PLASU Journal of Educational Research (PLASUJER) Vol. 1, No. 2, 2025
A Publication of the Faculty of Education, Plateau State University, Bokkos – Nigeria
ISSN: 3121-9624 https://journals.plasu.edu.ng

DIFFERENTIAL ITEM FUNCTIONING OF A GEOMETRY DIAGNOSTIC TEST ACROSS ITEM QUALITY AND SCHOOL TYPE IN PLATEAU STAT E, NIGERIA

STEPHEN BANWAR WANLOR
PRIMARY EDUCATION DEPARTMENT
FEDERAL COLLEGE OF EDUCATION, PANKSHIN
Stephenwanlor5@gmail.com
Phone: 08060921837

ZUMJI SEMSHAK DANLADI
PRIMARY EDUCATION DEPARTMENT
FEDERAL COLLEGE OF EDUCATION, PANKSHIN
zumjidanladi@gmail.com
Phone: 08062927665

KAYODE OYENIYI AJERE
PRIMARY EDUCATION DEPARTMENT
FEDERAL COLLEGE OF EDUCATION, PANKSHIN
kayodeajere@gmail.com

BEING A CALL FOR PAPERS FOR JOURNAL OF EDUCATIONAL RESEARCH
FACULTY OF EDUCATION, PLATEAU STATE UNIVERSITY, BOKKOS

ABSTRACT

This paper established differential item functioning of a geometry diagnostic test across item quality and school type in Plateau. Instrumentation and and cross-sectional survey research designs were used. The population was made of all Lower Basic three pupils. A sample of 1350 pupils were drawn using Probability Proportionate Sampling (PPS) method. The instrument was a 30-Items Multiple Choice Geometry Diagnostic Test.The validity of the instrument was estimated through item fit.. Two research questions and one hypothesis guided the study. Data analyses were done using the Xcaliber software. Some results of the findings revealed that 96.6% of the items fit the 3PLM. 50% of the items were of moderately difficult, 80% items discriminated well and 96.67% of the items had good guessing indices. It was concluded that the items had quality and fair to pupils based on school type and recommended among others that mathematics teachers can make use of the instrument to identify pupils with geometry learning misconceptions.

Key words: Lower basic education, geometry diagnostic test, school type differential item functioning, item quality

Introduction

Mathematics is one of the core subjects in the Nigeria s Lower Basic Education Curriculum that provides pupils with the fundamental knowledge towards the understanding of Science, Technology and Engineering. Mathematics often referred to as the study of science of structure, order and relations is inevitable for the understanding of science and technology in Nigeria Geometry is a branch of mathematics that deals with the study of points, lines, angles, shapes, measurement and relationships which helps in making mathematics a functional subject. The application of geometry is also wide in the social and political sciences, structures and sports. It helps pupils in their adult life as professionals in different fields such as surveyors, architects, and engineers. It is also useful to artisans like tailors, carpenters and masons (Hodanova & Nocar, 2016, Fabiyi, 2017). Hence, the need for its study by everyone alike. Fabiyi (2017) and WAEC Chief Examiners from 2017 to 2024 have reported that students performed worse in geometry than in any other area of mathematics on the Senior School Certificate Examination (SSCE).

Several factors have been found to account for this poor performance among which are poor quality teacher made achievement tests. It was found that secondary school mathematics teachers in Abuja do not determine the quality of their tests items before administering them to their pupils (Azuka, 2014) cited in Dadughun (2015). An item is said to have quality when its parameters like item difficulty usually represented by the letter b, item discrimination by the letter a, and Pseudo- guessing by the letter c are determined through application of the Classical Test Theory (CTT), the Item Response Theory (IRT) or the generalizability Theory (G-Theory). Item difficulty refers to the proportion of candidates getting the item correctly, item discrimination refers to the ability of the item to differentiate between high and low ability groups responding to the item while Pseudo-guessing refers to candidates guessing correctly an item in multiple choice question types. The over dependence on achievement test by mathematics teachers of basic education is another factor that does not provide them with the opportunity to check why pupils are failing a particular item. Hence the need for the adoption of a diagnostic test as a method of assessing pupils' performance in geometry.

