Gender Differences in Cognitive Intelligence Testing: An Exploration of the Differential Item Functioning among Adolescents in Nigeria

Taiwo Adebukola Kabir and Adeoye Oluremi Adedamola

Department of Counselling and Human Development Studies, University of Ibadan

DOI: https://dx.doi.org/10.47772/IJRISS.2023.70820

Received: 10 July 2023; Accepted: 15 July 2023; Published: 16 August 2023

ABSTRACT

This study was set to investigate gender differences in cognitive intelligence testing: an exploration of the differential item functioning among adolescents in Nigeria. Adopting the descriptive design of the survey type, the study sample consisted of secondary school students drawn using the multistage sampling procedure from Nigeria's six geopolitical zones. A five-section, 50-item cognitive intelligence test (? =.91), with sub-sections such as Attention to detail (0.90), Logic (0.86), Spatial management (0.83), Quantitative reasoning (0.84) and Associative memory (0.89) was used. DIF was determined through the DIF option of jMetrik version 4.2.1., with results revealing 9 items to be biased against testees on the basis of age, and 10, on the basis of gender. Following these results, it was recommended that test developers, reliant on the Item Response Theory (IRT) as the basis for test validation, should pay serious attention to the exploration of the DIF of items on the test, with a view to eliminating bias, and enriching the test for the purposes it was initially raised.

Keywords: Cognitive intelligence test, Differential item functioning, Adolescents in Nigeria

INTRODUCTION

Tests are at the heart of the teaching-learning process and play significantly pivotal roles. Asides from providing testers insight into the extent of learning achieved by testees and the ensuing opportunity to plan lesson guides, tests have also come to define how students learn, particularly how they and their teachers are motivated, and how teachers teach. Other forms of tests such as aptitude, vocational, personality, and intelligence tests, which could all be regarded as psychological tests, on their own part, are essentially deployed for diagnosis, job screening and placement, school admission, identification of specific behaviours, and psychometric research.

Tests, Ogidi (2020) posited, provides the systematic procedure used in the observation of an individual's behaviour, as well as its description by means of a numerical scale or category system, with Chikwe (2017) submitting that they have become instruments whose administration assists in the determination of the presence or otherwise of the construct under consideration. The position of these researchers aligns with that of Inko-Tariah and Ogidi (2017) who maintained that tests are a series of questions administered on a person or a group of persons in the bid to determine the traits present in the respondents.

In intelligence testing, the irreducible minimum, in the opinion of Herrnstein and Murray (1996), is that it is best measured as a cognitive ability, involving posing problems for people to solve, and examining resulting trends of correct answers, which researchers argued, are usually those identified by a specialisation's authorities, references, a panel of experts, or through consensus among testees (MacCann and Roberts, 2008). This maintains that intelligence testing is required in all areas of life.

The importance of cognitive tests in various spheres of life cannot be undermined. Intelligence tests have

been used to predict several important variables, which include academic achievement test scores, grades, and job performance among others. It has also been used to identify individuals with intellectual disability, learning disabilities, or intellectual giftedness (Kpolovie, 2015). Intelligence is pivotal to educational programmes (Mingat, Tan, and Sosale, 2003), and this implies that all academic activities require cognitive intelligence testing.

Cognitive intelligence is an important factor in national development (Taiwo and Ojuolape, 2023), noticeable in efficiency and effectiveness, innovation and creativity, increased productivity, and by extension, improved gross domestic product. With these visible contributions, any nation desirous of increased economic growth, productivity, income, wealth and wellbeing, will consciously institute strategies towards the fostering of talent and giftedness, remediation of mental retardation, as well as the diagnosis and cure of over and underachievement (Omoroguiwa and Iro-Aghedo, 2016). It is therefore not a surprise that the most developed nations are also the frontrunners in efforts directed at the discovery, development, and fostering of cognition.

