

Research Article DOI: <https://doi.org/10.47434/JEREDA.5.2.2024.162> eISSN:2735-9107

**GENDER RELATED DIFFERENTIAL ITEM FUNCTIONING ON BASIC
GENERAL MATHEMATICS EXAMINATION OF KANO STATE
COLLEGE OF EDUCATION AND PRELIMINARY STUDIES, KANO,
NIGERIA**

Hassan Ahmad Isma'il

Department of Psychology and Counselling, School of Education, Kano State College of
Education and Preliminary Studies, Kano, Nigeria

Email: hassanahmadismail1984@gmail.com

 <https://orcid.org/0009-0002-0889-5550>

Received: 19th March, 2024; **Revised:** 30th April, 2024; **Accepted:** 14th May, 2024

ABSTRACT

Introduction: Testing has been accepted in our society as a method of decision making in formal places such as schools, recruitments, promotions, job placement, researches and guidance purposes. Therefore, quality of assessment instruments is essentials in determining the achievement of the students.

Purpose: The study aimed to determine effect size and direction of bias items in Basic General Mathematics Examination.

Methodology: An ex-post facto research design was employed to conduct the study and a sample of 665 NCE II students were purposively selected from KASCEPS. The data obtained were analysed using logistic regression and IRT based analysis.

Result: The findings of the study revealed that 2 items (16 - 17) were found to have magnitude level of DIF pattern in relation to gender and they all favoured female students. The study concluded that GSE Basic Mathematics examination developed by the College examination team is fair to be used for large population.

Recommendation: Based on the findings of this study its recommended that Basic General Mathematics examination items developers should improve their effort in selecting best items with appropriate words and or symbols so as to measure the target knowledge and skills under the study. Basic general Mathematics test developers should ensure that sufficient and balanced content sampling is followed when developing GSE examination. This will go a long way in establishing better sampling and test-taking objectivity.

Keywords: Bias items, Effect size, DIF, Gender, and Basic General Mathematics



Cite paper as:

Isma'il, H. A. (2024). Gender related differential item functioning on basic general mathematics examination of Kano state college of education and preliminary studies, Kano, Nigeria. *Journal of Educational Research in Developing Areas*, 5 (2), 162 - 170.
<https://doi.org/10.47434/JEREDA.5.2.2024.162>.



Copyright © 2024 The author(s) of this article retain(s) the copyright

PUBLIC INTEREST STATEMENT

Colleges of education stand to benefit from the findings of this study so as to ensure that GSE Basic Mathematics examination receive special attention that will come up with items that measure only the skills and knowledge meant to be measured. The findings of the study will improve teachers' commitment in proper and best item selections.

INTRODUCTION

Assessment within educational setting is the process of gathering information and monitoring students' progress and consequently to make educational decision. Testing is a fundamental part of teaching and learning process used not only as the basis for assessment at the end of teaching and learning process but to guide teaching and aid in the development of curriculum as well as in the needs, learning difficulties, level of mastery and differences among learners. Testing has been accepted in our society as a method of decision making in formal places such as schools, recruitments, promotions, job placement, researches and guidance purposes. Therefore, quality of assessment instruments is essentials in determining the achievement of the students. Achievement tests are those tests design to measure the degree of students' mastery in a given area of knowledge, behaviour and or skills. Achievement test can be essay or objective type.

Tests questions may favour males more than females or vice-versa. Achievement test has to be planned is such a way that it represents the knowledge and or skills in questions. Such tests have to measure what they are supposed to measure and also give equal chance to every test taker irrespective of the gender differences. When a test gives equivalent chance to the group of examinees it said to be a fair test. The central aim of test fairness discussion is to make tests free from bias and to contribute to testing equality. A test may be considered biased when students having the same ability perform contrarily. Mathematics is taught in schools for the role it plays in in educational system and in the day to day social activities. Mathematics as a subject affects all aspects of human life at different levels. Mathematics is seen by

society as the foundation of scientific technological knowledge that is vital in social economic development of a nation.

