Determining the differential item functioning of non-verbal intelligence test using item response theory in Rivers State

Ogidi Reuben C. Ph.D
Department of Educational Psychology, Guidance and Counselling
Ignatius Ajaru University of Education, Rumuolumeni, Port Harcourt

Abstract

The study investigated the Differential Item Functioning of Non-Nérbal Intelligence Test using Item Response theory. The study adopted non-experimental research design. The population of the study consisted of445,231 Basic 5 pupils in public and private primary schools in Rivers East Senatorial District aged 9 years in the 2019/2020 academic session. The sample of the study comprised 800 pupils drawn through multi-stage procedure. The instrument for data collection was a non-verbal intelligence test standardized by the researcher in 2007. The instrument contained four sections: Figure Classification, Figure Analogies, Figure Synthesis and Matrices. Ruder-Richardson (K-R20) was used in determining the reliability of the instrument. The instrument yielded reliability coefficient of 0.71, 0.75, 0.70 and 0.73 respectively. Item Response Theory Patient Outcome (IRTPRO) was used in determining differential item functioning using method of Item Response Theory. The result of this study showed that items in Figure Classification and Figure Analogies had no DIF effect, some items in Figure Synthesis and matrices had DIF effect. Based on the results, recommendations were ma de that DIF should be used by test developers to eliminate bias amongst others.

Keyword: Differential Item Functioning, Intelligence Test, Nonverbal Intelligence Test

Introduction

Tests are very useful instruments in teaching and learning. They provide systematic procedure for observing a person's behaviour and describing it by means of a numerical scale or a category system. Astuu (2015) posits that test is a device or instrument for obtaining a sample of pupils' behaviour and is made up of a set of tasks, questions or problems intended to elicit particular types of behaviour. It may be designed to measure different areas of life of an individual such as competence, reasoning, skills, knowledge, achievement, ability, interests and attitude. Ali, Ezeadi and Ogbazi in Chikwe (2017) affirm that test is an instrument administered to determine the presence or absence of phenomenon being measured.

Inko-Tariah and Ogidi (2017) posit that test is a task or series of questions presented to an individual or a group of individuals in order to obtain the characteristics or traits possessed by them. It is a systematic and objective method used in obtaining a sample of an individual's behaviour, qualities or that of an object.
Ferguson in Ukwuije and Okpara (2012) affirm that test is a systematic procedure for comparing the behaviour of two or more persons. Test is an objective sample of some aspect of behaviour (Sheffzer & Stone in Ordu, 2016). Objectivity, in this context, refers to the requirements that its administration, scoring and interpretation are independent of the individual tester's subjective judgement. National council on Measurement in Education in Inko-Ta1iah and Ogidi (2017) affirms that test is a set of tasks or questions intended to elicit particular types of behaviour when presented under standardized conditions and to yield scores that have desirable psychometric properties.

It is used to measure and compare students' aptitude and achievement in various skills and subjects. Test scores help to determine the relative standing of schools regarding students' achievement in order to diagnose learning difficulties of students, to improve the teachers' instruction and to motivate the students to study. It is used to determine students' readiness to learn particularly in new classes. Information obtained from test may enable teachers modify their instructional strategies and also enable students develop and sustain their interest in school related activities. Asuru (2015) posit' that test is used to determine learners' strengths and weaknesses.

There are several types of test in the school system ranging from achievement, aptitude, personality, intelligence test etc. Aiken in Ogidi (2007) posit that intelligence tests are designed to measure an individuals' aptitude for scholastic work or other kinds of occupation requiring reasoning, verbal ability or the manipulation of objects. The items on intelligence tests represent attempts to assess individual differences in the effects of experiences common to nearly everyone in the culture. Intelligence test scores reflect how people resolve difficulties and make decisions about their lives amongst others.

Intelligence test may be classified into individual, group, culture fair, culture biased, non-verbal etc. Non-verbal intelligence tests are

developed without verbal symbolism (Ogidi, 2007). CllWevve (2017) posits that non-verbal intelligence tests do not require language and are not written in words rather such test are presented using figures. Onukwo (2002), InkoTariah and Ogidi (2017) affirm that non-verbal intelligence tests require the ability of an individual to process information, reason abstractly and solve problems without word expression although words can be used for writing the instructions of such a test. Walsu (2014) suggests that non-verbal intelligence test involves visual reasoning and the items may include recognizing visual sequence and noting relations between objects.

