

Using Item Characteristics Curve Approach in Detecting Item Bias in a Chemistry Achievement Test in Nigeria

Author's Details: Orluwene, Goodness W¹ Ph.D Queensoap, Memory^{2*} Ph.D

⁽¹⁾Department of Educational Psychology, Guidance & Counselling, Faculty of Education, University of Port Harcourt, Choba, Port Harcourt. Rivers State. ⁽²⁾Department of Arts Education, Faculty of Education, Federal University Otuoke, Yenagoa, Bayelsa State

Abstract

this study used Item Response Theory-Item characteristic Curve approach to detect item bias in a 60-item Chemistry Achievement Test (CAT). The CAT was administered to 400 students of senior secondary III in Niger State, Kwara State, Anambra State and Bayelsa State representing Hausa, Yoruba, Igbo and Ijaw ethnic groups. A descriptive-comparative research design was used to describe and compare examinees of the four ethnic groups. The two item parameters-item difficulty and item discrimination were estimated for each group with Proxy Algorithm and Point Biserial correlation methods while Microsoft Excel and Prism Graph pad 5.1 version, were used to compute the probability of correct answer, theta (θ) and construct the item characteristic curves respectively. The ICC curves when compared between the focal group and reference group flagged 56 items as showing differential item functioning (DIF) which were tagged biased, meaning that the ICC of Ijaw (focal group) was not identical with the ICCs of the other ethnic groups (reference group). The study was able to established, that the 60-item CAT has some items that showed biases and not all items that exhibited differential functioning were flagged biased. Therefore, the study concludes that there was an ethnic bias in the CAT. The study recommends, among others, that Item Response Theory-ICC approach should be used in educational testing in Nigeria.

Keywords: item characteristic curve (ICC), Chemistry Achievement Test (CAT), Item Difficulty and item discrimination.

1. Introduction

Do test items function in different ways for different groups of test-takers? Item functioning is intended to be invariant with respect to irrelevant aspects of test-takers, such as gender, ethnicity and socio-economic status. However, item functioning is expected to be altered by interventions targeted at those items, for instance, the use of calculators in arithmetic or statistic tests or the use of assistive devices on mobility items.

All things being equal, when interventions are put into control and test items shows deferring probability of success from test takers, such items can be said to occur differentially. Brown (2012) presented that differential item functioning (DIF) occurs when examinees from different groups show differing probabilities of success on the item after matching on the construct that the item is intended to measure. For example, Scheuneman in Brown posited that when students are asked to compare the weights of several objects, including football. It will be crystally obvious that girls are less likely to have handled a football, they will found the item more difficult than boys, even though they have mastered the concept measured by the item. Also, if students are asked to find a synonym to a word, “ebony”, and the students are grouped as black students and white students, the former group will be more likely to answer the item correctly than the later group because ebony may represent a dark-coloured wood and a name of a popular magazine targeted to African-Americans. These illustrations depict differential item functioning (DIF).

According to Abedalaziz (2011) “DIF is a collection of statistical methods utilized to determine if examination items are appropriate and fair for testing the knowledge of different groups of examinees “(p.2). The author further asserted that the use of DIF aids in the identification of test items that are potentially biased. DIF methods, therefore, assess the test-takers response patterns to specific test items. DIF occurs when a statistically significant difference is evident in the probability that the test takers from the different groups, who have the same underlying ability on the measured construct, demonstrate differing probabilities of correctly answering the item, remarked Abedalaziz (2011). It was also noted that the DIF analysis of one specific item is as independent as possible from the DIF analysis of the other test items (Zumbo in Abedalaziz).

Test experts have identified, despite the relevance of DIF methods, that an item showing apparent DIF is not sufficient to declare such item bias. Abedalaziz explicitly stated that DIF is required but is not sufficient for item bias. He explained further that if DIF is apparent, then its presence is not sufficient to declare item bias. This assertion was based on the fact that there might be the possibility that an item which shows a DIF may be a result of the actual difference between the

groups ability to respond to the item hence, such item showing DIF cannot be statistically declared bias. Camilli and Shephard in Abedalaziz (2011) added that in order to be able to determine whether an item that shows DIF is biased or not, further analysis have to be done. This will enable the test expert to determine whether the differences depended on differences of ability of the compared groups (not biased) or on the item measuring something else than intended (biased).

Consequently two distinct forms of DIF have been identified, namely uniform and non-uniform DIF. Hambleton, Swamanithan and Rogers in Abedalaziz posited that uniform DIF occurs when two ICCs differ, but are more or less parallel. That is, uniform DIF is likely to occur when ICCs have different “b” (difficulty) parameters and similar “a” (discrimination or slope) parameters. For the non-uniform DIF, it is identifiable when there is an interaction between the two ICCs at some point of ability value. This is indicative when the two ICCs have similar “b” parameters and different “a” parameters. Similarly Brown (2012) identified that uniform DIF exist when there is a lower probability of endorsing the item at all levels of ability which apparently affects the origin of scale while non-uniform DIF is noticed when ICCs of two comparing groups show a higher probability of the item at low level of trait, but lower probability at high level or vice versa. Brown argued that non-uniform DIF affects measurement unit and origin of scale.

These item characteristics curves (ICCs) are the functions that relate the probability of a correct answer on an item to the ability measured by the test containing the item (Abedalaziz, 2011). He further reiterated that if the un-idimensional assumption of the test is met, item characteristics curve defined by its item parameters will remain unchanged across sub population groups. In other words, an ICC estimated from any group will be equal to an ICC from another, and both will be equal to the ICC estimated from responses of all test-takers.

Similarly Adedoyin (2010) identified two technical properties that can be used to describe ICC, the values of item difficulty and item discrimination.

According to Adedoyin, the value of item difficulty denoted by “b” is a location parameter, indicating the position of the item characteristic curve in relation to the ability that is required for an examinee to have a 50% chance of getting the item right. Meanwhile, the item discrimination “a” provides information on how well an item separates examinees with high s

Furthermore, Adedoyin stated that;

the flatter the ICCs, the less the item is able to discriminate since the probability of correct responses at the low ability levels is nearly the same as it is at high ability levels. The steeper the curve the better the item can discriminate. (p.387)

It is, therefore, true to declare that if the ICCs for two populations are the same, the item is not biased. If the ICCs are different, the item is biased, that is, it is functioning or behaving differently across the groups.