Nigeria Certificate in Education (NCE) programme is a major sector of teacher education, generally aims at producing teachers with high personal and professional discipline and integrity, teachers who are dedicated with appropriate knowledge, skills and attitude that would facilitate easy achievement of national goals. Like other higher learning programmes of study, it encompasses general studies (GSE) and departmental courses (teaching subjects). The heaviness of credit loads and a number of courses in NCE programme and discrepancies of teaching subject(s) in Nigerian Colleges of Education motivate the adoption of General Studies Education (GSE). GSE examinations of Nigerian Colleges of Education are compulsory courses for every NCE candidate pursuing NCE certificate. From personal experience of the researcher, the reason for lack of information about the properties of G.S.E examination items of Kano State College of Education and Preliminary Studies is clear. GSE examination test items of Kano State College of Education and Preliminary Studies are of unknown properties, item analysis was never performed on GSE, and as such the examination is of unknown psychometric properties. That is the reason why this study analysed GSE test items' psychometric properties.

Differential Item Functioning (DIF) occurs when the responses of individuals having the same ability of interest show systematic differences simply based on their membership in a certain group (Ibrahim, 2017). According to Guilera, Gomez-Benito and Hidalgo (2009), an item is considered to exhibit DIF when examinees from different groups (e.g., ethnicity, culture or gender) have a different probability of endorsing an item, when these are matched on the attribute

measured by the item. Subgroups typically studied in DIF analyses are minority ethnic groups such as African Americans, Hispanics, and Asians, but subgroups based on other examinee characteristics such as gender, religion, and socioeconomic status are also studied. IRT is a psychometric approach emphasizing the fact that an individual's response to a particular test item is influenced by qualities of the individual and qualities of item. It provides procedures for obtaining information about individuals, items and test. IRT rests on two basic postulations: a) The performance of an examinee on a test item can be predicted or explained by a set of factors called traits, latent traits or abilities, and b) The relationship between examinee's item performance and the set of traits underlying the item performance can be described by monotonically increasing function called Item characteristic function or item characteristic curve (ICC).

Studies have been conducted to examine gender related Differential Item Functioning in different types of general examinations. In the study of Pius (2021) to test fairness and assessment of DIF of Mathematics Achievement Test for Senior Secondary Students in Cross Rivers state, Nigeria using IRT. The result of the study revealed that 6% of the total items were observed biased between male and female. Based on the analysis, the study observed that there was gender bias on some items in the Mathematics Achievement Test (MAT). DIF were not observed on 47 items representing 94% of the total items. Three items; 25, 40, and 41 were identified as exhibiting DIF among males and females.

Kamarulzaman and Johar (2011), on assessing gender Differential Item Functioning Analysis for the meaningful E-learning instrument in Malaysia. Who found that there no significant differences between male and female in meaningful e-learning. Enonwah, Akwa and Okon, (2014), on Differential items and Group Functions of Secondary Students Achievements in Mathematics in Cross-River State. The results from the study revealed that seven of the items and one of the nine topics (variations) were

significantly biased against school ownership types. It was therefore recommended that achievement test users should estimate the differential function of instrument to enhance appropriate interpretation of students' results. Abedalaziz, Leng, and Alahmadi, (2018) on detecting a gender-Related Differential Item Functioning Using Transformed Item Difficulty In DIF indexes, females showed a statistically significant and consistent advantage over males on items involving algebra, whereas males showed a less consistent advantage on items involving geometry and measurement, number and computation, data analysis, and proportional reasoning. However it was concluded that gender differences in mathematics may well be linked to content.

Shanmugan (2018) who examined gender differential item functioning for Mathematics in coeducational school cultures. The result of this study showed that boys and girls in a particular school culture have different probability of producing correct responses. The study also revealed that the DIF items on Mathematical assessing subtraction ability favoured girls than boys. The finding agrees with the findings of Faleye and Rasheed (2020), on differential item functioning of Osun state joint promotion examination. They found that one item (item 5) out of 40 items functions differentially by male students.

Queensoap and Orluwene (2017) examined Differential Item Functioning in a Chemistry Achievement Test for Students in Nigeria. The population consisted of SSS III in Niger, Kwara, Anambra and Bayelsa states. The study revealed that 46 items were identified having significant chi-square value, it was found that one large effect size items, 20 medium size effect 25 small size effect, thus there were variety degrees of item bias and DIF in the Chemistry Achievement Test (CAT) items. It is on the background that this study is aimed to determine effect size and direction of bias items in Basic General Mathematics Examination in Kano State College of Education and Preliminary Studies, Kano, Nigeria

STATEMENT OF THE PROBLEM

NCE students of Kano State College of Education and Preliminary Studies are repeatedly failing GSE 212 Basic General Mathematics. This failure leads some of the students to spill over one (1) or even two (2). Students across schools and course combinations are found to be among the victims. The College has recorded a massive failure that in the school of Arts and Social Science, only 26%, 10.5%, and 6.7% of the students passed the subjects in the 2015/2016, 2016/2017 and 2017/2018 academic sessions respectively (DEAR, KASCEPS, 2019). Therefore, this study wants to subject the GSE Mathematics tests Items into Psychometric Analysis of Differential Item Functioning to determine items with magnitude level of DIF and direction with regards to gender.