A diagnostic test is a test that is mostly administered by teachers before the commencement of an instructional programme to find out the level of mastery and non mastery or strength and weakness of the candidate(s) on the content area to be taught. This is aimed at providing the teacher with a comprehensive knowledge of who needs remedial instruction or not (Wanlor, Wakjissa & Mustapha, 2025). In a diagnostic test the teacher is interested in locating the different types of errors or misconceptions the candidates exhibit. Some of these misconceptions might be due to conceptual, procedural, computational or careless errors. A conceptual error occurs when a candidate uses a different approach to solve a problem or does not know how to solve the problem and leaves it unanswered. A procedural error occurs when the candidate knows the problem but omit certain steps and hence arrive at the wrong answer or unable to arrive at the solution to the problem. Computational errors on the other hand occur when candidate cannot perform the basic operation of arithmetic like addition, subtraction, multiplication and division appropriately and hence gets the answer wrong. A careless error occurs when a candidate wrongfully writes a wrong answer due to over confidence, copying wrong figures from the calculator, not arranging numbers according to their place values and so on. These must be checked by the teacher in order to know how to help such candidate(s) to measure up to those who are succeeding in the content area or topic (Wanlor, Mustapha & Wakjissa, 2025).

In all these, the demographic characteristics of the candidate such as school types may influence their performance and hence make some candidates with the same ability to get different scores.

School type in the context of this paper refers to whether the school is a public school or a private school. Public school are schools that are funded and administered either by the local government education authority, Plateau State Universal Basic Education Board or the Federal Ministry of Education or Universal Basic Education Commission (UBEC). Private schools are the schools that are funded by individuals, religious organizations, or communities. There might be disparity in the performance of candidate from these two groups probably because of the availability of learning materials to pupils, quality of teachers in such schools or due to quality of monitoring and evaluation. When performance of candidates varies on an item not because of ability but because of the types of school the candidates belong to is referred to as school type Differential Item Functioning (DIF).

Differential Item Functioning jeopardize with the validity and reliability of the test since the decisions taken on the results of such test items will be bias. Item response theory approach was used in this study through the application of the 3 Parameter Logistic Model (3PLM) to calibrate the developed diagnostic test items so as to identify and remove or modify items that were bias or flag off DIF characteristics, items with very high or low level of difficulty, poorly discriminating items and items that have high guessing parameters. Figures 1, 2, 3 and 4 display the interpretation of the item parameters for a diagnostic test item selection that teachers can use as guides after analyzing their data based on the 3PLM of IRT.

Difficulty value *(b)*	Interpretation
-2.50 to - 1.50	Very easy
-1.50 to - 0.50	Easy
).50 to 0.50	Moderately difficult
0.50 to 1.50	Hard
1.50 to 2.50	Very hard

Figure 1. Interpretation of Item Difficulty Indices for Diagnostic Tests. Source: American Educational Research Association. (2014).

Discrimination Value (a)	Quality of an Item
a ≤ 0.00	Very low discrimination (Not acceptable)
0.20 to 0.50	Low discrimination
0.50 to 1.00	Moderate discrimination (Adequate)
1.00 to 2.00	High discrimination (Excellent) no revision is required
a >2.00	Very high discrimination

Figure 2. Interpretation of Item Discrimination Indices for Diagnostic Tests. Source: American Educational Research Association. (2014)

Guessing Indices (c)	Interpretations
C = 0	No guessing
C ≥ 0.20	Low guessing and ideal

0.20< c ≤ 0.30	Moderate guessing.
C > 0.30	High guessing
C = 1	Complete guessing (item invalid)

Figure 3. Interpretations of Item Guessing Indices for Diagnostic Tests.
Source: American Educational Research Association (2014)

DIF statistics	Interpretation	Decision
Difficulty(b)	≥ 0.2	Significant DIF (item is more difficult for focal group)
	≤ -0.2	Significant DIF (item is less difficult for focal group)
Discrimination	≥ 0.1	Significant DIF (item discriminate more for focal group)
	≤ -0.1	Significant DIF (item discriminates less for focal group)
Pseudo-R2	≥ 0.02	Significant DIF (item functions differently for focal group)
	< 0.02	No significant DIF

