ASSEREN
Journal of Educational Research and Development
VOLUME 11, JUNE 2024 PAGE 61-70
ISSN: 2536-6899 E – ISSN: 2814 – 3248 Website: www.theasseren.org.ng

ASSESSING THE FUNCTIONALITY OF NATIONAL EXAMINATIONS COUNCIL PHYSICS TEST ITEMS AMONG NORTHERN AND SOUTHERN CANDIDATES IN NIGERIA

Metibemu, M. A., Orowole K. & J. G. Adewale

Submitted: 4 February, 2024 Revised: 11 May, 2024 Accepted: 31 May, 2023

Abstract

The performance disparity among Senior Secondary School students in Physics examinations, particularly in Nigeria, has raised concerns among stakeholders. Despite previous research attempts to proffer a solution to the issue, the persistent underperformance of Northern candidates compared to their Southern counterparts, remains a challenge. This study investigated the variation in the psychometric quality of the National Examination Council (NECO) of the 2017 Physics multi-choice test in Northern and Southern Nigeria. The study adopted a Non-experimental design of descriptive research type. The population comprised 1,051,472 candidates who took the NECO Physics examination in 2017. The instrument consisted of 60 multiple-choice questions from Physics Papers III. The study utilised factor analysis to identify underlying traits in the test and conducted model data fit analysis using multidimensional item response models (MIRT and difR) packages in R software version 4.2 for the analysis. The results revealed that 59 items exhibited DIF, with only item 1 showing no DIF between Northern and Southern candidates. Non-uniform differential item functioning was observed in 40 items, while 19 items displayed uniform DIF favoring the Northern examinees. The findings also indicated Physics Paper III items have three dimensions in the North and two in the South. It was concluded that the test items have more than one dimension and grossly exhibit DIF across the candidate groups. The study recommended that public examining bodies should assess test items during the development stage, considering DIF across or between candidate groups.

Keywords: Differential item functioning, Dimensionality, and Model Data fit.

Introduction

Science is widely recognized as the bedrock for contemporary technological advancements, employing observation, experimentation, and hypothesis formulation to derive conclusions and establish laws, theories, and principles that elucidate various phenomena and concepts. Physics instruction in senior secondary schools, Daramola and Omosewo (2012), gives pupils the chance to develop their critical thinking, analytical reasoning, and spirit of inquiry skills. The sciences of Astronomy, Biology, Mathematics, Engineering, Chemistry, Psychology, Literature, and Philosophy are all closely related to physics. Daramola and Omosewo (2012), argued that the progress of any nation is intricately tied to its robust technological infrastructure and development.

Physics acts as a catalyst for the technological progress of a nation, with its instruction in secondary schools and higher education institutions serving as a driving force behind the quality of a nation's technological advancements. The role of Physics in technology and national development is noteworthy as it plays a significant role in modern inventions that machines, equipment, and tools make life easy. Physics in senior secondary schools prepares students for jobs in astronomy, technology, engineering, and computer science, among other fields. Nigeria recognizing the importance of physics in national development, established the Ministry of Science and Technology and went on to enact a 60:40 science to humanities admission policy (Macmillan & Eray, 2015). Furthermore, Mattew, Ibrahim, and Yakubu (2021), stated that the establishment of universities of science and technology, technical colleges of education, and payment of a special allowance for science teachers in secondary schools are part of the Nigerian government's effort in science development.

Despite the importance of Physics and the tremendous investment of the government in physics education, the percentage of students who have demonstrated proficiency in physics certificate examinations appears to be dependent on the part of the country they reside. For example, the statistics of performance and entry of students from 2014 to 2018 in the West African Senior School Certificate Examination that candidates from the southern part of Nigeria performed better than their counterparts from the North. (Oguguo & Uboh, 2020). Table 1 presents the statistics of enrollment and performances in WASSCE from 2014 to 2018.

Table 1: Aggregated Position of Southern and Northern States and FCT for May/June WASSCE from 2014 to 2018 Years.

Year	South	North	National
2014	861213 (51.5%)	812554 (48.5%)	1673767
2015	834802 (53.0)	741086 (47.0%)	1575888
2016	805762 (53.2%)	709069 (46.8%)	1514831
2017	823426 (52.8%)	734923 (47.2%)	1558349
2018	842713 (54.2%)	711014 (45.8%)	1553727

Source: Federal Ministry of Education, Nigeria Digest of Education Statistics 2017; National Bureau of Statistics.