Ogidi (2007) posits that non-verbal intelligence tests eliminate cultural bias that affects the acceptability of verbal intelligence tests which have been criticized as being culturally loaded. Also, it is a more valid long-term measure of school potential of the low, medium and high achievers than verbal and performance tests. It is a legitimate aid in determining the range and strength of an individual's cognitive abilities, particularly for career planning. Non-verbal intelligence tasks tap a set of thinking skills basic to intellectual functioning and measures general intelligence. It comes handy while testing children because the items of such tests are unique in their appeal to children who generally exhibit delight in taking them. The importance of non-verbal intelligence is very obvious in studying children having language limitations or reading deficiencies.

Raby in Ogidi (2007) suggests that the following conditions be considered in constructing or designing a non-verbal intelligence test (i) such a test should consist of items that do not assume special training (ii) the test items should not take too long to solve. This means that the duration of the test of the testing of the instrument should be determined (iii) test items of the non-verbal intelligence should be devised so as to have a single correct answer (iv) test items of a nonverbal intelligence tests should follow the laws of neo-genesis. Neo-genesis means the production of new or novel contents, based on the relations observed between the elements of a given problem. This implies the miles of education of relations and the education of correlates (v) the age of the learners that the test is designed for should be specified. This is because non-verbal intelligence tests are age dependent.

For many years, measurement experts have attempted to find solution to the problem of bias of test items. A test is fair if it allows testees an equal opportunity to show the skills and knowledge they have acquired and which are of interest to the objectives of the test. There is therefore the need to solve the problem of test bias. This is especially due to the fact that test is of prime importance in instruction and as such test items should be fair to testees. Ogbebor and Onuka (2013) explained that bias is the existence of some in-relevant elements present in the items that causing differential performance for testees of the same ability but different demographic settings. Zumbo (1999) and Ogbebor (2012) posit that tests are considered biased because they contain sources of difficulty that are not relevant to the construct being measured and these extraneous sources affects test taker's performance. Test items are expected to be unbiased. Ojerinde, et al (2012) posit that a test is unbiased if the item difficulty index for one group is the same as that for the second group of interest Ojefinde et al (2013) explained that test fairness is a moral imperative for both the test takers and users. Thus, item bias and Differential Item Functioning (DIF) are similar. However, they are not the same. This is because Differential Item Functioning may occur without the judgement of unfairness. Differential Item Functioning analysis is one of the several processes that are used to ensure that items are free from bias (Bulus, 2018). It is utilized to investigate how items function in various sub-groups. Ogbebor and Onuka (2013) affirm that Differential Item Functioning is a statistical procedure used to assess the existence of item bias. Omorogiuwa and Iro-Aghedo (2016) suggest that DIF occurs when testees from different groups in a population have different likelihood of success on an item after they have been matched on the ability of interest. They explained that differential item functioning of an item can be understood as a lack of conditional independence between an item response and group membership given the same latent ability or trait.

Lyons-Thomas, et al (2014) posit that it is important to analyze whether items have DIF for two reasons: (i) that the presence of DIF signals potential bias and bias has an impact on the validity of inferences drawn from group comparisons. Therefore, DIF items, if confined to represent underlying bias, are often removed from future administrations of a test (ii) that items that exhibit DIF may have implications for curriculum and instruction especially if no reason for bias can be found. Omorogiuwa and Iro-Aghedo (2016) suggest that differential item functioning occurs when people from different groups with the same ability have systematically different responses to specific test items.

Several researchers (Adediwura, 2013; Obinne & Amali, 2014; Ani, 2014; Ahmad & Bazvard 2016; Bulus, 2018) suggested that DIF is an important tool in identifying the extent to which an item is measuring different abilities for number of subgroups. Abedalaziz (2010) investigated gender-related differential item functioning of mathematics test items. The researcher determined the DIF of mathematics items and concluded that the percentage of agreement among the three approaches in detecting DIF are relatively low.