PURPOSE OF THE STUDY

The purpose of this study is to determine:

1. The effect size of DIF pattern in the KASCEPS-2017 GSE 221 in measuring Basic General Mathematics (form A) in relation to gender.
2. The direction of biased items in KASCEPS-2017 GSE 221 in measuring Basic General Mathematics (form A) in relation to gender.

RESEARCH QUESTIONS

1. Which items in the KASCEPS-2017 GSE 212 basic general Mathematics (form A) test show magnitude level of DIF pattern in relation to gender?
2. What is the direction of biased items in the KASCEPS-2017 GSE 212 in measuring basic general Mathematics test (form A) in relation to gender?

METHODOLOGY

Design

This study is an ex-post-facto design which seeks to find out the factors that are associated with certain occurrences of already existing condition or state of affairs and searching back in the time for plausible causal factors

retrospectively (Cohen, Lawrence & Marrison, 2007). In this study the ex-post-facto design employed to examine whether items in the KASCEPS GSE-2017 Basic General Mathematics (form A) set by Kano State College of Education and Preliminary Studies in 2017 function differently by gender.

Population and Sample

The target population of this study consisted of all NCE II students in Kano State College of Education and Preliminary Studies. According to office Academic Secretary of Kano State College of Education and Preliminary Studies there were six hundred and sixty five (665) candidates who sat for GSE Mathematics and English language examination in 2017 (office of Academic Secretary, KASCEPS, 2019). Purposive sampling technique was employed to select NCE II students. The selection of NCE II students is based on the fact that it comprises major variables of the study male and female students, Science and Arts students. This by implication means the population encompasses Arts and Science students who sat for KASCEPS-GSE 2017 Basic general Mathematics (form A) examinations.

Large sample size allow researchers to better determine the averages values of their data hence avoid errors from testing small number of possibly atypical sample Borg and Gall 1996 (in Cohen, 2007) suggest that, as a general rule, sample sizes should be large where there are many variables, the sample will be broken down into subgroups the sample is heterogeneous in terms of the variables under study. Bujang at el (2018) recommended that for observational studies with the IRT logistic regression in the analysis, taking a minimum of sample size of 500 is necessary to derive the statistics that represent the parameter. In respect of this background, this study will use the entire NCEII students six hundred and sixty five (665) as sample of the study.

However, it is worth noting that the test items are population unit of analysis as to be observed from the students' responses which is observation unit of analysis. Therefore, the unit of

observation comprises the response of NCE students from two schools of Science and school of Humanities, who sat for basic general Mathematics GSE 212 (forms A) conducted in 2017. This is because the data to be collected would be from the students' performance and ideally the study is not after the students' performance on test items, but is after the items performance based on students responses.

Instrument for Data Collection

The study did not use any new developed instrument for the sake of generating data; this was because the study collected existing data obtained from a constructed and administered examination by College of Education. The examination was assumed to be faced and content validated by experts in the study area. National commission for colleges of education (NCCE) specify that, each NCE II and above question paper and result, must be moderated by experts from either a university and or a recognize higher learning institution. The data collection was in form of pro forma that used to collect responses of NCE II students of Kano State College of Education and Preliminary Studies KASCEPS GSE2017 basic general Mathematics GSE 212 (forms A) multiple choice examination sets, administered by GSE department of the college in 2017. The examination consisted of twenty 20 items. The paper was objective type with A, B, C, D options. Which cut across basic General Mathematics course contents for NCE II, the correct response will be scored as 1 and the wrong or any undecided response will be scored as 0.

Procedure for Data Collection

The researcher collected the data from the office of Directorate of Examination and Academic Records, Kano State College of Education and Preliminary Studies. The data was in form of binary item scores.