It is apparent from Table 1 that there is a disparity between the performance of northern and southern students' performance who have a credit pass in compulsory subjects including English and Mathematics and a credit pass in Physics for science students. In the years 2014, 2015, 2016, 2017, and 2018 the southern student has an aggregated percentage of 51.5%, 53.0%, 53.2%, 52.8%, and 54.2% respectively. On the other hand, in the year 2014, 2015, 2016, 2017, and 2018 the Northern have an aggregated percentage of 48.5%, 47.0%, 46.8%, 47.2%, and 45.8%. A cursory look at Table 1 revealed that there is a difference in the performance of students in the north and south in external examinations in Nigeria.

Several studies have attempted to investigate the cause of the poor performance of Physics in Nigeria. For instance, Duruji, Azuh, and Olarenwaju (2014), examined the influence of learning environments such as school facilities, on students' academic achievement in external examinations in secondary schools in Ota Ogun state Nigeria. It was found that learning environments have a significant role in students' performance. The work of Owolabi and Adedayo (2012), revealed that a teacher's qualification affects students' performance. Terpase (2019), found that such as motivation, self-efficacy and locus of control predicted students' academic achievement in physics. Abubakar (2020), investigated the influence of instructional materials on students' academic achievement in Physics, in Sokoto-Nigeria, the study revealed that inadequate instructional materials and periodical usage of teaching aids by Physics teachers influence the level of performance of the students. Despite the efforts of researchers, these differences in the performance of students persist over the years and have become an unanswered question to the government, teachers, and other stakeholders in Nigeria. However, there is a dearth of work on the causes of disparity in the performances of Northern and Southern students.

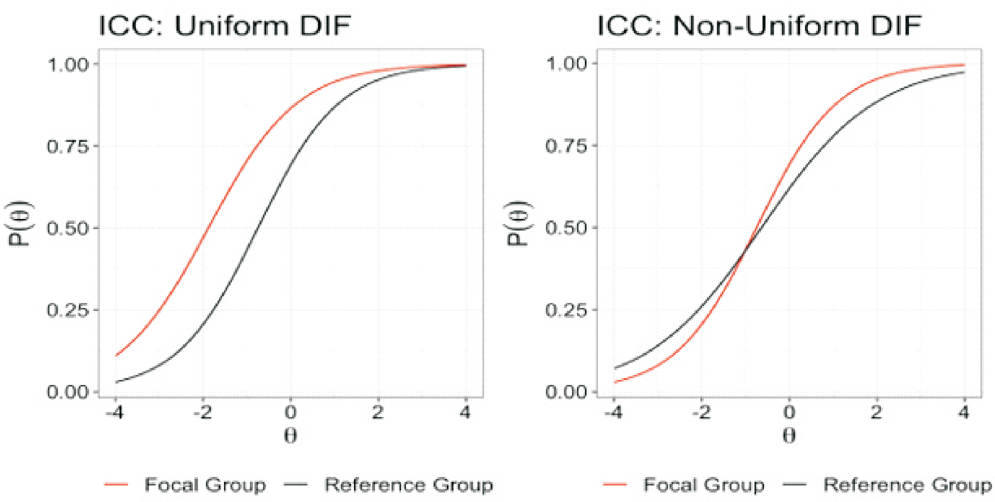
The functionality of test items used by public examining bodies is a critical determinant of students' performance in such examinations.

Test functionality tends to determine the number of traits a test is measuring, and the item parameters in terms of difficulty, discrimination, and guessing. Furthermore, when test functionality is considered, it tends to determine the presence of differential item functioning between groups of examinees in terms of ethnic groups, gender, location, and so on. The differential item functioning (DIF) refers to the different performance of an item for members of different groups that are equal in the ability that is measured by the test. The current practice for determining whether an instrument is biased is to examine the instrument at the item level to see whether one or more items may be considered biased. An item is biased when it favors one group over the other. Differential item functioning, according to Salehi and Tayebi (2012), happens when examinees from various groups exhibit differing possibilities of success on the item after matching the underlying ability that the item is meant to test.