Omorogiuwa and Iro-Aghedo (2016) investigated DIF by gender in National Business and Technical Examinations Board (NABTEB) in 2015 Mathematics Multiple-Choice Test Items (Dichotomous). The results of the analysis indicated that male and female testees fimctioned differently in 17 items (representing 34%), on the other hand, there was no difference in 33 items (representing 66%).

Lyons-Thomas, et al (2014) examined gender differential items function (DIF) across four jurisdictions that took part in a large-scale international assessment in Canada, Shangai, Finland and Turkey. They observed that some items performed differentially among the testees from the different countries (six items representing 12 percent had DIF effect while the other 44 items representing 88 percent had no DIF effect).

Ogbebor and Onuka (2013) investigated differential item functioning method as an item bias indicator. They used logistic regression statistics to identify items that have DIF against sub-groups such as public and private schools and urban and rural areas and discovered that eleven items favoured public schools while eleven items also favoured private schools.

Statement of the Problem

Several studies (Abedalaziz, 20 10; Omorogiuwa and Iro-Aghedo, 2016) related to differential item functioning were mainly based on the performance of students in various achievement test. There seems to be a paucity of works on determining DIF of non-verbal intelligence test because the researcher is yet to come across any. This to forms the major enthusiasm for this study. Also, the non-verbal intelligence test is an indigenous-tests. It is therefore the desire of the researcher to determine the DIF of the test based on school type, location and gender. Different techniques have been utilized to determine DIF.

Purpose of the Study

The main purpose of this study is to determine the DIF of the non-verbal intelligence test. Specifically, the study sought to:

1. investigate whether the items of nonverbal intelligence test do not function differentially for male and female pupils.
2. determine whether the items of nonverbal intelligence test do not function differentially for pupils from urban and rural areas in Rivers State.
3. examine whether the items of nonverbal intelligence test do not function differentially from pupils in public and private primary schools.

Hypotheses

Based on the purpose of the study, the researcher formulated the following null hypotheses.

1. Items of the non-verbal intelligence test do not function differentially for male and female pupils.

2. Items of the non-verbal intelligence test do not function differentially for pupils from schools in urban and rural areas.
3. Items of the non-verbal intelligence test do not function differentially for pupils from public and private schools.

Methods

The researcher adopted non-experimental research design. This is a type of research which describes "what is" by recording, analyzing and interpreting conditions that exist (Ajoku, 2006; Isangedighi, 2012; Nwankwo, 2013; Obilor, 2018). The process involves the collection of data in order to test certain hypotheses or answer research questions generated in the study.

The population of the study consisted of 445,231 (229,823 male and 215,408 female) Basic 5 Pupils in Public and Private Primal Y Schools in Rivers East Senatorial District. The area include: Port Harcourt City Local Government Area (PHALGA), Obia/Akpor Local Government Area (OBALGA), Emohua Local Government Area (EMOLGA), Ikwene Local Government Area (KELGA), Etche Local Government Area (ELGA), Omuma Local Government Area (OMOLGA), Okrika Local Government Area (OKLGA) and Ogu/Bolo Local Government Area. The study made used of in the study were aged 9 years in the 2019/2020 Academic Session. The researcher sampled 30 primary schools through multi-stage sampling procedure. First, simple random sampling was used to select Rivers East Senatorial District for the study. (Other senatorial districts in the state include: Rivers West and Rivers South-East senatorial districts). The researcher then employed stratified sampling technique and probability proportionate to size (PPS) to sample 800 pupils (429 male and 371 female pupils, 450 pupils from urban areas and 350 pupils from rural areas 461 pupils from public primary schools and 339 pupils from private primary schools) for the study from the various local government areas in the Senatorial District considering the gender of pupils, location and school type. The instrument for data collection was the non-verbal intelligence test developed and validated by the researcher. The instrument comprised four (4) sections: Figure Classification (25 items), Figure Analogies (25 items), Figure Synthesis (25 items) and Matrices (25 items). The items are arranged in the following order: Figure Classification I — 25, Figure Analogies 26-50, Figure Synthesis 51-75 and Matrices 76-100.