Intelligence test developers, over time, have been found to deploy a combination of verbal and nonverbal domains, or have gone purely nonverbal. Cognitive Intelligence Test for Nigerian Adolescents (CITNA), is a test developed and validated for use in Nigeria (Taiwo and Adeoye, 2023). The authors argued, went nonverbal due to their desire to stick to the acceptable view of cognitive intelligence as an individual's capacity for abstract and efficient reasoning, and the need to downplay language proficiency.

This explanation goes with the addition that much as English remains the nation's official language, the varying levels of its proficiency among constituent ethnic nationalities make a blanket evaluation of the intellection of an ethnic group on the basis of its language proficiency (Taiwo and Adeoye, 2003). The researchers concluded that the cognitive intelligence test will be useful for testees, including those with speech, language, or hearing impairment, which is a fulfillment of the goals of inclusive education.

Notwithstanding the lofty intentions of the developers of CITNA, it has become imperative to explore the test for the presence of bias or otherwise. This is because the presence of bias Ogbebor and Onuka (2013) noted, leads to differential performance by testees of the same quality of traits, who for one reason or the other, belong to different demographic settings, due to the existence of irrelevant elements in the items.

The position, earlier arrived at by Ogbebor (2012), was that items are considered biased because they contain sources of difficulty that are patently irrelevant to the construct under measurement, with these extraneous variables directly impacting performance. The challenge before test experts is, therefore, become the need for test fairness, as opposed to test biases (Taiwo and Ojuolape, 2023), a challenge Ojerinde (2013) described as a moral burden, not only for testers but also testees, submitting that testers must consciously devise means of constructing tests whose item difficulty index for one group, remains the same for all groups of interest who have taken the same test.

Researchers (Ogidi, 2020, Omoroguiwa and Iro-Aghedo, 2016, Taiwo, 2021) have worked on investigating DIF and have found out items on the test in no way functioned differently for either gender but did on the basis of school location and type. Omoroguiwa and Iro-Aghedo (2016) investigated DIF by gender in the 2015 National Business and Technical Examinations Board (NABTEB) Mathematics Multiple-Choice Items and indicated male and female respondents functioned differently only in 34% of the total item composition, and that the remaining 66% items were bias-free. Lyson-Thomas et al. (2014) whose examination of DIF across gender in four districts that participated in a large-scale assessment in Canada, China, Finland, and Turkey, submitted that only 12% of items had visible DIF effect, while 88% were free. In line with this, Ogbebor and Onuka (2013) deployed logistics regression statistics in the identification of items, possessing DIF across school types and locations and realised only a few items were biased.

Despite the need for the determination of the bias or otherwise of items on all types of tests, captured in psychometrics as differential item functioning, available evidence revealed that the focus has largely been on achievement tests, with little attention paid to other forms of psychological testing, especially intelligence tests (Abdalaziz, 2010). This necessitates a study on determining the differential item functioning of a cognitive intelligence test.

Intelligence tests, particularly those raised to evaluate a testee's capacity for critical thinking and efficient reasoning, have been established to play highly significant roles, in both personal and societal development, a position which further intensified the necessity to critically appraise items on such tests, to determine whether or not they remain valid for the purposes for which they were developed, initially (Taiwo and Adeoye, 2023).

At the core of the ideation and construction of the Cognitive Intelligence Test for Nigerian Adolescents, was the desire to deconstruct language proficiency, deploy everyday objects in the design of items (Taiwo and Adeoye, 2023), as well as travel the nonverbal route to capture a wide range of testees, in the bid to promote inclusive education (Taiwo and Adeoye, 2023). On the basis of these, it, therefore, became critical to explore the differential item functioning parameters of CITNA, with particular attention paid to the age and gender of testees.

Study Objectives

This study was set to:

1. Determine which and how many of the items on CITNA functioned differently on the basis of testees' age.
2. Determine which and how many of the items on CITNA functioned differently on the basis of testees' gender.

Research Questions

1. Which and how many of the items functioned differently on the basis of testees' age?
2. Which and how many of the items functioned differently on the basis of testees' gender?