Figure 4: Interpretation of DIF in Terms of the a, b and c Parameters for Diagnostic Tests.
Sources: Svetina and Rutkowski (2014)

Results of Empirical finding from Wanlor (2025) revealed that the chi- square goodness of fit statistics employed for analysis of middle basic pupils revealed that 100% of the diagnostic items fit the 3PLM. Similarly, Chamundeswari (2014) and the American Educational Research Association (2020). found the validity index of the instruments they developed were valid (item fit) while Dadughun (2015) found the items of the primary school mathematics diagnostic achievement test developed and calibrated to have good item fit statistics. Similar finding by Wanlor (2025) on item indices of a diagnostic test for middle basic pupils showed that 64% of the items were within the acceptable difficulty level of 0.30-0.50 (moderately difficult), 74% items discriminated well because they had index value of 0.31 and above and 98% of the items were good because they had guessing index value of 0.25 and below. Osadebe and Agbure (2018) found that school type and school ownership differential functioning exist in the 2014 BECE Social Studies multiple choice test. Alade, Aletan and Sokenu (2020) found that all the 50 items of the 2018 WASSCE Mathematics achievement test in Lagos State functioned deferentially based on school type. And Mawak, Mustapha and Wakjissa (2019) findings showed that items 41 and 59 show evidence of DIF based on school type. However, Gyamfi (2023) found that there was no presence of DIF in the items for both gender and types of schools (public or private). from secondary schools in Western region of Ghana.

Statement of the Problem

The performance of Lower Basic Education pupils in mathematics and in the content area of geometry in particular had been poor as revealed by the Basic Education Certificate Examination from 2017 to 2024 with only 10.94% of the students scoring a credit in the subject. While External WAEC Chief examiners also reported that pupils found geometry aspect of Mathematics in the WASSCE the most challenging aspect from 2017 - 2024. Hence the need to address these at the lower basic education level so that the pupils will master it before their transition to middle and upper basic as well as post basic education. Several factors are responsible for these among which are poor quality teacher made achievement test and difference in pupils' performance due to the type of school they attend. In the light of these, the study

developed a geometry diagnostic test that are of quality and free of item bias due to pupils' type of school which can be used to diagnose pupils learning difficulties in geometry.

Aim and Objective of the Study

The aim of the study is to develop a qualitative diagnostic test for Lower basic two pupils that can be used to diagnose pupils learning deficiencies in geometry content area of mathematics through differential item functioning of a geometry diagnostic test across item quality and school type, specifically the study sought to find out.

1. The fit of the items to the model used
2. The item analysis indices of the instrument
3. The differential item functioning based on school type.

Research Questions

The following research question were raised to guide the study.

1. What will be the fit of the items to the 3 Parameter Logistic Model?
2. What will be the item analysis indices of the instrument?

Hypothesis

The null hypothesis was tested at .05 level of significance

Ho: the item of the geometry diagnostic test will not function differentially based on the type of schools the pupils belong.

METHODOLOGY

The study adopted the Instrumentation and Cross- sectional survey research designs.The population was made up of public and private primary lower basic three pupils in Plateau Central Geo- political zone. Plateau Central Geo-Political zone is made up of five Local Government Education Authorities (LGEAs) from where three LGEAs were randomly sampled. The probability Proportionate Sampling (PPS) method was adopted. This method produced a sample size of 27 schools made up of 16 private and 11 public schools. These schools had sample of 600 pupils from private and 750 pupils from public schools. Thus, a sample size of 1350 pupils were selected for the study. All the pupils from a sampled school constituted the sample from such school. The instrument for data collection was a 30 Items Multiple Choice Geometry Diagnostic Test for Lower Basic Education three Pupils (MCGDTLBETP) developed by the researcher. The instrument was made up of sections A and B. Section contained the school types of the pupils while section B contained the 30 multiple choice items. Content validation was obtained through the construction of Table of Specifications and item fit as well as by research measurement and evaluation experts of the University of Jos and mathematics subject specialist from the Lower Basic Education Schools.The differential item functioning (DIF) was established though the z-test statistics of the Three Parameter Logistic Model (3 PLM) of Item Response Theory (IRT).

