions function differentially between male and female examinees?
2. What items in BECE 2024 Mathematics Examination questions function differentially between mixed and single-sex schools examinees?

**Methodology**

Ex-post facto research design was adopted for the study. (Nwankwo, 2006); ( Chinyere, 2017) noted that ex post facto research design involves collecting and analyzing data from some variables which are already in place (without manipulating any of them). Students’ performance was determined with respect to male and female (gender), private and public schools, (school ownership). The population of the study comprises of all 21126 junior secondary schools students in 14 Local Government Area public schools that sat for BECE 2024 Mathematics Examination in Zamfara State as at the time of the study. Of this number of examinees, 14756 (69.8%) were male, while 6370 (30.2%) were females respectively. Multi-stage sampling procedure with appropriate technique used to draw 400 registered students to serve as sample for the study. The sampling was done in stages in order to arrive at the size. A proportionate stratified random sampling technique was used to select 280 male and 120 female students from the selected schools in each of the Local Government Area in Zamfara state. The research instrument that was used for the study in data collections BECE 2024 mathematics multiple choice test questions, which comprises of 30 items that was supposed to cover the junior secondary school syllabus. Due to the standard of the test, the validity of the test (face and content validity) was ascertained. Logistic regression method was used to analyze the data. Logistic regression is a viable and flexible procedure for detecting DIF that does not require specific forms of item response function or large sample sizes (Chen and Jin, 2018).

**Results and Discussion of Findings**

***Research Question 1:*** *What items in BECE 2024 Mathematics Examination questions function differentially between male and female examinees?*

**Table 1: Logistic Regression analysis on BECE 2024 mathematics multiple choice examinations**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ITEM** | **B** | **S.E.** | **SIG.** | **EXP(B)** |
| **1** | **.487** | **.156** | **.234** | **1.356** |
| **2** | **.553** | **.166** | **.081** | **1.322** |
| **3** | **.650** | **.175** | **.000\*** | **1.665** |
| **4** | **.454** | **.159** | **.000\*** | **1.563** |
| **5** | **.454** | **.166** | **.001\*** | **2.333** |
| **6** | **.661** | **.165** | **.001\*** | **1.459** |
| **7** | **.0681** | **.177** | **.238** | **1.238** |
| **8** | **.788** | **.161** | **.062** | **1.167** |
| **9** | **.766** | **.157** | **.000\*** | **4.187** |
| **10** | **.834** | **.167** | **.007\*** | **2.235** |
| **11** | **.430** | **.181** | **.007\*** | **1.733** |
| **12** | **.486** | **.175** | **.091** | **1.567** |
| **13** | **.865** | **.169** | **.234** | **4.452** |
| **14** | **.775** | **.167** | **.278** | **3.149** |
| **15** | **.456** | **.170** | **.000\*** | **2.365** |
| **16** | **.1004** | **.183** | **.000\*** | **1.567** |
| **17** | **.496** | **.165** | **.069** | **2.0978** |
| **18** | **.754** | **.156** | **.004\*** | **3.181** |
| **19** | **.754** | **.176** | **.000\*** | **1.234** |
| **20** | **.655** | **.156** | **.187** | **1.987** |
| **21** | **.354** | **.176** | **.121** | **2.654** |
| **22** | **.764** | **.184** | **.012\*** | **2.361** |
| **23** | **.765** | **.165** | **.421** | **1.323** |
| **24** | **.549** | **.186** | **.211** | **1.553** |
| **25** | **.564** | **.167** | **.000\*** | **3.453** |
| **26** | **.651** | **.181** | **.000\*** | **3.221** |
| **27** | **.642** | **.173** | **.000\*** | **2.212** |
| **28** | **.655** | **.162** | **.003\*** | **1.645** |
| **29** | **.564** | **.149** | **.240** | **1.976** |
| **30** | **.456** | **.187** | **.000\*** | **1.771** |

From Table 1 above shows the items in relation to gender of students (male and female), identified by logistic regression method using SPSS version 23. Out of thirty items in mathematics multiple choice test items questions, DIF was present in 17 items. These items are item 3, 4, 5, 6, 9, 10, 11, 15, 16, 18, 19, 22, 25, 26, 27, 28, and item 30.

**Research question 2:** What items in BECE 2024 Mathematics Examination questions function differentially between mixed and single-sex schools examinees?