METHODS

The descriptive research design of the survey type was the framework of choice. The resaerchers preferred it as it is a form of quantitative research focused on the nature of the phenomenon without interest in its causative factors. In simple terms, surveys are concerned with "what, how, when and where", instead of "why", through the recording, analysis and interpretation of existing conditions (Obilor, 2018). For the purposes of this research, the researchers collected data from the responses of testees to items on CITNA, which was then analyzed to provide answers to the research questions raised.

The population of the study comprised all adolescents from Nigeria's six geopolitical zones which covers the Southwest comprising Ekiti, Lagos, Ogun, Ondo, Osun and Oyo, Southsouth covering Akwa-Ibom, Bayelsa, Cross River, Delta, Edo and Rivers, and the Southeast comprising Abia, Anambra, Ebonyi, Enugu and Imo. The Northcentral includes Benue, Kogi, Kwara, Nasarawa, Niger and Plateau, and Abuja the nation's capital, while the Northeast includes Adamawa, Bauchi, Borno, Gombe, Taraba and Yobe, and the Northwest covers Jigawa, Kaduna, Kano, Katsina, Kebbi, Sokoto and Zamfara. The need to enumerate all zones in Nigeria became expedient owing to the peculiarities of each zone, especially in terms of culture, geography, education, and other psychosocial factors which tended to shape the distinctive characterisation

of that zone's adolescents. Adolescents attending schools (as at the time of the study), easily reachable, and possessing the familiarisation with test writing, were the research participants.

Multistage sampling procedure was adopted for the determination of a sample size of 3900 secondary school students drawn from Nigeria's six geopolitical zones – Southwest, Southeast, Southsouth, Northwest, Northeast and Northcentral. In the first stage, all six geopolitical zones were enumerated. In the second stage, the simple random sampling technique was used to select a state from each zone (Lagos – Southwest, Enugu – Southeast, Edo – Southsouth, Kogi – Northcentral, Kano – Northwest and Adamawa – Northeast), totalling six states. In the third stage, the simple random sampling technique was used to select two local government areas from each selected state (Mushin and Agege – Lagos, Enugu East, and Igbo Etiti – Enugu, Akoko-Edo and Ovia Southwest – Edo, Adavi, and Lokoja – Kogi, Bichi, and Dala – Kano, as well as Fufore and Girei – Adamawa), totalling 12 local government areas. In the fourth stage, convenience sampling was used to select 30 schools from the selected 12 local government areas. The fifth stage was the use of the simple random sampling technique to select 130 students per school, totalling 3900 students across Nigeria.

The instrument for data collection, CITNA (Taiwo and Adeoye, 2023), a standardised 50-item, 5-section instrument had its reliability coefficient using KR-20 reliability as 0.91, while the five dimensions' reliabilities were Attention to detail (0.90), Logic (0.86), Spatial management (0.83), Quantitative reasoning (0.84) and Associative memory (0.89), which all depicted sound and acceptable levels of reliability. This was analysed using jMetrik version 4.2.1.

RESULTS

RQ1: Which and how many of the items functioned differently on the basis of testees' age?

The determination of items which functioned differently or otherwise, was undertaken using the DIF option of jMetrik. Particular attention was paid to the differences in the difficulty parameters (if any), of the items in terms of the focal group (younger students) or the reference group (older students), which Thompson (2022) had posited to be statistically represented as the reference group being the majority, and the focal group being the minority. In the case of CITNA, the reference group was represented by older students who constituted 51.9%, and focal group by younger students who constituted 48.1%.

Table 1: IRT analysis of DIF with respect to age

Item	DIF	P-value	Decision	Class	Description
A1	NaN	NaN	NaN	B-	Removed
A2	0.01	0.93	No DIF	A	Normal
A3	0.03	0.86	No DIF	A	Normal
A4	0.00	0.98	No DIF	A	Normal
A5	0.28	0.59	No DIF	A	Normal
A6	0.07	0.79	No DIF	A	Normal
A7	0.06	0.80	No DIF	A	Normal
A8	0.02	0.88	No DIF	A	Normal
A9	0.54	0.46	DIF	C-	Favoured older students
A10	NaN	NaN	NaN	B-	Removed