Penfield, Alvarez and Lee (2009) stated that the framework of differential item functioning (DIF) has become an integral component of test validation methodology and the study of test fairness. Differential item functioning (DIF) investigates the items in a test, one at a time, for signs of interactions with sample characteristic. Pine in Cohen (2006) opined that DIF is said to occur when the probability of a correct response differs for examinees in different groups conditional on ability. An item is determined to function differently when examinees of the same ability but in different groups (e.g. males & females) have different probabilities of success on the item. This definition of DIF is useful for dichotomously scored items, such as when responses are scored as either correct or incorrect.

Measures of physical activities are often scored in more than two categories. This is common with performance measures such as sit-ups, bowling, shooting basketball, running, and gymnastics. These kinds of task are often scored polytomously, that is, in more than two categories (Cohen, 2006). Consequently, Differential item functioning(DIF) needed to be defined more broadly for tasks such as these stated more generally, therefore, DIF occurs when the response function for that item or task differs in different groups (Cohen, Kim & Baker in Cohen). This later definition of DIF is more useful for performance types of items for which dichotomous scoring may not be either possible or desirable. However, Cohen revealed that DIF is not used simply to describe differences among groups. Rather, DIF is used to describe differences that are otherwise of the same ability. For example, males on average tend to be stronger than females. If we administer a running test over a 100m course, the most likely result will be that the average speed over this distance for a random sample of males will be less than that for a random sample of females. This represents a gender difference. It is not, however, an indication of DIF. More explicitly, Cohen explained that DIF does not occur simply because one group is faster, smarter or stronger than the other. In essence therefore DIF would occur if the same watch registered a faster time for male runners than for female runners even though both sets of runners actually covered 100m course in the same time.

Looney, Spray and Castelli in Cohen (2006) described DIF related to gender in a context of a basketball shooting task. DIF in a basketball shooting context arises because some shooting tasks are harder for females than males who are otherwise of the same level of basketball shooting ability. Another common realization of DIF occurs in a testing context where examinees that have the same raw scores (numbers correct) on the test, have different probabilities of getting an answer correct. Anastasi and Urbina (1997), in their own contribution related DIF as an analysis that concerned essentially with the relative difficulty of individual test items for groups with dissimilar cultural or exponential backgrounds. That is to say, DIF seeks to identify items for which equally able persons from different cultural groups have different probabilities of success. They stated that "equal ability means equal with regard to the construct that the test is designed to assess, or the criterion behavior it is designed to predict".

2. Methods

This study adopted a descriptive-comparative research design. Four matched groups were compared in terms of their probability of success on each item of the Chemistry Achievement Test (CAT). The matched groups were Yoruba, Igbo, Hausa, and Ijaws ethnic groups drawn from Kwara State, Anambra State, Niger State and Bayelsa State in Nigeria respectively. The Instrument used for this study was titled Chemistry Achievement Test (CAT). It was adopted from Orluwene (2007). The two parameter latent trait model was used in preparing the CAT. The instrument was made up of two Sections A and B. Section A was designed to elicit personal information from the respondents such as age, gender, school type etc. Section B consisted of 60 items with 5 options (multiple choices) answers to the questions (Orluwene, 2007). The instrument was dichotomously scored, 1 or 0 and was marked over 60. The instrument was revalidated through content and face validity. It was observed that the instrument was 82.62% reliable and 17.38% unreliable which was a good indication of consistency and high precision of measurement (Orluwene, 2007). Nevertheless, a pilot survey was done on a sample size of 20, made up of all ethnic groups at the Federal Government College, Odi. After administration and retrieval of the instrument, scoring and analysis was done with Cronbach Alpha reliability determination method which gave a reliability coefficient of 0.80. Since the reliability coefficient is consistent with Orluwene (2007) the instrument was therefore considered to be reliable. Four research assistants were recruited to administer the CAT instrument to respondent-examinees in their states. Marking scheme was prepared which, was in line with test designer's answers for the CAT. The data gathering procedures involved simultaneous administration of the test to the original intact classes and checking and scoring the test. Data gathered were not made up or influenced. Microsoft Excel and PrismGraph 5.1 were used to compute probability of correct answer, theta (θ) and Item characteristic curves respectively. Point biserial correlation and Proxy algorithm were used to estimate the discrimination index and difficulty index respectively (Queensoap, 2014).

3. Results

Research Question 5: Which are the items that show DIF when ICCs for the focal group are matched with ICCs for the reference groups?

Practically, item characteristics curve is a function that relates the probability of a correct answer on an item to the ability measured by the test containing the item. An ICC estimated from any ethnic group is expected to be equal or identical to an ICC from another (Abedalaziz, 2011). This study estimated the ICCs for the four comparing groups – Ijaw, Igbo, Yoruba and Hausa. Colours, such as light blue, dark blue, yellow and pink were used to identify the matched groups – Yoruba, Ijaw, Hausa and Igbo respectively. The ICCs of the 60 items were shown in figure 1-60 and brief interpretation given accordingly.

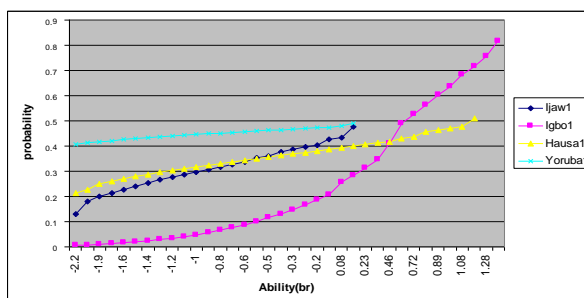


Figure 1: The one item characteristic curves for four matching groups: Ijaw ($a = 0.27, b = 0.50$) Yoruba ($a = 0.08, b = 0.60$) Hausa ($a = 0.20, b = 1.13$) and Igbo ($a = 1.07, b = 0.66$).

For item 1, the four ICCs were seen not identical. This was explained by the fact that the ICCs are significantly different from one another across the groups, then item one is said to show DIF. On visual inspection, the ICC of Hausa examinees was closer with the Ijaw ICC than the other ICCs, especially, the ICC of Igbo. In fact the ICCs of Ijaw and Yoruba meet at their high ability estimates.

