ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

The African Journal of Behavioural and Scale Development Research
AJB-SDR Vol. 4, No 1, 2022

Examining Item Biasness in Terminal Unified Mathematics Examination for Primary Schools in Lagos State

Rufai, Adewale Taiwo[1], Daniel Olutola Oyeniran[2]
Lagos State Universal Basic Education Board (LSUBEB) Maryland Schools Complex, Maryland, Ikeja, Nigeria[1]
Institute of Education, University of Ibadan[2]

Abstract
Item biasness analysis is an important element in assessing examination fairness and validity. The study examined the occurrence of item biasness in Lagos State Terminal Unified Mathematics examination for primary schools pupils. The examination contained 40 items of multiple choice which was developed by the examination unit of Universal Basic Education Board, Lagos State using primary 5 mathematics curriculum. Primary data of scores of 2018 primary 5 Mathematics 2nd Term Unified Examination was used for this research. A sample of 640 pupils selected through multistage sampling technique was used in the study. Two research questions guided this study. Data was analysed using descriptive and inferential statistics. The results displayed closeness of the mean and standard deviation scores for examinee groups, indicating the examinees have similar ability levels. Out of the 40 items, 16 and 17 items function differentially among the examinees based on gender and age respectively while differential item functioning cannot be estimated based on location because it was found to comprise examinees with different ability level because carrying out Differential Item Functioning (DIF) with groups with different ability level will distort the result. The results displayed a high level of DIF. Based on the findings, it can be concluded that the examination has high differentially functioning items and the overall fairness shows biasness in the items. In addition to other psychometric measures, bodies in charge of state or nationwide examinations should ensure items biasness are checked and corrected before being administered for actual assessment.

Keywords: Differential Item Functioning (DIF),Mantel-Haenszel method, Examination bias, Item bias, Unified Terminal Examination, Unified Examination, Mathematics

Introduction
Mathematics has been found to be the science of structure, order, numbers, space and quantity; it is a relationship which revolves around the elementary practice of counting, measuring and describing of shapes and objects (Soyemi, 2002). Maduabum and Odili (2006) defined Mathematics as the science of quantity and space. It has occupied a key position in the Nigerian Educational System. Mathematics is an important foundation which propels advancement of technology in Nigeria. Subsequently, mathematics is a core to enter technologically-based courses either in the university, polytechnic or college of education (Adebule, 2013).

Examination in mathematics is a vital aspect of the teaching and learning technique of mathematics education which allows the mathematics teacher to evaluate mathematics learners during and at the end of mathematics lessons. Examinations are routinely used for diverse purposes, such as selection, placement, recruitment, diagnosis, promotion among others by means of valid and reliable measurement instrument.

As regards how examination scores are used to make inferences about examinees' performance, it is essential to ensure fairness and validity of the items used to obtain those scores. Faleye and Dibu-Ojerinde (2006) and Ajayi and Ighoroje (2000) noted that show of student's interest in courses related to science at the institutions of higher learning is on decline, especially among females. Meanwhile, male students appear to enjoy upper hand beyond their female colleagues in science and mathematics assessments while females have better achievement in verbal aptitude (Adebule, 2004). In line with Barret (2001) and Adebule (2013), objective examination items are found to be biased in favour of males while the females encounter more difficulties with examination items especially those involving mathematical, spatial or high-thinking abilities. In addition, it was detected that high or average examination

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

scores by certain groups is either due to actual ability differences or bias in examination items or a combination of both (Lee, 1999).

Main goal of measurement and assessment in education is to make right and appropriate decisions for individuals. Since assessment is used to make some predictions about individuals, it is expedient that test should be valid and reliable and do so over time. Determining differential item functioning is a way of strengthening examination validity and reliability.

Item biasness refers to an item that functions differentially among students in different groups with identical abilities. That is, it is found to be exceptionally difficult or easy for members of a particular group. Henceforth, these items will be referred to as DIF items(Levi-Keren, 2016).According to Ojerinde, Popoola, Ojo & Onyeneho (2012), differential item functioning (DIF) is a condition when an item functions differently for respondents from one group to another. In other words, respondents with similar levels on a latent trait(ability/skill) but who belong to different populations, have a different probability of responding to an item. If, for example, in a mathematics examination, boys display higher probability of answering correctly more often than girls of equal ability level because the contents in the examination items are biased against girls, then the items are said to exhibit DIF and should be considered for modification or removal from the test. The DIF analysis is an essential effort in promoting the validity and fairness of the examination. Establishing DIF status of an item has been established to be very crucial in the process of ascertaining validity and reliability of standardized examination items. Differential item functioning can therefore be understood as a conditional dependence between an item response and group examination-taker belongs which depends on variable of concern given that there is same latent ability.