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS |Volume VII Issue VII July 2023

B1	0.00	0.99	No DIF	A	Normal
B2	0.00	0.97	No DIF	A	Normal
B3	0.09	0.77	No DIF	A	Normal
B4	0.68	0.41	DIF	C-	Favoured older students
B5	0.02	0.88	No DIF	A	Normal
B6	0.02	0.89	No DIF	A	Normal
B7	0.07	0.76	No DIF	A	Normal
B8	NaN	NaN	NaN	B-	Removed
B9	0.09	0.77	No DIF	A	Normal
B10	0.02	0.88	No DIF	A	Normal
C1	0.70	0.40	DIF	C-	Favoured older students
C2	0.02	0.89	No DIF	A	Normal
C3	0.08	0.78	No DIF	A	Normal
C4	0.02	0.88	No DIF	A	Normal
C5	0.36	0.55	No DIF	A	Normal
C6	0.02	0.90	No DIF	A	Normal
C7	0.04	0.85	No DIF	A	Normal
C8	0.01	0.93	No DIF	A	Normal
C9	0.04	0.84	No DIF	A	Normal
C10	0.15	0.70	No DIF	A	Normal
D1	0.16	0.69	No DIF	A	Normal
D2	0.34	0.47	No DIF	A	Normal
D3	0.00	0.98	No DIF	A	Normal
D4	NaN	NaN	NaN	B-	Removed
D5	0.00	0.95	No DIF	A	Normal
D6	0.14	0.71	No DIF	A	Normal
D7	0.02	0.89	No DIF	A	Normal
D8	0.49	0.48	DIF	C-	Favoured older students
D9	0.26	0.61	No DIF	A	Normal
D10	0.19	0.66	No DIF	A	Normal
E1	0.01	0.92	No DIF	A	Normal
E2	0.00	0.99	No DIF	A	Normal
E3	0.00	0.95	No DIF	A	Normal
E4	0.06	0.95	No DIF	A	Normal
E5	0.26	0.61	No DIF	A	Normal
E6	0.00	0.95	No DIF	A	Normal
E7	0.10	0.75	No DIF	A	Normal
E8	0.01	0.92	No DIF	A	Normal
E9	0.65	0.42	DIF	B+	Favoured younger students
E10	0.05	0.83	No DIF	A	Normal

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS |Volume VII Issue VII July 2023

In Table 1 above, IRT DIF statistics on students' item performance and age are depicted. While Column 2 shows the IRT DIF statistics for students, Column 3 reveals the p-value of the IRT DIF with respect to the varying difficulty parameter estimates when age is considered, Column 4 gave a decision based on the p-value, as Column 5 described the nature of the DIF, revealing whether it favoured the focal group (younger students) or the reference group (older students).

In analyzing the table, Class "A" implied no DIF, as B+ and C+ implied DIF favouring the focal group (younger students). B- showed removed items and C- implied DIF favouring the reference group (older students). Though 4 items (i.e. items A1, A10, B8, and D4) were removed due to the display of non-purified matching scores (NaN), Cognitive Intelligence Test for Nigerian Adolescents (CITNA) could be said to be free of DIF when age is considered, except for 5 items (items A9, B4, C1, D8 and E9). While 4 items (A9, B4, C1 and D8) favoured older students, item E9 favoured younger students. This implies that the cognitive intelligence test developed could be used to differentiate the intelligence of students based on age.

Cumulatively, while 9 items were dropped for the reasons of displaying non-purified matching scores (NaN) and for functioning differently (DIF), 41 items on Cognitive Intelligence Test for Nigerian Adolescents (CITNA) were confirmed pure with regards to exhibiting correctly, capacity to measure cognitive intelligence irrespective of testees' age. This implies that the cognitive intelligence test has the capacity to measure differential item functioning.

RQ2: Which and how many of the items functioned differently on the basis of testees' gender?