Results

Research Questions One

Table 1: Items Fit Parameter Indices of the Lower Basic Three Geometry Diagnostic Test

Item ID	Key	X^2	Df	p	Item ID	Key	X^2	df	p
1	B	11.2755	347	0.4918	16	D	12.8655	347	0.3565
2	B	9.3932	347	0.6537	17	C	9.8216	347	0.6181

3	A	10.3019	347	0.5769	18	C	12.2853	347	0.4046
4	C	9.7963	347	0.6202	19	D	13.208	347	0.3291
5	D	17.7332	347	0.2599	20	B	8.9556	347	0.6886
6	D	20.003	347	0.0785	21	A	10.9771	347	0.718
7	B	10.7351	347	0.5392	22	A	10.2947	347	0.5776
8	A	11.1104	347	0.5063	23	C	11.3182	347	0.5781
9	C	12.1655	347	0.4148	24	B	8.7031	347	0.8179
10	C	11.1434	347	0.5034	25	A	13.2406	347	0.4366
11	D	26.403	347	0.0213*	26	B	11.8994	347	0.5576
12	B	37.2927	347	0.413	27	C	15.6102	347	0.2834
13	A	12.9192	347	0.3522	28	A	9.2465	347	0.6256
14	D	18.9595	347	0.1269	29	B	13.8597	347	0.4793
15	D	15.5768	347	0.2754	30	B	15.3336	347	0.3799

Note: n = 1350.

Table 1 shows the item fit parameter indices of Lower Basic 3 Pupils Geometric Diagnostic Test (GDT). The result indicates that the p-values for all the items, except 1 item that is item 11 representing 3.33% is large (insignificant), indicating that 29(96.67%) of the items fit the model. The rule of thumb in model-data fit holds that items with significant values are misfit items for the model. However, items 11did not function in the intended manner hence it needs to be checked. Statistically, it means that there is no difference between expected and observed frequencies of correct answer to the item at various ability levels of Lower Basic 3 Pupils in Plateau Central Geo-political zone except for the above-mentioned misfit item.

Research Questions Two

Table 2: Item Analysis Indices of the Lower Basic Three Geometry Diagnostic Test

Item ID	Difficulty	Discrimination	Guessing	Item ID	Difficulty	Discrimination	Guessing
1	0.360	0.482	0.176	16	0.201	0.584	0.221
2	0.280	0.657	0.160	17	0.321	0.427	0.233
3	0.780	0.168	0.318	18	0.490	0.412	0.238
4	0.520	0.347	0.255	19	0.170	0.713	0.238
5	0.360	0.352	0.239	20	0.210	0.174	0.247
6	0.320	0.323	0.238	21	0.310	0.439	0.232
7	0.400	0.590	0.239	22	0.220	0.509	0.241
8	0.300	0.454	0.234	23	0.330	0.420	0.216
9	0.500	0.272	0.248	24	0.410	0.345	0.239
10	0.580	0.426	0.248	25	0.420	0.442	0.226
11	0.500	0.210	0.250	26	0.310	0.427	0.232
12	0.540	-0.076	0.255	27	0.250	0.533	0.229
13	0.720	0.244	0.250	28	0.350	0.494	0.233
14	0.280	0.576	0.230	29	0.270	0.552	0.249
15	0.360	0.442	0.238	30	0.370	0.361	0.237

Note: n = 1350

Table 2 presents the indices (a, b, and c parameter indices) of the Lower Basic 3 Pupils GDT items. The outcome indicates that 15 items (items 4, 7, 9, 10, 11, 12, 17, 18, 21, 23, 24, 25, 26, 28 and 30) representing 50% of the items are within the acceptable difficulty level of 0.30-0.50 (moderately difficult). Also, 13 items (1, 2, 5, 6, 8, 14, 15, 16, 19, 20, 22, 27 and 29,) representing 43.33% of the multiple-choice items were too difficult (0.0 – 0.29). Whereas, 2 items (3 and 13) represented 6.67% of the total items were very easy (0.720 – 1.0). In terms of discrimination indices, 24 (80%) items (1, 2, 4, 5, 6, 7, 8, 10, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, and 30) were good because they had index value of 0.31 and above. On the other hand, 5 (16.67%) items (items 3, 9, 11, 13 and 20) were "fair" (that is, within the index value of 0.10 – 0.30), while 1 (3.33%) item (item 12,) was "poor" hence fell within the index value of below 0.10. Nevertheless, items 12 has negative value which is too bad and need to be revisited because majority of pupils from low ability group were getting it correctly. In terms of pseudo-guessing (parameter c), 29 items represented 96.67% have index value of 0.25 and below except one item (item 3) which represented 3.33% had index value of above 0.25. The finding generally implied most of the GDT items had good indices in terms of difficulty, discrimination, and pseudo-guessing among Lower Basic Three Pupil in Plateau Central Geo-political zone.