An item cannot be said to show item biasness if individuals from different groups have a different probability to give a specific response; it shows item biasness can only be determined if individuals from different groups with equivalent ability have a unique chance of selecting a specific option (Camilli & Shepard, 1994).

Differential Item Functioning (DIF) is otherwise referred to as examination bias and unfairness. According to Adebule (2009), examination bias is realised due to certain decision, depending, to some extent, on if scores resulting from an examination is unfair or display incongruent effect on a particular group than the other. Gregory, (2006) opined that an examination can be described to be biased if the features of the estimation, the process involved in measurement or the inferences from the results of measurement cause inaccurate generalization concerning the psychomotor, cognitive and other affective aspect of a person or a set of examinees. There is absence of measurement bias in tests if the connection between latent trait and the observed scores which is being measured by the examination are similar across different groups (Stark, et al., 2001). Main differences that could be observed is due to differences among the sub-groups in the construct being measured. Stark et al (2001) asserted that differential item functioning is proven through violations of measurement equivalence. Lee (1999) presents that to examine bias at the level of item, large-scale assessment developers traditionally examine functioning of the items. Differential item functioning is a method of detecting systematic achievement disparities between persons with similar hidden genuine ability from different sub-groups (Camilli & Shepard, 1994; Raju & Ellis, 2002).

Investigating DIF is important, especially in the case of unified terminal examination for primary pupils to ensure cohesion and integration of education curriculum in the Universal Basic Education (UBE). Focus of most statistical procedures for determining DIF is increased validity and reliability of the examinations. In this regard, it seems crucial that studies on DIF should be carried out for the unified terminal exams that have a huge impact on the future of primary school children.

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

Mathematics plays a vital role in promoting science and technology across the globe and specifically in Nigeria and it is one of the means through which learners and schools are assessed. Even though multiple-choice examination is used as a common form of assessment for measuring pupils' achievement in mathematics in Lagos State Universal Basic Education Board, no evidence exists that the examination items used for the unified exam satisfy the psychometric property of fairness between different groups of examinees. In view of this, the study tends to examine whether Differential Item Functioning (DIF) exists in Terminal Unified Mathematics Examination for primary schools in Lagos State.

Purpose of the Study

Focus point of the study was to examine the occurrence of item biasness in Terminal Unified Mathematics Examination for primary schools in Lagos State. Specifically, the study seeks to determine:

1. Comparability level of the examinees in the various groups in Lagos State Terminal Unified Mathematics Examination for primary schools.
2. The occurrence of DIF in Lagos State Terminal Unified Mathematics Examination for primary schools based on Gender, Age and Location of the pupils.

Research Questions

Based on the problem stated above, the study sought to provide answers to the research questions below.

a. Will the examinees in the different groups be of similar levels of ability?
b. Does DIF exist in Lagos State Terminal Unified Mathematics Examination for primary schools based on Gender, Age and Location of the pupils?

Methodology

Ex-post facto descriptive design was adopted for this study because existing variables were examined in the study. The variables investigated were gender (male and female), age (11 years & below and above 11 years) and location (Lagos city and Lagos less city).

Secondary data was used for this study which comprises scores of primary 5 Mathematics 2nd Term Unified Examination of Lagos State Universal Basic Examination Board (LSUBEB), March 2018. Existing data for two Local Government Education Authorities each from two different geographical locations of Lagos State which comprised Lagos Less City and Lagos City was retrieved from the examining body. This makes a total of four Local Government Education Authorities for the study.

A simple random sampling was then used to select data from five primary schools from the data retrieved from four selected Local Government Education Authorities. The examination by Primary 5 pupils were used in each of the randomly-selected primary schools to make a total of 640 samples for the study.

The data was retrieved through photocopying of Optical Mark Reader (OMR) sheets for answering the examination. The Lagos State Terminal Unified Mathematics Examination contained 40 multiple choice items. Validation and standardization of the items were already carried out by the examination unit of LSUBEB with established reliability estimate of 0.91. Therefore, the instrument is very good based on instructional objectives and curriculum contents.

STATA package and Excel facility provided by the Microsoft office package was used for analysis purpose. Descriptive statistics was employed to show means and standard deviations of scores for the different groups of examinees in order to indicate whether the examinees are of comparable ability levels and Mantel-Haenszel method was used to check status of DIF in the items and items with 0.05 level of significant was adjudged to have DIF in Terminal Unified Mathematics Examination for primary schools in Lagos State.

Results

Research question one: Are the examinees in the different groups of similar levels of ability?