**Table 2:** Logistics Regression Analysis to show and detect items of the BECE 2024 mathematics multiple choice examinations that function differentially between mixed and single-sex schools examinees.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ITEM** | **B** | **S.E.** | **SIG.** | **EXP(B)** |
| **1** | **1.046** | **.187** | **.000\*** | **1.432** |
| **2** | **0.976** | **.137** | **.201** | **1.271** |
| **3** | **.766** | **.187** | **.000\*** | **1.547** |
| **4** | **1.002** | **.167** | **.000\*** | **1.454** |
| **5** | **1.110** | **.165** | **1.01** | **2.413** |
| **6** | **.928** | **.154** | **.066** | **1.445** |
| **7** | **.603** | **.168** | **1.020** | **1.323** |
| **8** | **1.107** | **.172** | **1.00** | **1.211** |
| **9** | **.919** | **.156** | **.000\*** | **3.432** |
| **10** | **.716** | **.176** | **.009\*** | **1.543** |
| **11** | **.821** | **.172** | **.072** | **2.433** |
| **12** | **1.211** | **.182** | **.103** | **1.382** |
| **13** | **.079** | **.168** | **2.02** | **3.546** |
| **14** | **.599** | **.154** | **.104** | **1.641** |
| **15** | **1.022** | **.165** | **.000\*** | **1.523** |
| **16** | **1.105** | **.177** | **.000\*** | **1.453** |
| **17** | **.881** | **.165** | **.301** | **2.218** |
| **18** | **.667** | **.170** | **.601** | **1.403** |
| **19** | **.913** | **.182** | **2.02** | **1.324** |
| **20** | **1.031** | **.166** | **.063** | **1.609** |
| **21** | **1.009** | **.169** | **.000\*** | **2.239** |
| **22** | **.789** | **.152** | **.001\*** | **1.657** |
| **23** | **.743** | **.149** | **1.04** | **1.541** |
| **24** | **1.121** | **.165** | **.022\*** | **1.608** |
| **25** | **1.011** | **.190** | **.101** | **3.117** |
| **26** | **.656** | **.172** | **.060** | **3.221** |
| **27** | **.910** | **.156** | **.000\*** | **1.514** |
| **28** | **.612** | **.149** | **.060** | **1.752** |
| **29** | **.711** | **.167** | **.061** | **1.839** |
| **30** | **.661** | **.177** | **2.01** | **1.648** |

Table2 above shows the items in relation to status of students (mixed and single-sex schools), identified by logistic regression method using SPSS version 23. Out of thirty items in mathematics multiple choice test items questions, DIF was present in 11 items. These items are item 1, 3, 4, 9, 10, 15, 16, 21, 22, 24 and item 27

**Discussion of findings**

The findings of the study shows the items in relation to gender (Male and female), identified by logistic regression have DIF. Out of 30 items in BECE 2024 mathematics questions in Zamfara State, DIF was present in 17 items representing 57% of the test items. These items are item 3, 4, 5, 6, 9, 10, 11, 15, 16, 18, 19, 22, 25, 26, 27, 28, and item 30. From the findings also, it is observed that these items that showed DIF are due to the structure, language of the questions and stem. These could be the characteristics that affected the examinees response in getting the item correctly. The finding, however agrees with the findings of Ememobong et al., (2020) on differential item functioning of 2018 Basic Education Certificate Examination (BECE) in mathematics male and female candidates. It is also in line with the study of Obiebi-Uyoyou, (2023), who reported that WAEC/SSCE 2021 mathematics multiple choice test items function differentially for male and female examinees, and that of Evans, (2022), which shows that 14 (representing 28%) items out of 50 items were gender bias and have group difficulty difference of +0.5. (IRT Analysis).

Also, the study shows Analysis of research question two which revealed that BECE mathematics multiple choice examination items used in 2024 DIF was present in 11 items representing 37% of the test items. These items are item 1, 3, 4, 9, 10, 15, 16, 21, 22, 24 and item 27. It is also observed that these items that showed DIF are due to either language or structure.

**Conclusion**

Any test conducted by examiners like BECE should be free from any inappropriate clues that can allow test takers to give answers to items in a different way in a test. If contrarily, it makes examinees to perform differently in a test. Based on the findings of the study, it was concluded that there was presence of gender, mixed and single-sex DIF in BECE 2024 mathematics multiple choice items in Zamfara State. It is therefore concluded based on this finding that there was a presence of DIF in the test items.

**Recommendations**

Based on the finding of the study, the following recommendations were made:

1. Test experts and developers should explore the use of DIF approach to detect items that are bias in mathematics test items.
2. Test experts and developer should explore the use of differential item functioning method, particularly the use of logistic regression analysis to detect items that are bias with regards to mathematics multiple choice questions.
3. Test designers should be trained and retrained on how to identify DIF items and how to write DIF-free items.
4. BECE should carry out differential item functioning analysis for all the test items as part of test development process.

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

**References**

Aborisade, O.J. & Fajobi, O .O. (2020). Comparative analysis of psychometric properties of mathematics items constructed by waec and neco in nigeria using item response theory approach. *Educational Research and Reviews. Vol. 15(1), pp. 1-7. DOI:105897 /ERR2019.3850.*

Adebule, S. O. (2013). A study of differential item functioning in ekiti state unified mathematics examination for senior secondary schools. *Journal of Education and practice. Vol 4, No.17.*

Adediwura, A. A., & Asowo, A.P. (2022). Examining the nature of item bias on students’ performance in national examination council (NECO) mathematics senior school certificate dichotomously scored items in Nigeria*. International Journal Of Contemporary Education. Vol.5, No. 1.*

Aggarwal, J. C. (2008). *Essential of examination system: evaluation, tests and measurement*: Vikas Publishing House PVT Ltd. Delhi.