DIF frequently indicates that the test or measure is tapping into a particular form of systematic but constructed irrelevant variance. Differential item functioning occurs when an item shows varying statistical properties for different manifest groups after the groups have been matched on a proficiency measure. Holland and Wainer (1995) and Rust and Golombok (2014) showed that DIF analyses are frequently used to support standardized tests and that items do not favor one or more focal or minority groups (e.g., females, Black people, etc.) over the reference group or a majority group (e.g., males, white people, etc.).

Assessment of differential item functioning (DIF), involves two distinct groups of examinees: the reference group and the focal group. The reference group holds primary significance in DIF analysis, while the focal group is considered the standard for comparison. Typically, the focal group comprises individuals from minority groups, whereas the reference group represents the larger population (Adegoke, 2016). The author argued that DIF analysis aims to identify any differentials in item performance between these two groups. Construct irrelevant elements encompass factors that are unrelated to the construct being measured and may introduce bias in the assessment process. These elements can include variables such as gender, age, geographical location, and ethnic background (Adegoke, 2016).

According to French and Miller (1996), differential item functioning (DIF) was divided into two categories by French and Miller (1996) uniform and non-uniform DIF. When an item consistently gives preference to one group over another over the whole ability range, this is referred to as uniform DIF. On the other hand, non-uniform DIF occurs when examinees' performance on an item interacts with their ability levels, changing the direction of DIF along the ability continuum.



Source (Ayala 2009): Example of Uniform and Non-uniform DIF item

Bichi (2016) conducted a gender base study to assess the differential Item functioning of the chemistry achievement test used in the Kano state promotion examination it was found that the items used by the state-based examination board favored males over their female counterparts. Opesemowo, Ayanwale, Opesemowo, and Afolabi (2023) conducted a study by subjecting the 2017 NECO Mathematics to differential item functioning and it was found that the items are multidimensional and exhibited differential item function for males and females. Oribhabor (2019) who subjected the test items of WAEC November/ December to differential item functioning and found that it is uni-dimensional the study revealed 12 items out of 50 items exhibited DIF. There seems to be a paucity of studies that have been conducted using geographical location as a demographic variable in the assessment of differential item functioning in Nigeria. Alade, Aletan, and Sokenu (2020) undertook a descriptive survey aimed at evaluating the conformity of 2018 WASSCE (SC) Mathematics objective items to the principles of unidimensionality and local independence within the framework of item response theory (IRT). Utilizing a sample of 1334 SS3 students from 18 schools in Lagos State, their findings revealed discrepancies from the expected unidimensional structure and lack of local independence among the test items.

In assessment, there are two frameworks used in detecting differential item functioning; Classical Test Theory (CCT) and Item response theory (De Ayale 2009). Classical test theory comprises three components; the observed score, the true score, and the error score. It is given by

$$X=T+E$$

Where T= true score, E= Error score, X= observed score

Classical test theory assumption is error scores and the true scores obtained by a population of examinees on one test are uncorrelated; Error scores on two different tests are uncorrelated; Error scores on one test are uncorrelated with the true scores on another test (Bichi, 2016).

However, Hambleton and Jones (1993) pointed out the shortcoming of the Classical test theory assessment framework which does not invoke a complex theoretical model to relate an examinee's ability to succeed on a particular item, instead, it collectively considers a pool of responses of examinees on an item. According to Kiany and Jalali (2009), the drawback of classical test theory is circular dependency. The shortcomings of circular test theory are person statistic or observed score which is sample dependent, and the item statistics or item difficulty and item discrimination) are dependent on the examinee (examinee). There are theoretical challenges associated with classical test theory in specific measurement situations such as computerized adaptive testing and test equating.

These limitations of the Classical Test Theory call for a more robust framework in detecting the differential item functioning of an achievement test used by public examining bodies in Nigeria. The dawning of the 21st century brought the Item Response theory to the limelight. Item Response Theory (IRT) is a general statistical theory about the examinee, the item, test scores, and how performance relates to the abilities that the test's items measure (Hambleton & Jones, 1993). IRT focuses on item-level information rather than test-level information that CTT focuses on. In item response theory, there are four models commonly used. They are one, two, and three parameters and the four-parameter model. It can be rewritten as 1-PL, 2-PL, 3PL, and 4PL. In item response theory one-parameter model is commonly called the Rasch model. (De Ayala 2009).