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)
The African Journal of Behavioural and Scale Development Research
AJB-SDR Vol. 4, No 1, 2022

***Table 1: Score d*istribution of Lagos State Summative Unified Mathematics Examination**

Scores	Gender		Age		Location		Total
	M	F	11 years and below	Greater than 11 years	Lagos Less City	Lagos City	
0-19	104	136	180	60	108	132	240
20–30	92	96	116	72	36	152	188
31- 40	80	132	152	60	156	56	212
Total	276	364	448	192	300	340	640
Mean	23.0	24.38	23.60	24.23	25.43	22.34	23.79
StD.	8.94	9.38	9.22	9.20	10.83	7.20	9.21
p-value	0.059		0.427		<0.001		

Table 1 shows closeness in the mean scores of examinees based on gender (t = -1.88, p = 0.059) and age group (t = -0.794, p = 0.427), however, there is a larger gap between the mean scores based on location (t = 4.288, p < 0.001). This indicates that the examinees are of similar ability levels when considering gender and age group but not on location. Hence, differential item functioning can only be performed based on gender and age group and not on location since the ability level is not similar.

Research question 2: Does DIF exist in Terminal Unified Mathematics Examination for primary schools in Lagos State?

Table 2. Chi-square Method for Detecting DIF in Terminal Unified Mathematics Examination for primary schools in Lagos State

Item	Gender		Age	
	Chi^2	Prob.	Chi^2	Prob.
1	0.01	0.9311	2.08	0.1496
2	0.19	0.6621	4.17*	0.041
3	8.52*	0.0035	14.40*	0.0001
4	14.65*	0.0001	0.58	0.4466
5	5.37*	0.0205	17.87*	0.00
6	0	0.9481	1.82	0.1775
7	4.02*	0.0451	11.68*	0.0006
8	1.38	0.2406	0.62	0.4326
9	10.08*	0.0015	1.71	0.1912
10	1.83	0.1759	1.09	0.2974
11	26.62*	0.00	20.20*	0.00
12	0.53	0.4655	4.46*	0.0347
13	0	0.9702	5.48*	0.0193
14	13.11*	0.0003	0.01	0.9216
15	3.88*	0.0488	0.01	0.9167
16	0.3	0.5823	10.06*	0.0015
17	2.34	0.1264	0.04	0.834
18	6.34*	0.0118	7.98*	0.0047
19	1.19	0.2751	1.17	0.28
20	8.48*	0.0036	4.31*	0.0379
21	1.3	0.255	4.86*	0.0275
22	0.01	0.9116	1.50	0.2206
23	8.28*	0.004	4.70*	0.0301
24	0.07	0.797	0.00	0.957
25	0.64	0.4224	6.87*	0.0088
26	20.56*	0.00	1.78	0.1827
27	3.52	0.0605	2.40	0.1211
28	0.37	0.5429	4.95*	0.0261
29	3.3	0.0694	0.20	0.6522
30	3.57	0.059	0.04	0.8446
31	18.29*	0.00	33.31*	0.00
32	3.44	0.0636	16.79*	0.00
33	5.71*	0.0169	11.14*	0.0008
34	1.58	0.2089	3.13	0.077
35	0.25	0.62	0.22	0.6388
36	0.13	0.7186	1.40	0.2367
37	1.55	0.2136	0.64	0.4227
38	1.03	0.3101	0.40	0.528
39	5.52*	0.0188	2.66	0.1031
40	9.22*	0.0024	0.00	0.9508

*** DIF items (p≤0.05)**

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

Table 2 shows items analysis to detect items with DIF. For item to flag DIF, the Mantel-Haenszel chi-square coefficient significant value must be less than 0.05. When this is applied to Unified Mathematics Multiple-Choice Examination items, the coefficient significant value less than 0.05 flagged under gender were items 3, 4, 5, 7, 9, 11, 14, 15, 18, 20, 23, 26, 31, 33, 39 and 40 while for age items identified to have DIF are 2, 3, 5, 7, 11, 12, 13, 16, 18, 20, 21, 23, 25, 28, 31, 32, and 33 representing 40% and 42.5% respectively of the 40 items displaying DIF.

Conclusion can therefore be made that the Lagos State Terminal Unified Mathematics Examination for primary schools exhibited about average DIF items and the overall fairness should be improved to ensure the results from the items truly shows the knowledge of the examinees without bias based on any criterion used.

Discussion

Unified Terminal Examination is a long-time policy of State Universal Basic Education Board (SUBEB). It is an examination that enables cohesion and integration of education curriculum. It is designed to provide equal level ground of assessment for all learners and to achieve a better academic outcome in placement examination both internally and externally. This grand idea is aimed and directed towards evaluating students' achievement by means of valid and reliable examination items. Meanwhile, items biasness will deny the students opportunity for demonstrating skills, knowledge and attitudes acquired in the teaching and learning process make them have lower score than their actual ability.

It is found through the study that items administered during the Unified Terminal examination in Lagos State especially as it relates with this study to be biased based on gender and age. The result is in resonance with the study of Adebule (2013) who established that the examinees in different groups (gender and age) are of equal levels of ability. The examinees, regardless of their numbers in each sub-group, their scores intervals are similar across different groups which indicate equal levels of ability.