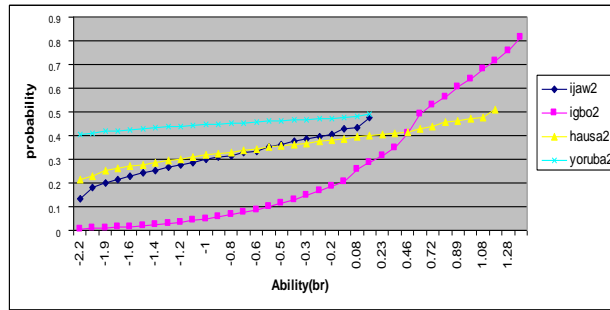


Figure 2: The two item characteristics curves of the four matching groups: Ijaw ($a=0.22, b=0.26$), Yoruba($a=-0.10, b=0.30$), Hausa($a=-0.02, b=0.21$) and Igbo($a=0.16, b=-0.52$).

This item showed a DIF similar to item one. The ICCs are not the same various for the groups. The item function similarly as in item one. It presents the Yoruba group having higher probability for correct answer at lower ability level than the other ICCs. For the Ijaw and Hausa groups their ICCs cross at one point of the ability θ . It is obvious that the item displays a substantial DIF with a very large area between Igbo and Ijaw ICCs. This item therefore, may not be an equivalent measure of the same latent variable for both groups.

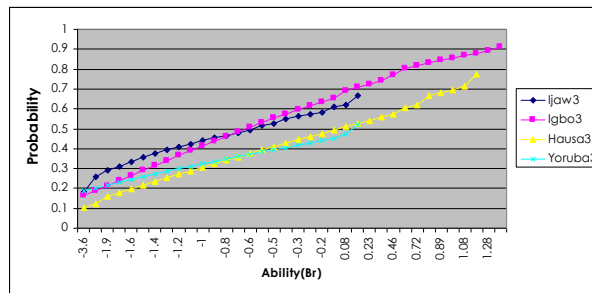


Figure 3: The three item characteristic curves of the four matching groups, Ijaw ($a=0.33, b=-0.94$), Yoruba ($a=0.37, b=0.17$), Hausa ($0.51, b=-0.19$) and Igbo ($a=0.64, b=-0.66$).

Figure 3 above showed a very small area between the ICCs of various groups. The graphs on visual inspection are not the same thereby indicating a DIF. The Ijaw ICC crosses the Igbo ICC at one point of the ability θ , indicating a non-uniform DIF. This indicates that, before the cross over point, the item favours the Ijaw group and after the ICC's cross, the item starts to favour the Igbo group.

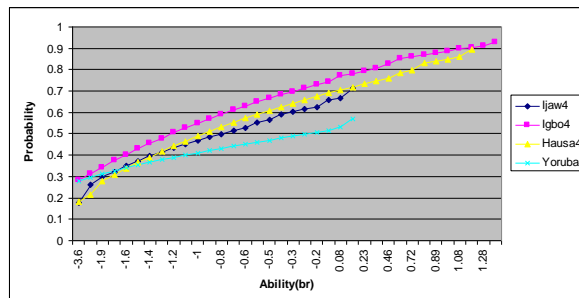


Figure 4: The four item characteristics curves of the four matching groups, Ijaw ($a=0.37, b=-1.17$), Yoruba ($a=0.29, b=-0.27$), Hausa ($a=0.55, b=-1.08$) and Igbo ($a=0.56, b=-1.19$).

Looking at figure 4, the ICC for Ijaw examinees and ICC for Hausa examinees seem much more similar than the others with Ijaw when compared. The Yoruba examinees require higher ability value to have a 50% chance of setting the item right, indicating that the item was more difficult for the Yoruba group than others. Also the figure showed that the item discriminate similarly. However, the item shows a DIF.

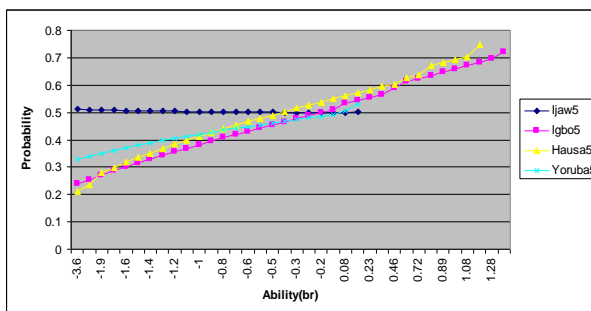


Figure 5: The five Item characteristic curves of four matching groups: Ijaw ($a=-0.01, b=-0.61$), Yoruba ($a=0.20, b=-0.06$), Hausa ($a=0.36, b=-0.55$) and Igbo ($a=0.34, b=-0.15$)

For item 5, the ICCs for Yoruba, Igbo and Hausa are similar but differ from the Ijaw ICC. The Ijaw ICC was flatter than the others, indicating that it shows less ability to discriminate among test takers. That is, the probability for correct response of low ability levels was nearly the same as it is of high ability levels. On visual inspection, the ICCs were not identical hence it shows a DIF.

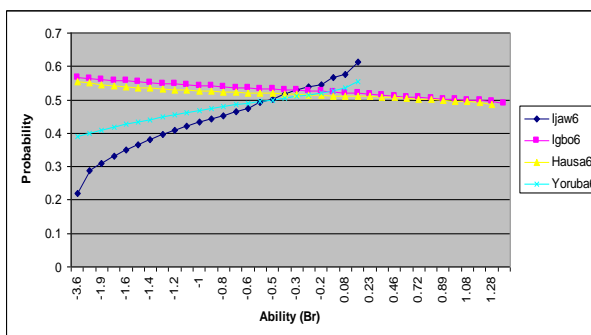


Figure 6: The six item characteristic curve of the four matching groups, Ijaw ($a=0.26, b=-0.76$), Yoruba ($a=0.16, b=-0.52$), Hausa ($a=-0.04, b=0.49$) and Igbo ($a=-0.05, b=1.04$).

The ICC curves for Ijaw, Yoruba, Hausa and Igbo are not the same, though, those of Hausa and Igbo appears similar while Yoruba and Ijaw behave similarly. The item indicated that for Yoruba and Ijaw examinees require similar ability value to have 50% chances of getting the item right. This suggests that, the item seems easier for the Ijaw and Yoruba examinees than the Hausa and Igbo. Since the graph of the ICCs revealed that they are not the same, the item can be flagged as showing DIF.