The instrument was validated using judgement of 15 experts in Measurement and Evaluation. Based on the scoring of items of the test influenced the pruning down of the items to 25 for each of the sections after analysis. KuderRichardson20 was utilized (with the aid of SPSS) in determining the reliability of the instrument. Each section of the instrument yielded the following reliability coefficient: Figure Classification 0.71, Figure Analogies = 0.75, Figure Synthesis 0.70 and Matrices 0.73. The instrument was administered through the assistance of 5 research assistants trained for the study. To forestall cheating, pupils sitting on the same row were given separate sections of the test during the administration of the test and vice versa. The administration of the instrument lasted for three months. The pupils' scripts were scored and the scores generated were subjected to item analysis. Item Response Theory Patient Report Outcome (IRTPRO) was used for determining differential item functioning, using Wald test method of Item Response Theory.

Results Table I: Differential Item Functioning (DIF) of Non-verbal Intelligence Test Items based on Gender

Item	Wald test	P	Bias Against	Item ID	Wald test	P	Bias Against	Item ID	Wald test	P	Bias Against
4	1.41	0.53		41	2.22	0.33		78	0.88	0.34	
5	0.85	0.84		42	2.08	0.55		79	1.45	0.40	
6	1.19	0.32		43	2.19	0.84		80	1.32	0.59	

7	1.17	0.19		44	3.11	0.30		81	2.43	0.33		
8	2.42	0.35		45	2.87	0.32		82	2.38	0.62		
9	2.58	0.4346		2.13	0.41	83	2.22	0.72	10 3.11	0.55	47	
	0.98	0.35	84	2.54	0.35	11 1.81	0.34	48	0.75	0.28	85	
		1.79	o. 73	122.59	0.42	49	1.59	0.95	861.63	0.35	13 1.88	
		0.22	50	1.68	0.47	87	1.54	0.61	14 3.17	0.90	51	
	1.78	0.45	88	1.33	o. 73	152.14	0.47	52	1.33	0.23	89	
										0.94	0.58	
16	2.26	0.46		53	2.44	0.56		90	1.05	0.48		
	171.93	0.32	54	2.21	0.48	91	1.28	0.46	18 1.86	0.53	55	
								2.96	0.65921.43		0.32	
19	I .49	0.11	56	2.11	0.31	93	1.69	0.35				
20	2.25	0.48	57	0.88	0.88	94	1.44	0.27				
21	2.21	0.68	58	0.43	0.76	95	1.32	0.29				
22	3.00	0.34	59	0.99	0.94	96	0.44	0.22				
23	.64	0.92	60	0.73	0.96	97	3.65	0.46				
24	0.36	0.97	61	1.03	0.36	98	0.82	0.41	25 2.51	0.74	62	3.38
		0.27	99	1.13	0.38							
26	1.81	0.32		63	4.41	0.44		100	1.74	0.31		
27	0.69	0.28		64	4.22	0.26						
28	1.84	0.35		65	3.18	0.28						
29	0.83	0.52		66	4.44	0.29						
30	3.41	0.36		67	0.59	0.24						
31	3.39	0.25		68	0.64	0.28						
32	3.68	0.26		69	1.05	0.26						
33	3.33	0.23		70	2.38	0.61						

34	0.31	0.25	71	1.27	0.52
35	I. 41	0.54	72	0.82	0.53
36	0.87	0.63	73	3.25	0.83
37	0.33	0.30	74	0.76	0.49
38	0.59	0.34	75	4.09	0.72
39	3.31	0.55	76	1.39	0.83
40	2.11	0.28	77	0.81	0.82

Information from analyzed data in Table I shows that three items (3%) were removed and were not calibrated. These items include items l, 2 and 3. However, the remaining 97 items (97%) were not biased against the male or female students. This revealed that no item functioned differentially against male and female pupils. The null hypothesis earlier stated is therefore accepted.