Method of Data Analysis

In this study, two different statistical analyses were conducted. First, regression analysis was conducted to determine items with magnitude DIF. Logistic Regression (LR) was employed using SPSS for data analysis for the detection of DIF by school type. Second, IRT based analysis to determine the direction of bias items in relation to gender. The researcher used the presence of the DIF in LR approach is to be determine by testing the improvement in model fit that occur when a term for group membership and a term for interaction between test score and groups membership are added to the regression model. To discover the items that exhibit magnitude amount of DIF effect size, Jodoin and Gierl (2001) classified the effect levels of DIF that are determine with logistic regression (LR) in the Following ways:

- I. A Level: If $R^2 < .035$, a negligible level of DIF is present
- II. B Level: If $.036 < R^2 < .070$, a medium level of DIF is present
- III. C Level: If $R^2 > .071$, a magnitude level of DIF is present

RESULTS

Research Question 1: Which items in the KASCEPS-2017 GSE 212 basic general Mathematics test (form A) show magnitude level of DIF pattern in relation to gender?

Table 1: Summary of Logistic Regression Analysis for KASCEPS-2017 GSE 212 Basic General Mathematics Test (Form A) by Gender

Effect Level	Item Numbers	Percentage s
$R^2 < .035$, a negligible level of DIF	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 18, 19, 20	90.0
$.036 < R^2 < .070$, a medium level of DIF	Nil	0.0
$R^2 > .071$, a magnitude level of DIF	16, 17	10.0
Total		100.0

In finding out which items in the KASCEPS-2017 GSE 212 General Mathematics (form A) test show magnitude level of DIF pattern in relation to gender. The responses of the students were subjected to binary logistic regression analysis using SPSS. The outcome of the analysis was presented in the above table. The result revealed 90.0% of the items have negligible level DIF pattern, none of items representing

0.0% was found to have moderate level DIF pattern and 2 items (16 and 17) representing 10.0% were found to have magnitude level of DIF pattern.

Research Question 2: What is the direction of biased items in KASCEPS-2017 GSE 212 basic general Mathematics test (form A) in relation to gender?

Table 2: Summary of Item Bias Analysis of KASCEPS-2017 GSE 212 (Form A) by Gender

Item No.	Parameter <i>b</i>		Direction of Bias
	Male	Female	
16	4.17	0.51	Favours Female
17	6.01	0.78	Favours Female

Table 2 above depicts the summary of item bias analysis of KASCEPS-2017 GSE 212 general Mathematics test (form A) in relation to gender performed using IRTPRO software. The outcome from the analysis indicated that both items 16 and 17 are in favour of female students group with *b* parameter estimates of 0.51 and 0.78 as against that of male students with 4.17

and 6.01. The ICCs for item 16 and 17 are not identical for both groups, the ICC for male group shifted more to the right. This is a sign that, the items could not differentiate well among the groups meaning that the items were more difficult for male group. It can therefore, be inferred that items 16 and 17 were biased towards female group.