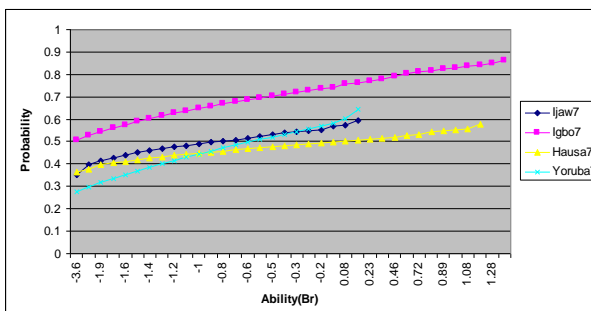


Figure7: The seven item characteristic curves for the four matching group; Ijaw ($a=0.15, b=-1.21$), Yoruba ($a=0.37, b=-0.65$), Hausa ($a=0.13, b=-0.19$) and Igbo ($a=0.29, b=-2.22$).

Figure 7 above showed that the ICCs of the various ethnic groups are not identical hence, indicating a DIF. Looking at the graphs, their flatten shape suggest similar discrimination value and different difficulty parameter. This can be termed a uniform DIF. Though, figure 7 also showed that Yoruba ICC crosses both Ijaw and Hausa ICCs at differing points on the ability estimates.

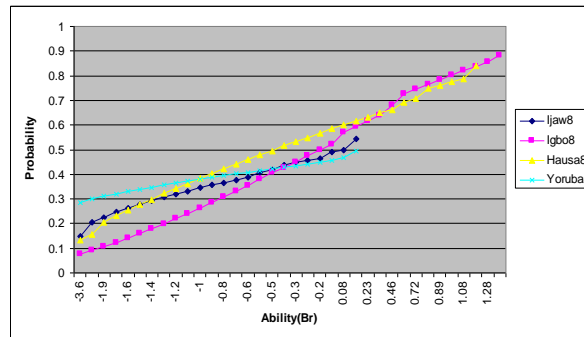


Figure 8: The eight item characteristic curve of the four matching group, Ijaw ($a=0.29, b=-0.07$), Yoruba ($a=0.21, b=0.37$), Hausa ($a=0.53, b=-0.60$) and Igbo ($a=0.37, b=-0.15$).

The item characteristic curves shown in figure 8 indicated similar ICCs for Ijaw and Yoruba while in overall assessment the graph revealed that the ICCs are not the same. This relates that the item shows a DIF, that is, the Ijaw examinees functions in the test differently than other groups. This item was seen been easier for Hausa and Igbo examinees than Ijaw and Yoruba examinees.

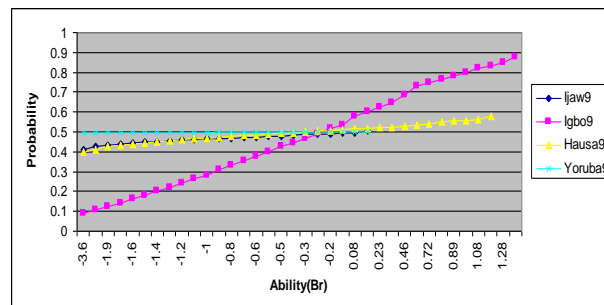


Figure 9: The nine item characteristics curves for the four matching group, Ijaw($a= 0.06, b= 0.05$) Yoruba ($a=0.01, b=-0.32$), Hausa ($a=0.11, b=-0.46$) and Igbo ($a=0.69, b=-0.20$)

Figure 9 displayed ICCs of Ijaw, Yoruba and Hausa similar but all appearing different from Igbo. One peculiar attribute of this item was that all the ICCs cross one another. The ICC for Igbo crosses all other ICCs, indicating that the test item favour the other examines at low ability values than the Igbo examines while the Igbos were more favoured at the high ability values. In any case, the item showed a DIF.

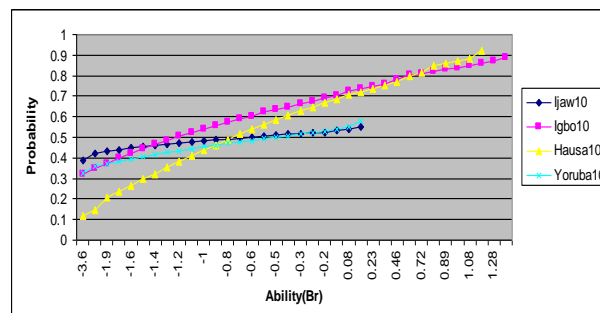


Figure 10: The ten item characteristic curves for the four matching groups, Ijaws ($a= 0.10, b=-0.94$), Yoruba ($a= 0.23, b=-0.52$), Hausa ($a=0.67, b=-0.91$) and Igbo ($a= 0.45, b=-1.19$).

On visual inspection of the graphs of the four ICCs revealed non uniform DIF where ICCs cross each other at one point to the other. Figure 10 showed that the item was easier for the Ijaw examinees than the other examinees because it requires low ability at the beginning to have higher probability for correct answers. Although, a close watch will reveal that it will take a higher ability value to have 50% chance of getting the right answer. Thus, at higher ability the item does not favour the Ijaw examinees.

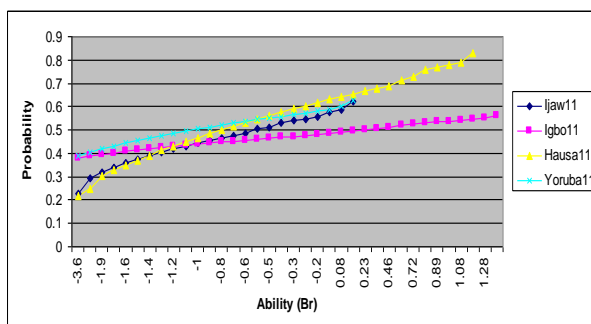


Figure 11: The eleven item characteristic curve for the four matching groups, Ijaw ($a =0.26, b=-0.85$), Yoruba ($a=0.23, b=-1.04$), Hausa ($a= 0.43, b=-0.91$) and Igbo ($a=0. 12, b=0.23$).

Based on the parameters, difficulty and discrimination index, Figure 11 showed similar ICCs. Looking at it closely, DIF was not apparent because ICCs of the reference ethnic group appear similarly and a very close area index with that of the Ijaw ICC.