Agi, C. I., Hager, A. E. A. & Amuche, B.E. (2024). Evaluation of item bias using differential item functioning (dif) technique in neco conducted economics examination in Taraba State, Nigeria. *International Journal of Research and Innovation in Social Science (IJRISS) Vol.VIII Issue III. DOI: 10.47772.*

Akihito, Kamata & Brandon, K. V. (2004). An introduction to differential item functioning analysis. *learning disabilities: A contemporary Journal. Vol.2(2), pp.49-69.*

Bandele, S. O., & Aborisade, O. J. (2018). Examining the item bias of mathematics examinations constructed by WAEC and NECO in Nigeria. *International Journal of Quantitative and Qualitative Research Methods, 6*(2), 1-7.

Barnabas, C. M. (2012). Analysis of gender-related differential item functioning in mathemetics multiple choice items administered by west african examination council (WAEC). *Journal of Education and Practice. Vol.3, No.8.*

Bashir, M. I. (2023). Psychometric properties analysis of mathematics west african senior school certificate examination in Dala Education Zone, Kano State, Nigeria. *International journal of Advanced Academic Research. Vol.9,Issue1. ISSN: 2488-9849.*

Chen, H. & Jin, K. (2018). Applying logistic regression to detect differential item functioning in multidimensional data. *Journal of Frontiers in Psychology, 9*(1), 48-57.

David, Andrich & Curt, Hagquist (2012). Real and Artificial Item Functioning. *Journal of Educational and Behaviourl Statistics Vol.37, No.3, pp. 387-416. Htt://jebs.aera.net.*

Effiom, A. P. (2021). Test fairness and assessment of differential item functioning of mathematics achievement test for senior secondary students in Cross River State, Nigeria Using Item Response Theory*. Global Journal of Educational Research. Vol 20, 2021:55-62.*

Ememobong M. E., Isaac, O. U. & Eni, I. E. (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. Vol. 12, No.1&2.*

Evans, G. U., Uko, M. P. & Ekim, R. E. D. (2022). Investigation of differential item functioning of basic education certificate examination (BECE) mathematics items in Akwa Ibon State, Nigeria. *The African journal of behavioural and scale development research. AJB-SDR. Vol. 4, no. 1. ISSN: 2714:3449*

Federal Republic of Nigeria (2013). National Policy on Education, 2014 Edition, Lagos, NERDC.

Henry, O. Owolabi (2019)*.Introduction To Educational Measurement and Evaluation:* Wils & Bankleys Publisher (Nig) Ltd. Ilorin.

Ibrahim, A. (2017). Relative effectiveness of generalized mantel-haenszel, simultaneous item bias test and logistic discriminate function analysis for detecting differential item functioning in ordinal test items. *Ife Psychologia (An International Journal), 25*(1), 104-132.

Jimoh, M. I., Daramola, D. S., Oladele, J. I., Sheu, A. L. and Ogunjimi, M. O. (2020). differential item functioning of mathematics joint mock multiple-coice test items in Kwara State, Nigeria. *Journal of Multidisciplinary Studies. Vol. 9 Issue No. 1, pp. 67-92. Doi: http://dx.doi.org/10.7828/jmds.v9i1.1451*

Karami, H. (2012). An introduction to differential item functioning. *The International Journal of Educational and Psychological Assessment, 11(2) 012, 59-79.*

Kunnan, A. J. (2004). Test fairness. In M. Milanovic, & C. Weir (Eds.), *European language testing in a global context*, (pp. 27–48). Cambridge: CUP

Odili, J. N. (2014). Effect of language manipulation on differential item functioning of test items in biology in a multicultural setting. *Journal of the Association for Educational Association in Africa,* 4, 269-286.

Oladele, B. K. Adegoke, B. A. & Longjohn, D.A. (2020). Assessment of differential item functioning in public examinations mathematics constructed-response tests. *Journal of Positive Psychology and Counseling. Vol. 4.2020.*

Omirin, M. S., & Audu, Godwin (2023). Detection of gender-related differential item functioning of West African Senior School Certificate Examination multiple choice mathematics items in Ekiti State. *Journal of Research in Science Education (JORISE) Vol.4No.1*

Obiebi-Uyoyou, Orevaoghene (2023). Assessment of differential item functioning in mathematics multiple choice test questions in Senior Secondary School Certificate Examination in Delta Central Senatorial District. *International Journal of Research in Education and Sustainable Development. Vol.3, issue 2.*

Solomon, C. O., Momoh, O. C., & Ikeanumba, C. B. (2024). Differential item functioning detection methods: an overview. *International Journal and Research Publication and Reviews. Vol.5, No.2. pp.1555-1564.*