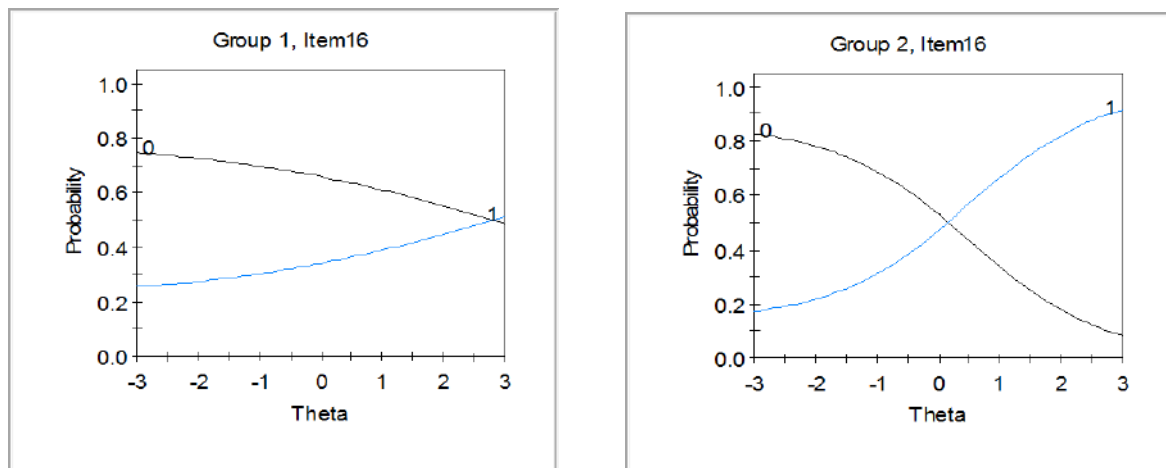


Figure 1a&b: Effect size by gender in DIF Male

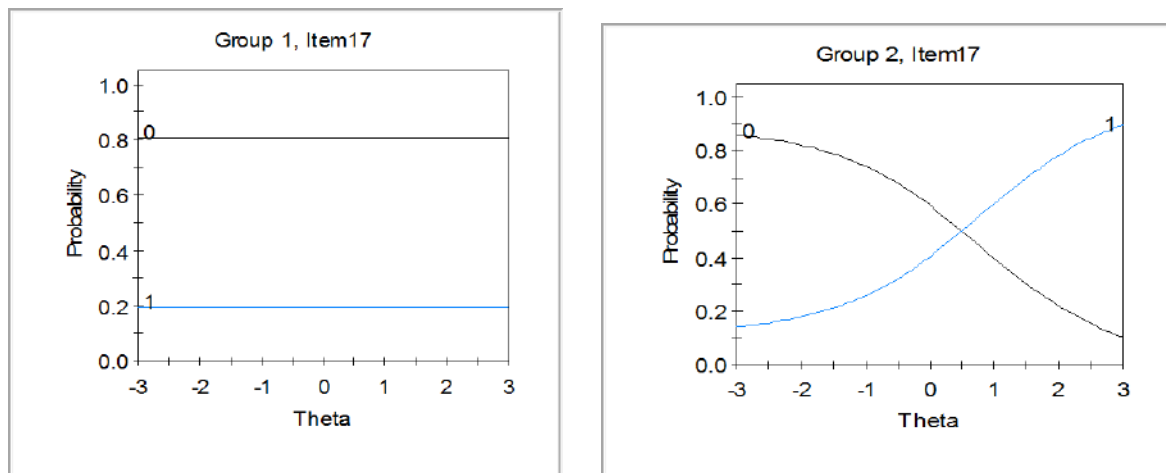


Figure 2: Direction of DIF by gender Male and Female

DISCUSSION

The finding of the study revealed that, two items (16, 17) were found to have magnitude level of DIF pattern in GSE 212 in relation to gender. This finding agrees with the study of Pius (2021) who revealed that 6% of the total items were observed biased between male and female. Based on the analysis, the study observed that there was gender bias on some items in the Mathematics Achievement Test (MAT). DIF were not observed on 47 items representing 94% of the total items. Three items; 25, 40, and 41 were identified as exhibiting DIF among males and females. The finding disagrees with the findings of Faleye and Rasheed (2020), on differential item functioning of Osun state joint promotion examination. They found that one item (item 5) out of 40 items functions differentially by male students. The study also agrees with the

findings of Queensoap and Orluwene (2017) examined Differential Item Functioning in a Chemistry Achievement Test for Students in Nigeria. The population consisted of SSS III in Niger, Kwara, Anambra and Bayelsa states. The study revealed that 46 items were identified having significant chi-square value, it was found that one large effect size items, 20 medium size effect 25 small size effect, thus there were variety degrees of item bias and DIF in the Chemistry Achievement Test (CAT) items.

The finding also showed that, both items 16 and 17 are in favour of female students' group in GSE 212 in relation to gender. The result agrees with the findings of Abedalaziz, Leng, and Alahmadi, (2018) who found in DIF indexes, females showed a statistically significant and consistent advantage over males on items involving algebra,

whereas males showed a less consistent advantage on items involving geometry and measurement, number and computation, data analysis, and proportional reasoning. However it was concluded that gender differences in mathematics may well be linked to content. Also the study agrees with the findings of Shanmugan (2018) who examined gender differential item functioning for Mathematics in coeducational school cultures. The result of this study showed that boys and girls in a particular school culture have different probability of producing correct responses. The study also revealed that the DIF items on Mathematical assessing subtraction ability favoured girls than boys. The finding disagrees with the findings of Faley and Rasheed (2020), on differential item functioning of Osun state joint promotion examination. They found that one item (item 5) out of 40 items functions differentially by male students.

CONCLUSION

This study was conducted to investigate gender related Differential Item Functioning of Basic General Mathematics examination of Kano state College of Education and Preliminary Studies. Based on the findings of this study it is concluded that GSE Basic Mathematics examination developed by the College examination team is fair to be used for the large population. Thus, the instrument was said to be fair to serve its purpose.