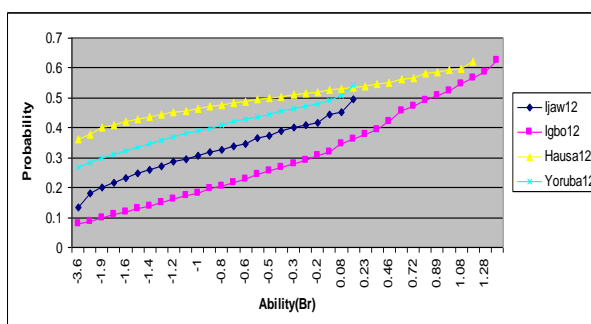


Figure 12: The twelve item characteristic curve for the four matching groups, Ijaw ($a=0.28, b=0.34$), Yoruba ($a=0.28, b=-0.06$), Hausa ($a=0.16, b=-0.60$) and Igbo ($a=0.48, b=0.85$). Figure 12 showed ICC that took the form of uniform DIF. That is, the ICCs have different “b” (difficulty) and similar “a” (discrimination slope). This was very obvious between ICC for the Yoruba and Ijaw, though their graphs seem similar or identical. Also, the ICC for Igbo looks the same with ICC for Ijaw. However, the ICC for Hausa and that of Ijaw was apparently not identical. The item, therefore, showed a DIF.

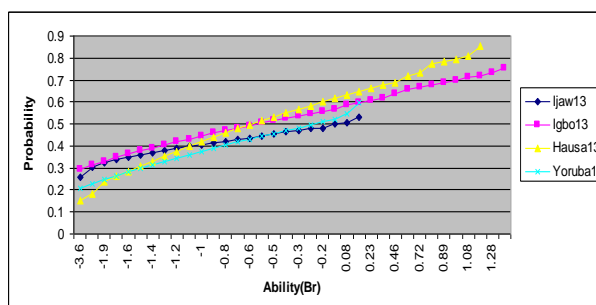


Figure 13: The thirteen item characteristic curves for the four matching groups, Ijaw ($a=0.18, b=-0.13$), Yoruba ($a=0.41, b=-0.27$), Hausa ($a=0.52, b=-0.76$) and Igbo ($a=0.32, b=-0.57$).

The ICCs in figure 13 showed that all the ethnic groups have a similar ‘‘b’’ which brought them so close and crossing over one another at various point of the ability values. Nevertheless, the ICCs for Ijaw examinees showed resemblance with that of ICC for Yoruba while differ with ICCs of Hausa and Yoruba hence the item showed a DIF.

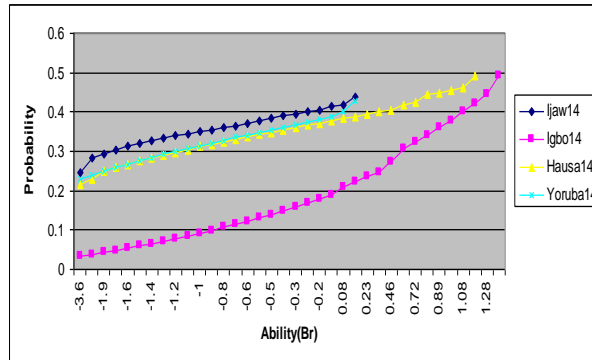


Figure 14; the fourteen item of graph 14 characteristic curves for the four matching groups, Ijaw ($a=0.13, b=1.42$), Yoruba ($a=0.22, b=1.08$), Hausa ($a=0.19, b=1.32$) and Igbo ($a=0.54, b=1.52$).

Figure 14 showed that the ICCs for Ijaw examinees have a very large area with the ICC for Igbo while very small area index with ICCs for Yoruba and Hausa. It was observed that the item seem easier for Ijaw, Yoruba and Hausa than Igbo. The Igbo examinees required a higher ability value to have 50% probability of correct response. The Ijaw, Yoruba and Hausa ICCs were identical but differ with the Igbo hence showed a DIF.

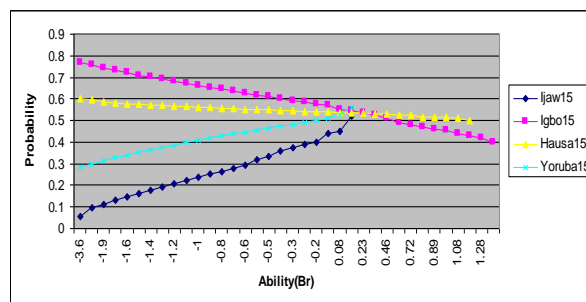


Figure 15: The fifteen item of the graph 15 characteristic curve for the matching groups, Ijaw ($a=0.44, b=0.19$), Yoruba ($a=0.28, b=-0.22$), Hausa ($a=-0.06, b=1.32$) and Igbo ($a=-0.26, b=0.55$). The four ICCs in figure 15 showed an apparent DIF among matching ethnic groups. A visual inspection of the ICCs revealed that Hausa and Igbo examinees behave similarly while Ijaw and Yoruba behave similarly. However, the ICCs were not the same. The item was observed much more difficult for Ijaw and Yoruba than the other two ethnic groups; as well discriminate better. For the Igbo and Hausa, the higher ability sub-groups have low probability for correct answers.

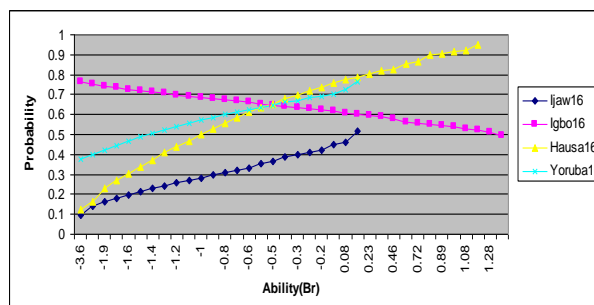


Figure 16: The sixteen item characteristic curve for the matching groups, Ijaw ($a=0.35, b=0.19$), Yoruba ($a=0.40, b=-1.42$), Hausa ($a=0.73, b=-1.13$) and Igbo ($a=-0.19, b=1.43$).

It was observed to state that the ICCs in figure 16 were not identical. The item was less difficult for Igbo examinees and as well discriminates poorly than the other ethnic groups.