Table 2a: Differential Item Functioning (DIF) of Non-verbal Intelligence Test Items based on Location

Item ID	Wald test	P	Bias Against	Item ID	Wald test	P	Bias Against	Item ID	Wald Test	P	Bias Against
4	0.59	0.24	Bias Against	41	1.42	0.53	Bias Against	78	0.88	0.34	Bias Against
5	0.64	0.28		42	0.85	0.84		79	1.45	0.40	
6	1.05	0.26		43	1.19	0.32		80	1.32	0.59	
7	2.38	0.61		44	1.17	0.19		81	2.43	0.33	
8	1.27	0.52		45	2.42	0.35		82	2.38	0.62	
	0.82	0.53		46	2.58	0.43		83	0.16	0.00 urban	
10	325	0.83	47	3.11	0.55	84	2.54	0.35			
11	0.76	0.49	48	1.81	0.34	85	1.79	0.73			
12	4.09	0.72	49	2.59	0.42	86	1.63	0.35			
13	1.39	0.83	50	1.88	0.22	87	1.54	0.61			
14	0.81	0.82	51	1.64	0.92	88	2.21	0.48			
15	1.33	0.73	52	0.36	0.97	89	2.96	0.65			

16	0.94	0.58	53	2.51	0.74	90	2.11	0.31			
17	1.05	0.48	54	1.81	0.32	91	0.88	0.88			
18	1.28	0.46	55	0.69	028	92	0.43	0.76			
19	1.43	0.32	56	1.84	0.35	93	0.99	0.94			
20	1.69	0.35	57	0.83	0.52	94	0.73	0.96			
21	1.44	0.27	58	0.31	025	95	1.030.36				
22	1.32	0.29		59	1.41		0.5496			3.38	0.27
23	0.44	0.22		60	0.87	0.63			97	4.41	0.44
24	3.65	0.46		61	0.33	0.30			98	4.22	0.26
25	0.82	0.41	62	0.59	0.34	99	3.18	0.28			
26	1.13	0.38	63	3.31	0.55	100	4.44	0.29			
27	1.74	0.31	64	2.11	028						
28	3.17	0.90	65	2.22	0.33						
29	2.14	0.47	66	2.08	0.55						
30	2.26	0.46	67	2.19	0.84						
31	1.93	0.32	68	3.11	0.30						
32	1.86	0.53	69	0.16	0.00 urban						
	1.49 0.11 70 2.13 0.41 34 2.25 0.48 71 0.98 0.35										
35	2.21	0.68		72	0.75	0.28					
36	3.00	0.34		73	1.59	0.95					
37	3.41	0.36		74	1.68	0.47					
38	3.39	0.25		75	1.78	0.45					
39	3.68	0.26		76	1.33	0.23					
40	3.33	0.23		77	2.44	0.56					
3.33		0.23		77	2.44						0.56

Table 2b: DIF Items in Non-verbal Intelligence Test based on Location

NMT Application by School Location	Figure Classification (25	Figure Analogies	Figure Synthesis	Matrices

	items)	(25 items)	(25 items)	(25 items)
	IRT	IRT	IRT	
Biased against the pupils in urban areas	0	0	0	0
Biased Against the pupils in the rural areas	0	0	1	1
Total items with DIF	0(0%)	0(0%)	1(4%)	1(4%)
Total item without DIF	22(100%)	25(100%)	24(96%)	24(96%)

IRT

Tables 2a and 2b show that three items in Figure Classification (representing 12%) were removed and were not calibrated. These items are l, 2 and 3. Tables 2a and 2b revealed that no item in Figure Classification and Figure Analogies had DIF effect. However, item 9 in Figure Synthesis and item 8 in Matrices functioned differentially in favour of pupils who schooled in urban areas. Thus, one (l) item each in Figure Synthesis and Matrices functioned differentially against pupils in Rural areas in favour of pupils in urban areas. In a nutshell, Tables 2a and 2b showed that two items functioned differentially in favour of pupils who schooled schools in urban areas while the other items were free from DIF effect. The null hypothesis earlier stated is therefore rejected.

Table 3a: Differential Item Functioning (DIF) of Non-verbal Intelligence Test Items based on school type