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1+e^{D(\theta - b_i)}} \text{ or } [1+e^{-D(\theta - b_i)}]^{-1} \text{ ------------ 1-PL}$$

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1+e^{Da_i(\theta - b_i)}} \text{ or } [1+e^{-Da_i(\theta - b_i)}]^{-1} \text{ ------------ 2-PL}$$

$$P_i(\theta) = c_i+(1-c_i)\frac{e^{Da_i(\theta - b_i)}}{1+e^{Da_i(\theta - b_i)}} \text{ or } c_i+(1-c_i)[1+e^{-Da_i(\theta - b_i)}]^{-1} \text{ ------------ 3-PL}$$

$$P_i(\theta) = c_i+(\gamma_i-c_i)\frac{e^{Da_i(\theta - b_i)}}{1+e^{Da_i(\theta - b_i)}} \text{ or } c_i+(\gamma_i-c_i)[1+e^{-Da_i(\theta - b_i)}]^{-1} \text{ ------------ 4-PL}$$

Hambleton and Swaminathan (1985 p. 49) highlighted the equation four-parameter model. From the equations above there are emerging parameters common to Item response theory. $P_i(\theta)$ is the probability that an examinee with ability level θ answers item I correctly while *bi* is the item difficulty parameter, a_i is the item discrimination parameter, D1. is a scaling factor and C_I is the guessing parameter. i.e the possibility of the examinee guessing or picking the correct option which is independent of his/her ability while *yi* is the possibility of high-ability examinee not answering the test item correctly, it may be due to carelessness or having information beyond the assumed test item writer; that is not the correct answer.

The sub-optimal performance of Physics students from the northern region of Nigeria in Public examinations poses a significant concern for the government and stakeholders involved in Physics education. Several studies have been conducted such as pedagogical methods and skills adopted by Physics teachers in the classroom, the effects of teacher qualification on Physics student performance, and the effect of school location on outcomes of Physics practicals among other studies. Research efforts have predominantly focused on assessing differential item functioning (DIF) in test items utilized by public examining bodies, employing both Classical Test Theory (CTT) and Item Response Theory (IRT). However, it is evident that despite the proposed interventions such as instruction materials, employing professional teachers, and equipped laboratories among many other panaceas it appears that the academic achievements of students in the northern region of Nigeria significantly differ from their counterparts in the southern region. Consequently, it becomes crucial to conduct a comprehensive differential item analysis on the 2017 NECO test items to ascertain the potential favoritism towards students from the southern region, potentially leading to a disadvantage for their northern counterparts.

Research Questions

This study attempts to answer the following research questions;

1. How many NECO Physics Achievement Test questions function differentially (DIF) for North and South using the IRT assessment framework?
2. What is the dimensionality of the NECO Physics Achievement Test (PAT) measured in North and South?

Methodology

The study utilised a non-experimental design of descriptive research type, to assess differential item functioning (DIF) in the 2017 National Examination Council (NECO) Physics examination. The population comprised 1,051,472 candidates who took the NECO Physics examination in 2017. The research instrument employed was the Physics Papers III (Objective Test questions). The dimensionality assumption of the test employed by the National Examination Council was evaluated through model-data fit analysis using full information factor analysis. Various dimensions were hypothesized to underlie the datasets and the fitness of the data to the dimension were compared. The model exhibiting lower values of AIC, BIC, HQ, and - 2Loglikelihood was selected as the appropriate dimension underlying the data. The assessment of the differential item functioning (DIF) was done using Lord's chi method (1985). The level of significance was set at 0.05, corresponding to a detection threshold of 5.9915. The analysis was conducted using version 4.2 of R programming language software. MIRT and difR packages was used for the analysis.

Research Question 1: How many NECO PAT questions function differentially function (DIF) for North and South using the IRT assessment framework?