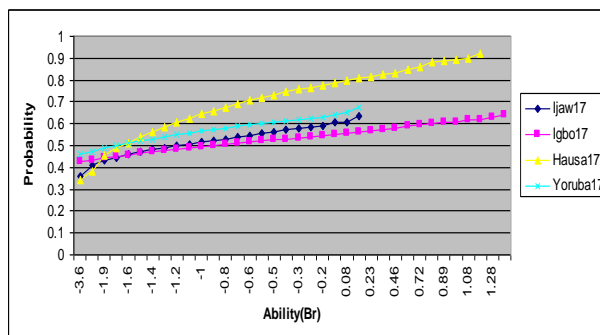


Figure 17: The seventeen item characteristic curve for the matching groups, Ijaw ($a=0.17, b=-1.60$), Yoruba ($a=0.21, b=-1.73$), Hausa ($a=0.47, b=-1.87$) and Igbo ($a=0.14, b=-0.90$).

For item 17, the ICCs for Ijaw, Igbo, and Yoruba are identical while Hausa ICC differs with them. The ICCs were indicating a non-uniform DIF.

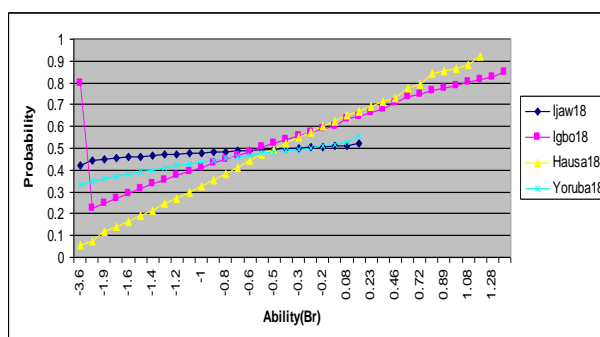


Figure 18: The eighteen Item characteristics curve for the matching groups, Ijaw ($a=0.06, b=-0.56$), Yoruba ($a=0.22, b=-0.32$), Hausa ($a=0.80, b=-0.60$) and Igbo ($a=0.50, b=-0.57$).

Looking at the ICCs in figure 18, it can be deduced that the ICC of Ijaw examinees for item 18 was not identical with the ICCs of other ethnic groups, especially that of Hausa and Igbo. More so, the ICCs display a cross-over point which made the ICCs to take the form of a non-uniform DIF, indicating a similar ‘‘b’’ parameter and different maximum shape ‘‘a’’.

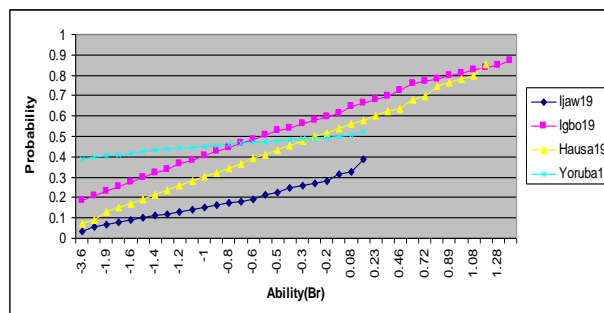


Figure 19: The nineteen item characteristics curve for the matching groups, Ijaw ($a=0.45, b=0.88$), Yoruba ($a=0.13, b=-0.12$), Hausa ($a=0.65, b=-0.37$) and Igbo ($a=0.55, b=-0.57$).

For item 19, the ICCs presented were observed similar, though not the same. Looking at them it could be noticed that Ijaw and Yoruba ICCs can be interpreted the same. However, the test item favours the Yoruba examinees more than the Ijaw examinees. The item discriminates poorly among Yoruba examinees group than the Ijaw. When considering Hausa, Igbo versus Ijaw ICC, the item discriminates similarly but seems easier for the Hausa and Igbo examinees than the Ijaw examinees. This test item showed a DIF.

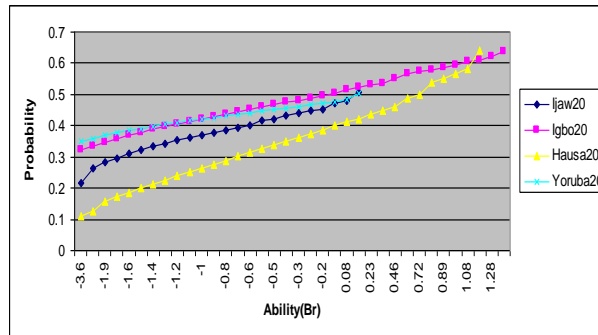


Figure 20: The twenty item characteristics curve for the matching groups, Ijaw ($a=0.20$, $b=0.19$), Yoruba ($a=0.15$, $b=0.24$), Hausa ($a=0.40$, $b=0.38$) and Igbo ($a=0.21$, $b=-0.10$).

The ICC curve for Ijaw and the other ethnic groups are not the same though, that of Igbo and Yoruba show some closeness. That is, the area between the curves is very small and the parameter would be nearly equivalent. The test item seems more difficult for the Hausa examinees than the other ethnic groups but also favors Igbo and Yoruba more than the Ijaw examinees. Thus, it showed a DIF among the groups.

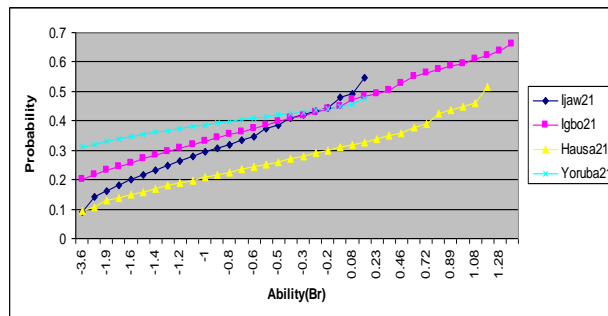


Figure 21: The twenty-one item characteristics curve for the matching groups, Ijaw ($a=0.31$, $b=-0.01$), Yoruba ($a=0.17$, $b=0.60$), Hausa ($a=0.35$, $b=1.13$) and Igbo ($a=0.33$, $b=0.28$).

This item, looking at the ICC curves, displayed a DIF because the various ICCs were not identical. Though, the area between the ICC for Ijaw and others is very small. The test item, generally, seems more difficult for the Hausa group and easier for the Yoruba group. The item was flagged showing DIF.