To determine those items which functioned differently between male and female students using the IRT Framework, the DIF option of jMetrik was computed, with the analysis showing the differences in the difficulty parameters (if any), of the items in terms of the focal group (male students) or the reference group (female students), following the postulation of Thompson (2022), that the majority is statistically represented by the reference group (female students who constituted 51.5% of the sample size in the case of CITNA), and the minority being the focal group (male students who constituted 48.5% in the case of CITNA).

Table 2: IRT analysis of DIF with respect to gender

Item	DIF	P-value	Decision	Class	Description
A1	NaN	NaN	NaN	B-	Removed
A2	0.04	0.83	No DIF	A	Normal
A3	0.08	0.78	No DIF	A	Normal
A4	0.05	0.92	No DIF	A	Normal
A5	0.38	0.64	No DIF	A	Normal
A6	0.08	0.85	No DIF	A	Normal
A7	0.06	0.84	No DIF	A	Normal
A8	0.12	0.87	No DIF	A	Normal
A9	0.75	0.24	DIF	C-	Favoured female students
A10	NaN	NaN	NaN	B-	Removed
B1	0.06	0.96	No DIF	A	Normal
B2	0.23	0.86	No DIF	A	Normal

B3	0.14	0.79	No DIF	A	Normal
B4	0.78	0.23	DIF	C-	Favoured female students
B5	0.26	0.84	No DIF	A	Normal
B6	0.16	0.92	No DIF	A	Normal
B7	0.09	0.76	No DIF	A	Normal
B8	NaN	NaN	NaN	B-	Removed
B9	0.13	0.87	No DIF	A	Normal
B10	0.04	0.87	No DIF	A	Normal
C1	0.28	0.78	No DIF	A	Normal
C2	0.84	0.17	DIF	B+	Favoured male students
C3	0.08	0.76	No DIF	A	Normal
C4	0.03	0.87	No DIF	A	Normal
C5	0.37	0.58	No DIF	A	Normal
C6	0.07	0.91	No DIF	A	Normal
C7	0.06	0.85	No DIF	A	Normal
C8	0.97	0.38	DIF	C-	Favoured female students
C9	0.08	0.87	No DIF	A	Normal
C10	0.24	0.78	No DIF	A	Normal
D1	0.17	0.79	No DIF	A	Normal
D2	0.36	0.48	No DIF	A	Normal
D3	0.02	0.94	No DIF	A	Normal
D4	NaN	NaN	NaN	B-	Removed
D5	0.03	0.93	No DIF	A	Normal
D6	0.13	0.78	No DIF	A	Normal
D7	0.03	0.84	No DIF	A	Normal
D8	0.27	0.75	No DIF	A	Normal
D9	0.53	0.95	No DIF	A	Normal
D10	0.19	0.66	No DIF	A	Normal
E1	0.06	0.97	No DIF	A	Normal
E2	0.08	0.89	No DIF	A	Normal
E3	0.04	0.92	No DIF	A	Normal
E4	0.08	0.96	No DIF	A	Normal
E5	0.24	0.68	No DIF	A	Normal
E6	0.07	0.85	No DIF	A	Normal
E7	0.85	0.12	DIF	C-	Favoured female students
E8	0.03	0.94	No DIF	A	Normal
E9	0.83	0.32	DIF	B+	Favoured male students
E10	0.68	0.87	No DIF	A	Normal

Table 2 above showed the IRT DIF statistics on students' item performance and gender. From the table, Column 2 could be said to provide the IRT DIF statistics for students, Column 3 the p-value of the IRT DIF with respect to the varying difficulty parameter estimates considering gender, Column 4 the decision based on the p-value, and Column 5 describing the nature of the DIF, as to whether or not it favoured the focal group (male students) or the reference group (female students). Forty items showed that there was no differential item functioning based on the gender of the respondents. This implies that cognitive intelligence could differentiate performance between male and female students.