Table 1: Differential Item Functioning Analysis of NECO 2017 PAT for North and South

Item	Statistics	P-Value	Remark	Item	Statistics	P-Value	Remark
1	358.652	0.000	DIF	31	67.5014	0.000	DIF
2	215.763	0.000	DIF	32	23.8904	0.000	DIF
3	2079.46	0.000	DIF	33	10.3882	0.000	DIF
4	163.212	0.000	DIF	34	91.1575	0.000	DIF
5	26.812	0.000	DIF	35	131.642	0.000	DIF
6	35.1754	0.000	DIF	36	491.092	0.000	DIF
7	208.096	0.000	DIF	37	33.942	0.000	DIF
8	302.558	0.000	DIF	38	337.824	0.000	DIF
9	247.379	0.000	DIF	39	205.007	0.000	DIF
10	15.3814	0.000	DIF	40	4.5114	0.1048	NO DIF
11	179.032	0.000	DIF	41	168.755	0.000	DIF
12	2220.14	0.000	DIF	42	81.8849	0.000	DIF
13	2526.23	0.000	DIF	43	437.095	0.000	DIF
14	92.2483	0.000	DIF	44	118.373	0.000	DIF
15	55.8683	0.000	DIF	45	48.3479	0.000	DIF
16	146.46	0.000	DIF	46	425.49	0.000	DIF
17	16.3471	0.000	DIF	47	509.251	0.000	DIF
18	227.58	0.000	DIF	48	61.4252	0.000	DIF
19	1613.65	0.000	DIF	49	246.132	0.000	DIF
20	503.695	0.000	DIF	50	1988.88	0.000	DIF
21	283.567	0.000	DIF	51	163.713	0.000	DIF
22	185.838	0.000	DIF	52	412.584	0.000	DIF
23	28.7314	0.000	DIF	53	339.08	0.000	DIF
24	88.5469	0.000	DIF	54	3038.29	0.000	DIF
25	248.952	0.000	DIF	55	166.201	0.000	DIF
26	178.175	0.000	DIF	56	36.6774	0.000	DIF
27	573.227	0.000	DIF	57	2695.63	0.000	DIF
28	27.7153	0.000	DIF	58	614.12	0.000	DIF
29	476.468	0.000	DIF	59	1545.05	0.000	DIF
30	229.064	0.000	DIF	60	680.385	0.000	DIF

Table 1 shows the Lord's DIF statistics using the Item Response Theory framework. This statistic is tested at a significance level ($p < 0.05$) and a detection threshold of 5.9915. When an Item Statistic is greater than the threshold (5.9915) the item indicates DIF, an item statistic with a value less than 5.9915 items did not exhibit DIF. Fifty-nine items exhibited DIF. The Item characteristics Curve of the test items used by the National Examinations Councils in 2017 showed that 40 items (1,2,4,7,8,11,12, 13,14,15,16,17,19,20,21,29,30,31,32,34,35,36, 42,44,45,46,47,49,50,52,53,54,56,57,58,,59, 60) displayed non-uniform differential item functioning for the northern examinee. While 19 items (3,5,6,9,10,18,22,23,24,25,26,27, 28,33,41,43,48,51,55) displayed uniform differential item functioning for the southern examinee.

Research Question 2: What is the dimensionality of the Physics Achievement Test (PAT) measured in North and South?

Table 2(a): Dimensionality of the 2017 NECO Physics Multiple Choice Test in the South

No. of dimension	AIC	SABIC	HQ	BIC	-2Logik
1	8331242	8332495	8331781	8333067	8330882
2	8120284	8121948	8121000	8122707	8119806
3	8182571	8184638	8183460	8185582	8181976
4	8205613	8208077	8206673	8209202	8204906

Table 2(b): Dimensionality of the 2017 NECO Physics Multiple Choice Test in the North

No. of dimension	AIC	SABIC	HQ	BIC	Logik
1	12142058	12143394	12142610	12143966	12141698
2	11743200	11744974	11743933	11745734	11742722
3	11627824	11630029	11628735	11630973	11627230
4	11648517	11651145	11649603	11652270	11647810

Tables 2(a) and 2(b) show the number of dimensions of the Paper III objective test of the NECO achievement test 2017. The result shows that the two-dimension model was the most appropriate model found to explain the number of traits/dimensions underlying the Physics test among the southern candidates. The implication is that two cognitive tasks are responsible for the difference observed in the results of southern candidates. The result showed that the three-dimensional model was the best fit to explain a number of traits/dimensions underlying the Physics test among the northern candidates. The implication is that three cognitive tasks are responsible for the difference observed in the results of northern candidates. This implies that the observed difference in the performance of southern and northern candidates was because the test did not measure the same cognitive task.