Item ID	Wald Test		Blas Against	Item	Wald test		Bias Against	Item ID	Wald test		Bias Against
4	331	0.55		41	2.59	0.42		78	2.25	0.48	
5	2.11	0.28		42	1.88		0222.21	79			
6								80		222 0.33	43
7	0.923.00	0.34						81			
8								82		2.08 0.55	44
9	0.360.973.41	0.36						83			
								84		2.19 0.84	45
	2.510.74339	0.25						85			
								86		3.11 0_30	46
	1.81	0.323.68		0.26				87			
10	023	0_32	47	0.69	0.283.33	0.23		88			
11	2.13	0.41	48	0.350.43	0.76			89			
12	098	0_35	49	0.83	0.520.99	0.94		90			
13	0.75	0_28	50	0.31	0.250.73	0.96		91			
14	1.59	0.95	51	1.41	o .540.16	O _OOprivate		92			
15	168	0.47	52	0.87	0.633.38	0.27					
16	1_78	0.45	53	0.33	0.304.41	0_44					
17	133	0_23	54	0.14	0.00 private	4.22	0.26				
18	2.44	0.56	1.69	0.353.18	0.28						
19	0.59	024	56	1_44	0.27	93	4.44 0.29				
20	064	0.28	57	1.32	0.29	1.79	0.73				
21	1.05	026	O .44	0.22	95	1.63	0.35				

22	238	0_61	59365	0.461.54		0.61			96	
23	127	0.52	60	0.82		0.41		221	97	0.48
24	0.82	0.53	610.16	0.00 private	2.96	065			98	
25	325	0_83	1.74	0.312.11	0.31				99	
26	0.76	0_49	63	3.17	0.90	100	0_88			
	4.09	0.72		64	2_14	0_47				
28	139	0.83	65	226	0_46					
29	0.81	0.82	661.93	0_32						
30	133	0.73	671.86	0.53						
31	O .94	0.581.49	0.11 32	1.05	0_48	69I	.42	0.53		
33		128	0.46	70		0.85	0.84			
34		1_43	0_32	71	1.19	0.32				
35		0_34	72	1.17	0.19					
36		1.45	73	2_42	0.35					
	132	0.59		74	2.58	0.43				
38	2.43	0.33	75	3.11	0.55					
39	238	0.62	76	1.81	034					
40	051	0.71	77	2.54	0.35					

Table 3b: Set of DIF Items in Non-Verbal Intelligence Test based on School Type

NVIT Application by School Type	Figure Classification (25 items) IRT	Figure Analogies (25 items) IRT	Figure Synthesis (25 items) IRT	Matr ices (25 items) IRT
Biased against the pupils in Public schools Biased Against the pupils in Private school			2	
Total items with DIF Total item without DIF				

Tables 3a and 3b show that three items in Figure Classification representing (12%) were removed and were not calibrated. These items are l, 2 and 3. Tables 3a and 3b indicate that no item in Figure Classification and Figure Analogies had DIF effect. However, items 4 and Il in Figure Synthesis and item 13 in Matrices functioned differentially in favour of pupils from private schools. In a nutshell, Tables 3a and 3b reveals that 3 items functioned differentially in favour of pupils from private schools while the other items (95 items) were free from DIF effect. The null hypothesis earlier stated is therefore rejected.

Discussion of Results

The items of the non-verbal intelligence test did not function differentially for any of the gender. The 97 items that were calibrated were free from DIF effect. The finding of this study disagrees with Omorogiuwu and Iro-Aghedo (2016) that the items of the National Business and Technical Examination Board in 2015 functioned differently. However, the difference between the two studies is that the former study focused on achievement test while the present study is on non-verbal intelligence test (the researcher could not lay hand on any other study that discussed the DIF of non-verbal intelligence test).

Also the results of the study revealed that the items of the non-verbal intelligence test functioned differentially in favour of pupils from urban schools (3 items). In addition, the results of the study indicated that the items of the non-verbal intelligence test functioned differentially in favour of pupils in private primary schools. This revealed that few items in the non-verbal intelligence test were biased

against pupils in pnma1Y schools in Rural areas and public primary schools. This could be due to the fact that pupils in urban areas and in private schools have the likelihood of enjoying greater success in school-related activities. This result is in agreement with Ogbebor and Onuka (2013) that test items favour students in urban areas and in private schools.

Conclusion

Based on the results of the study, it was established that differential item functioning is a useful tool in detecting test bias not only in achievement test but also in intelligence test such as the non-verbal intelligence test.

Recommendations

Based on the results of the study, the following recommendations were made

1. It is recommended that measurement or test experts should acquaint themselves the opportunity of obtaining knowledge of the processes involved in DIF.
2. Test experts should explore the use of DIF in test development. This will make for fairness of the test items to different subgroups of test takers.
3. It is also necessary that test developers should consider the background of the test takers while developing tests in order to minimize test bias.
4. Test developers should consider developing standardized intelligence test in the education system rather than the present situation that skewed in favour of achievement test. The items of such standardized test should reflect life experiences relevant to the testees for which the test is designed for.