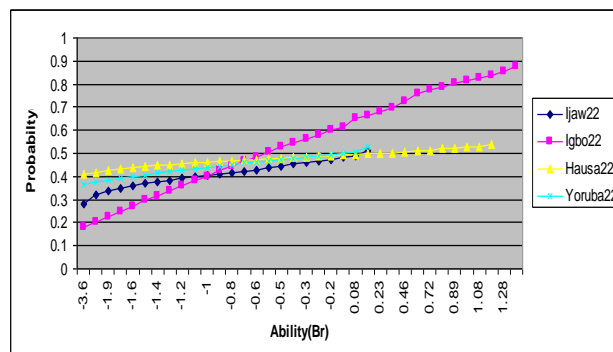


Figure 22: The twenty-two item characteristics curve for the matching groups, Ijaw ($a=0.15$, $b=0.12$), Yoruba ($a=0.16$, $b=-0.12$), Hausa ($a=0.08$, $b=0.00$) and Igbo ($a=0.56$, $b=-0.57$).

Item 22 displays item characteristics curve that have very small area between them. The ICC for Ijaw was not identical with other ICCs, especially that of Igbo. The item discriminated better among Igbo examinees than the other ethnic examinees. The item was easy for all ethnic groups but more difficult to the Ijaw because the Ijaw examinees require a higher ability value to obtain a 50% probability for a correct response. This item shows a DIF.

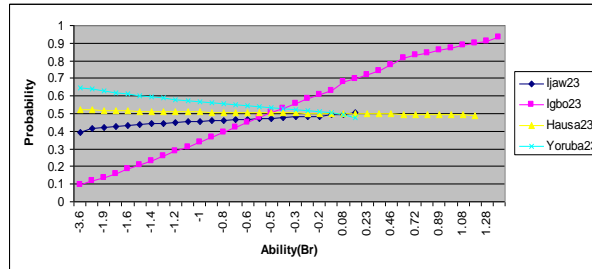


Figure 23: The twenty-three item characteristic curve for the matching groups, Ijaw ($a=0.07$, $b=0.12$), Yoruba ($a=-0.17$, $b=-0.06$), Hausa ($a=-0.02$, $b=-0.09$) and Igbo ($a=0.78$, $b=-0.48$).

The item 23 represented in figure 23, showed item characteristic are not identical; indicating a DIF among ethnic groups. Looking at the Igbo ICC, it discriminated better while the ICCs of others especially that of Hausa and Yoruba was very poor. The item seems more difficulty for the Igbo examinees because they required a higher ability value to get 0.5, proportion of those getting the item right.

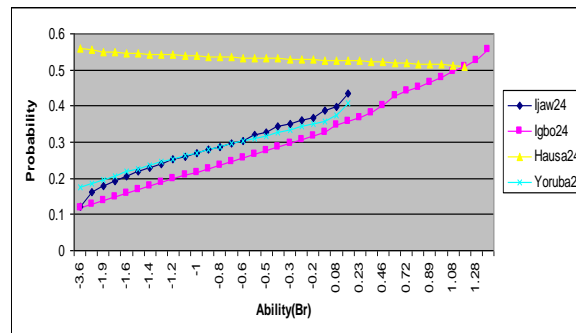


Figure 24: The twenty-four item characteristics curve for the matching groups, Ijaw ($a=0.26$, $b=0.88$), Yoruba ($a=0.28$, $b=-1.08$), Hausa ($a=-0.03$, $b=1.19$) and Igbo ($a=-0.36$, $b=1.11$).

The ICC curves for the Ijaw, Yoruba and Igbo were identical presenting a small area between the curves but the ICC of Hausa differ with them, indicating that the item was very easy for the examinees in all ability level, especially the low ability group. This item related, Ijaw ICC with Hausa ICC, as displaying a uniform DIF.

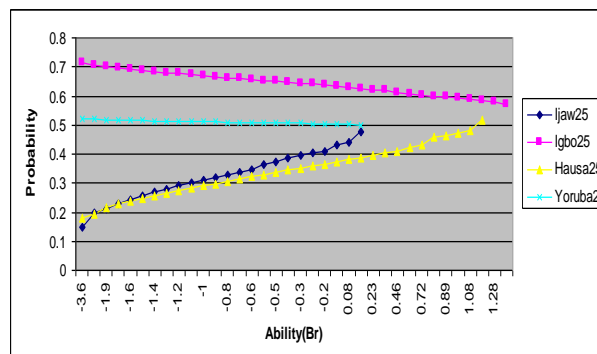


Figure 25: The twenty-five item characteristic curve for the matching groups, Ijaw ($a=0.25, b=0.50$), Yoruba ($a=0.02, b=0.30$), Hausa ($a=0.24, b=1.05$) and Igbo ($a=-0.10, b=3.17$).

Item 25 was relatively easier for the Igbo examinees than the other ethnic groups. The low ability group in the Igbo group has high probability for correct response than the low ability group in the other ethnic group. The ICC of the Igbo and the Hausa looked identical but differs with the ICCs of Yoruba and Igbo. There is an apparent DIF among the ethnic groups. For the Yoruba and Igbo examinees, the item was easier for the low ability group than the high ability, showing a poor discrimination of item while that of Ijaw and Hausa, the item discriminates similar and better than the other ethnic groups.

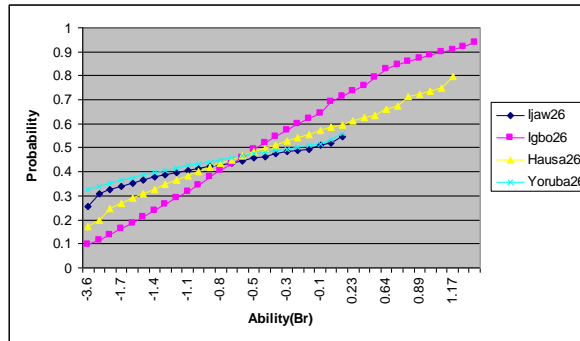


Figure 26: The twenty-six item characteristics curve for the matching groups, Ijaw ($a=0.19, b=-0.30$), Yoruba ($a=0.23, b=-0.32$), Hausa ($a=0.44, b=-0.60$) and Igbo ($a=0.80, b=-0.52$).

All item characteristic curves shown in figure 26 exhibited similar item discrimination and item difficulty. The item was an easy one, all examine groups have a negative item difficulty value while discrimination among low, moderate and high ability group was clear but more prominent in the Hausa and Igbo examinees. This has made the ICCs of Hausa and Igbo not identical with the Ijaw ICC, though they met ability less than -0.5 . ICC of Yoruba was slightly identical with ICC of Ijaw. Hence, there was a differential item functioning between Ijaw examinees, and Hausa and Igbo examinees.