Hypothesis

Table 3: Differential Item Function for Lower Basic 3 Pupils Geometry Diagnostic Test Based on School Type

Item ID	z-test	P	Bias Against	Item ID	z-test	p	Bias Against
1	0.5658	0.5715	N/A	16	0.6214	0.5343	N/A
2	0.4002	0.687	N/A	17	0.4026	0.6871	N/A
3	1.3413	0.1792	N/A	18	0.7602	0.4473	N/A
4	1.8123	0.0690	N/A	19	0.255	0.7978	N/A
5	0.5671	0.5705	N/A	20	0.7602	0.4471	N/A
6	0.6170	0.5370	N/A	21	0.454	0.6490	N/A
7	0.2562	0.7976	N/A	22	0.8336	0.4043	N/A
8	1.0003	0.3170	N/A	23	0.1660	0.8672	N/A
9	0.2108	0.832	N/A	24	0.4551	0.640	N/A
10	0.8288	0.4070	N/A	25	0.401	0.6876	N/A
11	0.9117	0.3612	N/A	26	0.4534	0.6501	N/A
12	0.5670	0.5705	N/A	27	0.6275	0.5301	N/A
13	1.4881	0.1360	N/A	28	0.3523	0.7245	N/A
14	0.8300	0.4062	N/A	29	0.5092	0.6104	N/A
15	0.9041	0.3650	N/A	30	0.566	0.5706	N/A

Note: n = 1350. N/A = Not Applied

Table 3 results indicate the differential item function of GDT items based on school type. The results also indicated that the probability value (p-value) of the z-test for all the items were greater than 0.000 (p = 0.069 to 0.904). Consequently, the null hypothesis was accepted and inferred that GDT items among Lower Basic Three Pupils in Plateau Central Geo-political Zone did not function differently based on school type.

Discussion

The finding from research question one reveals that 29(96.67%) of the items fit the model with the exception of 1 (3.33%) that did not fit the 3PLM indicating that all the items of the lower basic GDT were valid. This finding corroborates that of Chamundeswari (2014), the American Educational Research Association (2020) and Wanlor (2025) who found that all the items of the instrument they developed fit

the models used. Similarly, the finding from research question two reveals that 50% of the items were of moderate level of difficulty, 80% have high discrimination and 96.67% of the items have low pseudo guessing parameter below 0.25. These items possess excellent discrimination, are of moderate difficulty and very low guessing parameter which are in tandem with Wanlor (2025) findings that 64% of the items were within the acceptable difficulty level,74% of items discriminated well and 98% of them had good guessing index value of 0.25 and below. Thus, the items of the test have quality

Finally, the analysis of the hypothesis tested indicated that the probability value (p-value) of the z-test for all the items are greater than 0.000 (p = 0.069 to 0.904). This implies that none of the items of the LBETPGDT did not function differentially based on school type which is in line with that of Gyamfi (2023) who found that there was no presence of DIF in the items for types of schools (public or private). from Western region of Ghana and contrasted that of Alade, Aletan and Sokenu (2020) who found that all the 50 items of the 2018 WASSCE Mathematics achievement test in Lagos State functioned deferentially based on school type while findings of Mawak, Mustapha and Wakjissa (2019) showed that some items show evidence of DIF based on school type.