The results displayed 14 items under gender and 17 items under age representing 40% and 42.5% respectively of the 40 items displaying DIF which shows absence of fairness of the examination based on gender which is more of female and age of the pupils which are those below 13 years. Thus, it could be confirmed that Unified Terminal Mathematics Examination in Lagos State is not free from differentially functioning items for the period considered in this study. Hence, the result agrees with Braimah (2011) who reported biased items in the 2005 to 2007 Unified Mathematics Examination of Edo state.

Given the high proportion of items differential functioning in the examination, there is going to be a certain disparity in the final scores obtained by the pupils which will influence their final placement and opportunities and therefore their trajectory in their educational endeavours. The pupils might get discouraged in themselves for their score in the subject and accept their low ability in Mathematics as suggested by scores from the examination.

Conclusions

It can therefore be said that Unified Terminal Mathematics Examination of Lagos State is not free from differentially functioning items and the examination function differentially between the examinees based on sub-groups of gender and age which lead to bias and report of results that are not true representative of the student's ability.

Recommendations

Since examination is used to establish learners' success, it is important that such assessment tool be valid, reliable and free of item bias. Determining differential item functioning is a way of strengthening examination validity, reliability and fairness of the items. Meanwhile, examination test fairness can be checked, it must first be ascertained that the examinees have equivalent ability level. Therefore, examination practitioners in the examination bodies, ministry of education and people concerned with

ISSN: 2714-2965 (Print)
ISSN: 2714-3449 (Online)

examination development should endeavour to ascertain items are calibrated appropriately and carry out DIF analysis from a pilot study before administration of examination so that items that exhibit differential functioning for different groups of examination takers can be identified for possible corrections and/or a replacement.

References

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 www.iiste.orgISSN 2222-1735 (Paper) ISSN 2222-288X (Online)Vol.4, No.17, 2013

_____ (2009). Identification of biased items in school certificate Mathematics achievement test. *International Journal of Research in Education*. *1*(2): 69—73.

_____ (2004). Gender difference on a locally standardised rating scale in Mathematics. *Nigerian Journal of Counselling and Applied Psychology*. *2*(1): 22—29.

Ajayi, O. B. &Ighoroje, A. D. (2000). Female enrolment for information technology training in Nigeria. Available online at www.wigsat.org. Retrieved on 20/6/2021.

Barret, S. (2001). Differential item functioning: A case study of first year Economics students. *International Education Journal*. *2*(3).

Braimah, M. (2011). A study of differential item functioning in Mathematics examination on selected secondary school students in Edo State. An unpublished M.Ed. proposal, University of Ado-Ekiti.

Camilli, G., & Shepard, L. A. (1994). Methods for identifying biased items. Thousand Oaks, CA: Sage.

Echernacht, G. (1974). A quick method of determining test bias. *Educational and Psychological Measurement*. *34*(2): 271—280.

Faleye, B. A. & Dibu-Ojerinde, O. O. (2006). A review of the enrolment and performance of male and female students in Education Economics programme Obafemi Awolowo University, Ile-Ife. *Journal of Social Sciences*. *12*(2): 143—146.

Gregory, R. J. (2006). Psychological testing: History, principles, and applications. 4th edition. India: Pearson Education Inc.

Lee, V. (1999). Identifying differential item functioning on the NELS history achievement test. Available online at www.cse.ucla.edu. Retrieved on 24/6/2021.

Levi-Keren, M. (2016). Bias factors in mathematics achievement tests among Israeli students from the Former Soviet Union. *Cogent Education* (2016), 3: 1171423http://dx.doi.org/10.1080/2331186X.2016.1171423

Maduabum, M.A. and Odili, G.A. (2006). Analysis of students' performance in General Mathematics at the Senior School Certificate Level in Nigeria from 1991-2002. *Journal of Research in Curriculum and Teaching*. *1*(1), 64-68.

Ojerinde, D., Popoola, K., Ojo, F., Onyeneho, P. (2012). Introduction to Item Response Theory Parameter Models, Estimation and Application. Abuja: Marvelous Mike PressLtd.

Raju, N. S., & Ellis, B. B. (2002). Differential item and test functioning. In F. Drasgow & N. Schmitt (Eds.), Measuring and analyzing behavior in organizations: Advances in measurement and data analysis (pp. 156188). San Francisco: Jossey-Bass.

Soyemi, B.O. (2002). Student Dictionary of Mathematics. *Nigeria Educare Foundation*.

Stark, S.; Chemyshenko, S.; Chuan, D.; Lee, V. & Wadlington, P. (2001). Detection of differential item/test functioning using IRT. Available online at www.umu.se. Retrieved on 20/6/2021.