References

Adebalaziz, N. (2010). A gender related differential item functioning of Mathematics test items. The International Journal of Educational and Psychological Assessment, 5: 101-116.

Adediwura, A.A (2013). A comparative study of item response Theory and generalized linear model methods of detecting differential item functioning in dichotomous test. Research Journal in Organizational Psychology and Educational Studies, 2(6):308-316.

Ahmad, A. & Bazvand, A.D (2016). Gender differential item functioning on a national field-specific test: The case of PhD entrance exam of TEFL in Iran. Iranian Journal of Language Teaching Research, 491): 63-82.

Ajoku, L.I (2006). Foundations of educational research and statistics. Port Harcourt: Pearl publishers.

Ani, E.N (2014). Application of item response theory in the development and validation of multiple-choice tests in Economics. Retrieved on 8/01/2020 from http://www.sciedu.ca/itemresponsetheory-development-validation-of• multiple choice tests

Asuru, V.A (2015). Measurement and evaluation in education and psychology. PortHarcou1t: Pearl publishers Int'l.

Bulus, C.H (2018). Development and calibrating of a diagnostic test in quantitative aspect of economics in senior secondary schools in Plateau State, Nigeria. Unpublished PhD thesis, University of Jos.

Chikwe, C.K (2017). Fundamentals of test, measurement and evaluation in education. Port Harcourt: Emmanest ventures.

Ink0-Tariah, D.C & ogidi, R.C (2017). Fundamentals of psychological testing for psychologists, counsellors and educationists. Port Harcourt: Rodi printing and publishing company.

Isaagedighi, A.J (2012). Essentials of research and statistics in education and social sciences. Calabar: Eti-Nwa associates.

Lyons-Thomas, J., Sandilands, D. & Ercikan, K. (2014). Gender differential item functioning in Mathematics in four international jurisdictions. Education and Science,39(172):

Nwakwo, O.C (2013). A practical guide to research writing. Choba: University of PortHarcou1tPress.

Obilor, E.I (2018). Fundamentals of research methods and statistics in education and social sciences. Port Harcourt: SABCOS printers and publishers.

Obinne, A.D.E & Amali, A.O (2014). Differential item functioning: Implication for educational testing in Nigeria. International Review of Social Sciences and Humanities, 7(1):52-65.

Ogbebor, U. & Onuka, A. (2013). Differential item functioning method as an item bias indicator. Educational Research, 4(4): 367-373.

Ogbebor, U.C (2012). Differential item functioning of economics question paper of National Examinations Council in Delta State, Nigeria. Unpublished M-Ed thesis, University of Ibadan.

Ogidi, R.C (2007). Development and standardization of non-verbal intelligence test (WIT). Unpublished Ph.D thesis submitted to the Department of Educational Psychology, Guidance and Counselling, Faculty of Education, University of Port Harcourt, Port Harcourt.

Ojefinde, D., Popoola, K., Ojo, F. & Onyeneho, P. (2012). Introduction to item response theory: Parameter model, estimation and application. Abuja: Marvelous Mike press.

Omorogiuwa, K.O. & Iro-Aghedo, E.P (2016). Determination of differential item functioning by gender in the National Business and Technical Examination Board (NADTEB) 2015 Mathematics Multiple Choice Examination. International Journal of Education, Learning and Development, 4(10):25-35.

Onunkwo, G.I.N (2002). Fundamentals of educational measurement and evaluation. Oweni: Capepublishers Int'l.

Ordu, S.N (2016). Counselling psychology principles: Theories and practice. Port Harcourt: Rodi printing and publishing company.

Ukuije, R.I & opara, M.I (2012). Test and measurement for teachers. Port Harcourt: Chadik printing press.

Walsu, P.C (2014). Measurement and evaluation for schools and colleges. New York: Robson publishers.

Zumbo, B.D (1999). A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-like (ordinal) item scores. Ottawa, Canada: Directorate of Human Resources Research and Evaluation.