Conclusion

The study made available to Lower Basic Education mathematics teachers geometry items that have good item fit, provide them with high quality items in terms of difficulty, discriminating and guessing parameters obtained through the Use of the 3PLM of IRT. None of these items were found to exhibit DIF based on school type. It can help in informing the Plateau State Universal Basic Education Board of the necessity to develop, calibrate and store items of diagnostic value for effective assessment and instructional strategies and also help the Lower Basic Education mathematics teachers to identify pupils with geometry learning misconceptions and to organize remedial teaching to those that need it.

Recommendations

The following recommendations were done based on the research questions and hypothesis:

1. Lower Basic Education teachers in Plateau State should always find the validity of the instrument they develop through the use of the item fit statistics

3. Lower Basic Education teachers in Plateau State should always ensure that they find the difficulty, discrimination and guessing parameters of test items they developed in order to ensure that all the items are of good quality

4. PSUBEB should embark on the training and retraining of Lower Basic Education teachers on how to developed quality diagnostic test that can guide them to organize remedial instruction for pupils.

References

Alade, O. M., Aletan, S. & Sokenu, B. S. (2020). *Assessing the differential Item Functioning of 2018 WASSCE Mathematics achievement tests in Lagos State*, *Nigeria.* AJB-SDR, 2 (2), 8-18.

American Educational Research Association. (2014). *Standards for educational and psychological testing*. Wasshington, DC: American Educational Research Association.

Chamundeswari, S. (2014). Conceptual errors encountered in mathematical operations in algebra among students at the secondary level. *International Journal of Innovative Science, Engineering & Technology, 1*(8), 24 – 35.

Dadughun, S. I. (2015). *Development and calibration of a Primary School Mathematics diagnostic test based on item response theory* [Doctoral thesis, University of Nigeria, Nsukka, Nigeria].

Fabiyi, T. R. (2017). Geometry concepts in mathematics perceived difficult to learn by senior secondary school students in Ekiti State, Nigeria. *Journal of Research and Method in Education, 6* (1), 83-90.

Gyamfi, A. (2023). Differential Item Functioning of Performance-Based Assessment in Mathematics for high schools. *Jurnal Evaluasi dan Pembelajaran, 5*(1), 20 – 30. Retrievedfrom https://jepjurnal.stkipalitb.ac.id/index.php/hepi.

Hodanova, J. & Nocar, D. (2016). Mathematics importance in our life. *Proceedings of INTED Conference, Valencia, Spain*. Retrieved from https://www.researchgate.netpuplication/298705287.

Mawak, J. J., Mustapha, A. Y. & Wakjissa, S. G. (2019). Demographic characteristic and differential item functioning in economics achievement test in Plateau State, Nigeria. *Journal of Humanities, 4*(2), 53-69.

Osadebe, P. U. & Agbure, B. (2018). Assessment of differential item functioning in social studiesmultiple-choice questions in basic education certificate examination. *European Journal of Education Studies, 4(*9), 236-251. Retrieved from www.oapub.org/edu.doi: 10.5291/zenodo.1301272.

Svetina, D. & Rutkowski, L. (2014). Detecting differential item functioning using generalized logistic regression in the context of large-scale assessments. *An IERI – International Educational Research Institute Journal,2*(4), 1-17.

Retrieved from https://doi.org/10.1186/s40536-014-0004-5.

West African Examination Council (WAEC, 2017 - 2024). *Chief examiners annual reports.* Retrieved from https://www.waecgh.org>examiner.

Wanlor, S.B. (2025). Item quality and gender-based differential functioning in geometry diagnostic test for middle basic pupils in Plateau State, Nigeria *Asian journal of Education and Social Studies. 51*(8), 794-804. Retrieved from: https://doi.org/10.9734/ajess/2025/v51i82278

Wanlor, S.B., Mustapha, A. Y., & Wakjissa, S. G. (2025). Assessing the geometry proficiency of upper basic education students for sustainable national development in Plateau State, Nigeria. *Greener Journal Educational Research. 15*(1), 53-59. Retrieved from https://gjournals.org/GJER

Wanlor, S.B., Wakjissa, S. G. & Mustapha, A. Y. (2025). Development and calibration of a geometry diagnostic test for upper basic education students in Plateau State, Nigeria. International *journal of research and innovation in social science*. Retrieved from https://dx.doi.org/10,47772/IJRISS.2025.90300202