RECOMMENDATIONS

Basic on the findings of the study, the following are recommended:

1. Basic General Mathematics examination items developers should improve their effort in selecting best items with appropriate words and or symbols so as to measure the target knowledge and skills under the study.
2. Basic general Mathematics test developers should ensure that sufficient and balanced content sampling is followed when developing GSE examination. This

will go along the way in establishing better sampling and test-taking objectivity.

3. KASCEPS-2017 GSE 212, BASIC GENERAL MATHEMATICS could be used freely among different sub-groups of examinees since it showed almost the same level of validity in the construction of its items and area of fairness as well.

CONFLICT OF INTEREST: The author declared no conflict of interest.

ACKNOWLEDGEMENTS

The author like to acknowledge the contributions of entire staff of Directorate of examinations and academic records of Kano State College of Education and Preliminary studies for the release of all necessary information and data in good time.

DISCLAIMER STATEMENT

This study is original work of the thesis which is part of the unpublished Ph.D thesis in Tests and Measurement submitted to the school of Postgraduate studies in Bayero University, Kano, Nigeria.

NOTES ON AUTHOR

Hassan Ahmad Isma'il is a Ph.D holder in Tests and Measurement working with Kano State College of Education and Preliminary Studies, he has M.Ed Degree in Tests and Measurement and Bachelor's degree in Education in the Departments of Education, Faculty of Education, Bayero University, Kano. He is interested in teaching, own a professional teaching certificate by Teachers Registration Council of Nigeria and conduct researches in the areas like statistics, analysis of Psychometric properties of an assessment instruments, marriage counseling and child development.

REFERENCES

- Abedalaziz, N., Leng, C. H. & Alhamadi, A. (2018). Detecting gender-related of differential item functioning using transformed item difficulty. *Malaysian Online Journal of Educational Science*, 2 (1), 16-22.

- Bujang, M. A. (2018). Sample size guideline for logistic regression from observational studies with large Population: Emphasis on the accuracy between statistics and parameters based on real life clinical data. *Malaysian Journal of Medical Science*, 25 (4), 122- 130. <http://doi.org/10.21315/mjms2018.25.4.12>.
- Cohen, L., Manion, L., & Marrison, K. (2007). *Research methods in education* (6th Edition). Taylor & Francis e-Library.
- Enonwah, C. I., Akwa, A. M., & Okon, C. E. (2014). Differential item and group functioning of secondary students' achievement in Mathematics in Cross-Rivers state. *African Education Indices*, 7 (1), 23-44.
- Faleye, B. A., & Rasheed, A. A. (2020). Differential item functioning of 2015/2016 Biology multiple Choice items of Osun State joint promotion examination. *Alhikmah Journal of education*, 7 (1), 55-67. [http://www.alhikima.edu.ng>article >view](http://www.alhikima.edu.ng>article>view).
- Guilera, G., Gomez-Benito, J., & Hildago, M. D. (2009). Scientific production on the Mantel- Haenszel procedure as a way of detecting DIF. *Psicothema*, 21(3), 492-498.
- Ibrahim, A. (2017). Empirical comparison of three methods for detecting differential item functioning in dichotomous test items. *Journal of teaching and teacher education*, 5(1), 1-18.
- Jodoin, G. M., & Gierl, M. J. (2001). Evaluation type I error and power rates using an effect size measures of logistic regression procedure for dif Detection. *Applied Measurement in Education*, 14, 329-349.
- Kamarulzaman, R. F., & Johar, N. A. (2011). *Gender differential item functioning (GDIF) Analysis for meaningful e-learning instrument*. *Journal of Educational Measurement*, (32), 131-144.
- KASCEPS, (2019). *Semester result on academic records*. Issued on 18/8/2019.
- Pius, E. A. (2021). Test fairness and assessment of dif of Mathematics achievement test for senior secondary students in Cross Rivers state, Nigeria using IRT. *Global Journal of Education Research*, 20, 55-62.
- Queensoap, M., & Orulwene, G. W. (2017). Examining DIF in Chemistry achievement test for students in Nigeria. *International Journal of Education and Evaluation*, 3(7), 78-99.
- Shanmugan, K. S. (2018). Determining gender differential item functioning for mathematics in coeducational school culture. *Malaysian journal of learning and instruction*, 15(2), 83-109.