Class "A" implied no DIF, B+, and C+ implied DIF Favouring the focal group (male students), B- showed removed items, and C- implied DIF Favouring the reference group (female students). Although 4 items (i.e. items A1, A10, B8, and D4) were removed for displaying non-purified matching scores (NaN), Cognitive Intelligence Test for Nigerian Adolescents (CITNA) could be said to be DIF-free with respect to gender, save for 6 items (i.e. items A9, B4, C2, C8, E7 and E9). While 4 items (A9, B4, C8, and E7) favoured female students, 2 items (C2 and E9) favoured male students.

In all, 10 items were dropped for reasons of displaying non-purified matching scores (NaN) and for functioning differently (DIF), as 40 items on the Cognitive Intelligence Test for Nigerian Adolescents (CITNA) were confirmed pure with regards to exhibiting correctly, capacity to measure intelligence, irrespective of testees' gender. This ascertains that the cognitive intelligence test developed is a good measure to differentiate the gender and age of adolescents.

DISCUSSION

In the case of CITNA, a comparison is made between DIF values and the p-value of items on the test. For DIF to exist, its value must be greater than the p-value, indicating that the particular item is sensitive to that particular intervening variable. When not, then there is no DIF, indicating that items do not function differently according to the variable in question.

For CITNA, the comparison between ages was split into the focal group (younger students) and the reference group (older students). Of the 50 items deployed on 3002 Nigerian secondary school students, while 4 items were deleted for showing non-purified matching scores, 5 exhibited DIF. In summary, therefore, 41 items were found to be age-neutral, having exhibited no DIF in terms of age. This determination of an instrument's differential item functioning with respect to age, in Nigeria, is in tune with earlier efforts by Kpolovie and Emekene (2016), in their study on the item response theory validation of Advanced Progressive Matrices in Nigeria, as well as that of Omoroguiwa and Iro-Aghedo (2016) that items on the 2015 NABTEB Examinations functioned differently, but disagrees with Ogidi (2020), whose investigation of the DIF on the basis of gender for a non-verbal intelligence test using item response theory in Rivers State reported items did not function differently.

For CITNA, a comparison is made between DIF values and the p-value of items on the test. For DIF to exist, its value must be greater than the p-value, indicating that the particular item is sensitive to that particular intervening variable. When not, then there is no DIF, indicating that items do not function differently according to the variable in question.

Of the 50 items deployed on 3002 Nigerian secondary school students, while 4 items were deleted for showing non-purified matching scores, 6 exhibited DIF, with 4 favouring female students and 2 favouring male students. Cumulatively, therefore, while 10 items were deleted for reasons highlighted above, 40 items were found to be gender-neutral, having exhibited no DIF in terms of gender.

This result aligns with those of Alade, Aletan and Sokenu (2020), whose DIF Analysis of the 2018 West African Senior School Certificate Examination (WASSCE) Mathematics Achievement Tests in Lagos State,

Nigeria revealed some form of bias according to gender, and those of Iweka (2018) on the use of differential item functioning analysis for bias analysis in test construction, with respect to gender.

CONCLUSION

This study successfully explored the differential item functioning (DIF) parameters of the Cognitive Intelligence Test for Nigerian Adolescents (CITNA) on the basis of age and gender, finding out that while 9 items, of the 50 on the test, were biased against testees on the basis of age, 10 displayed bias on the basis of gender. This study also established the capacity of DIF in the exploration of test item bias in intelligence tests. The study has an implication for counselling practices, especially when dealing with the gender and age of the clients. The study also has an implication for diagnostic issues in the areas of comparison of groups of individuals.

RECOMMENDATIONS

Based on this study, the following were recommended:

1. Teachers in secondary schools should on a regular basis conduct a performance test that could identify a lower and upper achiever.
2. Having identified students with low academic performance using Differential Item Functioning, this would be a good measure for inclusive education for all categories of students. So, there should be a policy that will take care of students with learning impairments.
3. Differential Item Functioning is a veritable tool for the elimination of test item bias. In that regard, test developers, and Psychometricians need knowledge of this tool to categorize people.
4. Test developers, test experts, and psychometricians should incorporate DIF into the constitution of test items right from the inception of cognitive testing of the respondents.
5. Similarly, testers should consider the background of testees right from the conception of test items so as to provide immunity against bias.