Discussion of Findings

The study agrees with the work of Opesemowo, Ayanwale, Opesemowo, and Afolabi (2023) who found that NECO Mathematics test items used in 2017 was multidimensional. The underlying cause of differences in the performance of candidate who sat for the examination. The result of the study is in tandem with Alade et al. (2020), who found that the WAEC 2018 test items did not satisfy the assumption of unidimensionality and that there is more than one dimension that accounted for the variation observed in examinees' responses to the mathematics test items. On the other hand, the study disagrees with disagree with Ogunsanmi, Ibikunle, and Shogbesan (2017) who found that 2017 WAEC Physics multiple-choice items satisfied unidimensionality and local independence assumptions of Item Response Theory (IRT). Similarly, the results disagree with Oribhabor (2019) who subjected the test items of WAEC November/December into dimensionality test using item response theory approach and found that and found that the test items is uni-dimensional and local independence which satisfy item response assumption. The differences in result could be due to the methods of assessing differential item functioning, the author employed Raju Area Measure technique in the assessment .

The findings of research question two revealed that fifty-nine items exhibited differential item functioning while only item one item did not exhibit differential item functioning between Northern and Southern candidates. Non-uniform differential item functioning was observed in 40 items, while 19 items displayed uniform DIF favoring the Northern examinees. This implies that the PAT used by NECO 2017 has 59 items that function differently between the South and the North. The results of the findings disagree with Bichi (2016) who conducted a gender base study to assess the differential item functioning of the chemistry achievement test used in the Kano state promotion examination it was found that 25 items of the 40 items displayed DIF while 13 and 12 favored boys and girls respectively. In the same vein, the study is similar to Adegoke (2016) who subjected the Physics Achievement Test (PAT) to differential item functioning analysis using boys as a focal group and girls as the reference group, the study revealed that 33 items exhibited DIF while 18 items and 15 items favored boys and girls respectively. Moreover, the study disagrees with Oribhabor (2019) who subjected the test items of WAEC November/December to differential item functioning and found that 12 items out of 50 items exhibited differential item functioning. The author's concern was gender using male as the focal group and female as the reference group. Furthermore, the results also disagree with Ogunsanmi, Ibikunle, and Shogbesan (2017) who found that 12 items functioned differentially among males ($p<0.05$) while 20 items functioned differentially among urban and rural examinees at ($p<0.05$). The variation in result might be due to difference in the focal and reference group. Most of the study focused on gender and school location while this study focused on the Northern and Southern candidates as focal and reference groups respectively.

Conclusion

The Physics Paper III objective test functions differently among the northern and southern candidates who sat for the test because it measures different cognitive tasks i.e., two dimensions in the south and three dimensions in the north. Fifty-nine of the tests exhibited differential item functioning between the reference and manifest group while only item 40 did not show differential item functioning.

Recommendations

The study recommended that:

1. Examining bodies in Nigeria should always carry out a model data fitness at all times during test development instead of subjecting the test to an assumed model.
2. Public examining bodies should subject their test items to differential item functioning analysis using the IRT framework before administration.
3. Training programs on Item response theory applications in test development should be organized for staff in public examining bodies.

References

Abubakar, M. B. (2020). Impact of instructional materials on students' academic performance in Physics, in Sokoto-Nigeria. In *IOP Conference Series: Earth and Environmental Science 476(1)* 012071

Adegoke B. A (2010). Effect of continuous assessment on low-achieving Secondary School Students' Performance in Physics and study habits. *African Journal of Educational Research*. *18*, 47-60

Adegoke B. A (2016). Assessment of differential functioning and gender differences in Physics Objective Test. Nigeria *Journal of clinical and counseling psychology* 22(1), 11-30

Alade, O. M., Aletan, S., & Sokenu, B. S. (2020). Assessing the dimensionality and local independence of WASSCE 2018 Mathematics objective tests scores in Lagos State, Nigeria. *The African Journal of Behavioural and Scale Development Research*, *2*(1).