REFERENCES

1. Abdelaziz, N. 2010. A gender-related differential item functioning of Mathematics test items. The International Journal of Educational and Psychological Assessment 5:101-116.
2. Alade, O. M., Aletan, S. and Sokenu, B. S. 2020. Assessing the differential item functioning of 2018 WASSCE mathematics achievement tests in Lagos State, Nigeria. AJB-SDR 2.2:8- 20.
3. Chikwe, C. K. 2017. Fundamentals of test, measurement, and evaluation in education. Port Harcourt: Emmanuel ventures.
4. Herrnstein, R. J. & Murray, C. 1996. The Bell Curve: Intelligence and class structure in American life. Free Press.
5. Inko-Tariah, D. C. and Ogidi, R. C. 2017. Fundamentals of psychological testing for psychologists, counselors, and educationists. Port Harcourt: Rodi printing and publishing company.
6. Iweka, F. 2018. Use of differential item functioning (DIF) analysis for bias analysis in test construction. International Journal of Education, Learning, and Development 6.3: 80-91.
7. Kpolovie, P. J. 2015. The indispensability of intelligence testing in the repositioning and revitalization of Nigerian education. Global Academic Journal 9:1-13.
8. Kpolovie, P. J. and Emekene, C. O. 2016. Psychometric Advent of Advanced Progressive Matrices – Smart Version (APM-SV) for use in Nigeria. European Journal of Statistics and Probability 4.3:20-60.
9. Lyson-Thomas, J., Sandilands, D. and Ercikan, K. 2014. Gender differential item functioning in Mathematics in four international jurisdictions. Education and Science 39.172:20-32.
10. MacCann, C. and Roberts, R. D. 2008. New paradigms for assessing emotional intelligence: Theory and data. Emotion 8.4:540-551.

11. Mingat, A., Tan, J. P. and Sosale, S. 2003. Tools for education policy analysis. Washington DC: The WorldBook.
12. Obilor, E. I. 2018. Fundamentals of research methods and statistics in education and social sciences. Port Harcourt: SABCOS printers and publishers.
13. 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.
14. Ogbebor, U. and Onuka, A. 2013. Differential item functioning method as an item bias indicator. Educational Research 4.4:367-373.
15. Ogidi, R. C. (2020). Determining the differential item functioning of non-verbal intelligence test using item response theory in Rivers State. AJB-SDR 2.1:83-93.
16. Ojerinde, D. 2013. Classical test theory (CTT) vs item response theory (IRT) An evaluation of the comparability of item analysis results. Retrieved from www.ui.edu.ng/sites/profojerinde'slectures.pdf on 20 November 2022.
17. Omorogiuwa, K. O. and Iro-Aghedo, E. P. 2016. Determination of differential item functioning by gender in the National Business and Technical Examinations Board (NABTEB) 2015 Mathematics Multiple Choice Examination. International Journal of Education, Learning and Development 4.10:35-45.
18. Serpell, R. and Haynes, B. P. 2004. The cultural practice of intelligence testing: Problems of international export. In R. J. Stenberg and E. L. Grigorenko (Eds). Culture and Competence: Contexts of Life Success. American Psychological Association:163-185.
19. Taiwo, A. K. and Ojuolape, M. O. 2023. Comparison of African Indigenous and Western Intelligence Tests using Validation Processes of Bakare Progressive Matrix and Wechsler Adult Intelligence Tests. International Journal of Research and Innovation in Social Science. Vol. VII Issue VI, Pg 739 – 750. ISSN No. 2454-6168. Doi: 10.47772/IJRISS.
20. Taiwo, A. K. and Adeoye, O. A. 2023. Development of an Indigenous Cognitive Intelligence Test for Nigerian Adolescents. An unpublished thesis in the Department of Counselling and Human Development Studies, University of Ibadan.
21. Thompson, N. 2022. Differential item functioning (DIF). Retrieved fromwww.assess.com/ differential item functioning on 26 November 2022.