Bichi, A. (2016) Differential item functioning analysis: Evidence of gender-biased items in chemistry achievement test. *Public examination in sub-Saharan Africa: Issues, Challenges, and Prospects*. 67-80

Daramola, & Omosewo, E. O. (2012). An appraisal of the new Nigerian senior secondary school physics curriculum. *Journal of Education and Practice*, *3*(8), 191-194.

Ayala, D. E. (2009). The theory and practice of item response theory.

Duruji, M., Azuh, D. & Olarenwaju, F. (2014). The learning environment and academic performance of secondary school students in external examinations: A study of selected schools in Ota. In *EDULEARN14 Proceedings* (pp. 5042-5053). IATED.

French, A. A., & Miller, T. R. (1996). Logistic regression and its use in detecting differential item functioning in polytomous items. *Journal of Educational Measurement*, 33(3), 315-332.

Gana, C. S., Christian, S. U., & Terpase, A. A. (2019). Students' psychological predictors of academic achievement in physics. *International Journal of Research and Innovation in Social Science, 12(3)* .23-28

Hambleton, R. K., & Jones, R. W. (1993). An NCME instructional module on the comparison of classical test theory and item response theory and their applications to test development. *Educational measurement: issues and practice*, *12*(3), 38-47.

Hambleton, R. K., Swaminathan, H., Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*, 75-99.

Holland, P. W., & Wainer, H. (1995). Differential item functioning. *Psicothema*, *7*(1), 237-242. Kiany, G. R., & Jalali, S. (2009). Theoretical and practical comparison of classical test theory and item-response theory.

Mcmillan, M. J., & Eray, P. L. (2015). Physic Education in Nigerian Secondary Schools: Problems and Prospects. *International Journal of Research in Science, Technology and Mathematics*. *3*(2) 115-123

Federal Ministry of Education, Digest of Education Statistics, August 2017.

Noortgate, W. V. D., & Boeck, P. D. (2005). Assessing and examining differential item functioning using logistic mixed models. *Journal of Educational and Behavioral Statistics*, 30(40), 443-464

Nworgu, B. G., & Odili, J. N. (2005). Analysis of differential item functioning in WAEC SSCE Biology multiple choice test. *International Research Journals*, *4*(4) 367-373

Oguguo, O. S. U. & Uboh, A. V. (2020). State-based analysis of candidates' WASSCE participation and achievement of five credits passes and above including mathematics and English language in Nigeria. *International Journal of Advanced Academic Research*, *6*(6).

Ogunsanmi, O. A., Ibikunle, Y. A., & Shogbesan, Y. O.(2023) Evaluation of differential item functioning (DIF) in 2017 West African Examinations Council (WACE) physics multiple choice test in

Osun state, Nigeria. *Al-Hikmah Journal of Arts & Social Sciences Education, 5*(1) .

Okoronka, U. A., & Wada, B. Z. (2014). Effects of analogy instructional strategy, cognitive style and gender on senior secondary school students' achievement in some physics concepts in Mubi Metropolis, Nigeria. *American Journal of Educational Research*, *2*(9), 788-792.

Opesemowo, O., Ayanwale, M. A., Opesemowo, T., & Afolabi, E. (2023). Differential bundle functioning of national examinations council mathematics test items: An exploratory structural equation modeling approach. *Journal of Measurement and Evaluation in Education and Psychology*, *14*(1), 1-18.

Oribhabor, C. B. (2019). Assessment of the unidimensionality and differential item functioning in the 2017 West African Examination Council (WAEC) November/December mathematics multiple choice test items. *Benin Journal of Educational Studies*, *25(1),* 1-13.

Owolabi, O. T., & Adedayo, J. O. (2012). Effect of teacher's qualification on the performance of senior secondary school physics students: Implication on technology in Nigeria. *Journal of English Language Teaching*, *5*(6), 72-77.

Rust, J., & Golombok, S. (2014). *Modern psychometrics: The science of psychological assessment*. Routledge.

Salehi, M., & Tayebi, A. (2012). Differential item functioning (DIF) in terms of gender in the reading comprehension subtest of a high-stakes test. *Iranian Journal of Applied Language Studies* , *4*(1), 135-168.