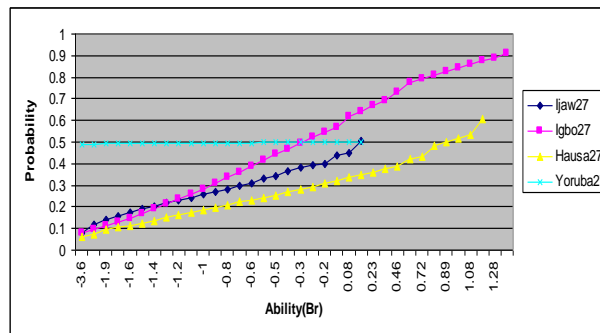


Figure 27: The twenty-seven item characteristic curve for the matching groups, Ijaw ($a=0.38, b=26$), Yoruba ($a=0.01, b=0.01$), Hausa ($a=0.48, b=0.70$) and Igbo ($a=0.78, b=-0.29$).

Figure 27 showed that, ICC for Hausa examinees meeting with the ICC for Ijaw at 0.08 ability level showing a poor discrimination among examinees. Igbo ICC crosses Ijaw ICC at -1.4 ability levels but shows good discrimination among examinees. Hausa ICC looks similar to that of Ijaw ICC however the item proves more difficult for the Hausa examinees than all other groups while Yoruba ICC indicated an item that was very easy for examinees. Thus, the item showed a differential item functioning between Ijaw examinees and other group examinees.

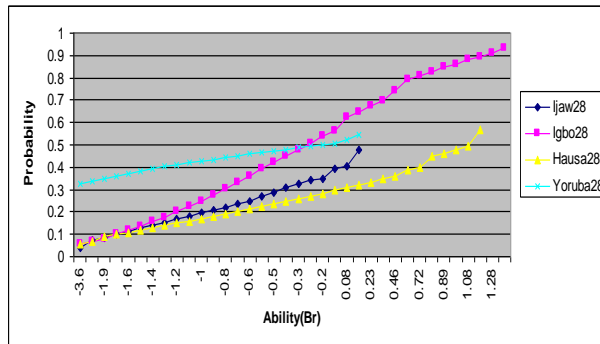


Figure 28: The twenty-eight item characteristic curve for the matching groups, Ijaw ($a=0.47$, $b=0.41$), Yoruba ($a=0.22$, $b=0.22$), Hausa ($a=0.46$, $b=0.89$) and Igbo ($a=0.88$, $b=0.25$).

The four matching group item characteristic curves in figure 28 showed a differential item functioning. This was deduced from the fact that the ICCs of other ethnic groups were not identical. A visual inspection of the graph showed that the item was easier for Hausa examinees than other group examinees. Also the discrimination was poorer among the Yoruba group than others. However, Igbo and Hausa ICCs crossed that of Ijaw ICC at low ability level below -1.4.

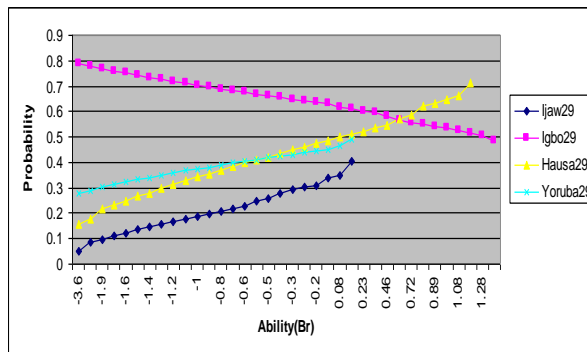


Figure 29: The twenty-nine item characteristic curve for the matching groups, Ijaw ($a=0.38$, $b=0.88$), Yoruba ($a=0.22$, $b=0.37$), Hausa ($a=0.39$, $b=-0.14$) and Igbo ($a=-0.22$, $b=1.34$).

A visual inspection of figure 29 identified item 29 as exhibiting a differential item functioning among ethnic groups. Looking at ICC for Igbo examinees, the item was easier for the low ability group than the high ability level, indicating a negative discrimination. Other ethnic examinees showed the item discriminating fairly.

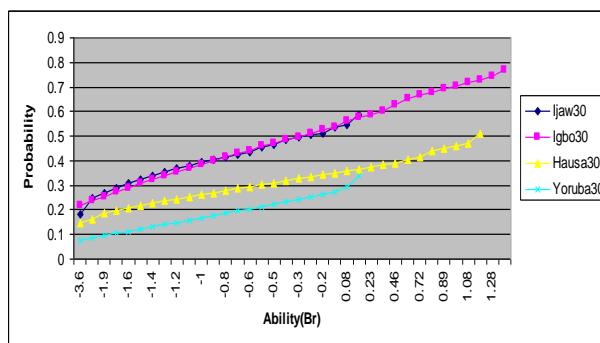


Figure 30: The thirty item characteristic curve for the matching groups, Ijaw ($a=0.28$, $b=0.46$), Yoruba ($a=0.43$, $b=1.20$), Hausa ($a=0.27$, $b=1.13$) and Igbo ($a=-0.40$, $b=-0.29$).

Figure 30 above showed an item that indicated a differential item functioning among ethnic groups. Igbo ICC and Ijaw ICC overlap each other along the ability level while there was a large area between the Ijaw and that of Yoruba and

Hausa ICCs. Looking at the display of the ICCs, there was a non-uniform DIF between Ijaw ICC and Igbo ICC while a uniform ICC between Ijaw ICC and Yoruba and Hausa ICCs.

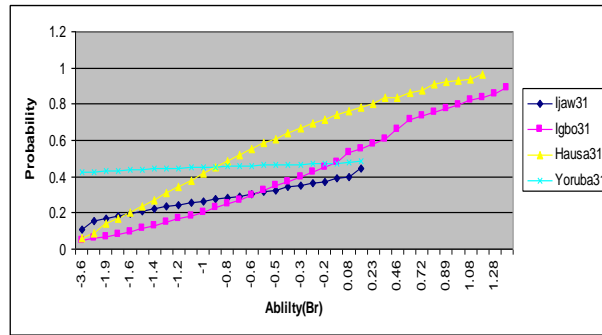


Figure 31: The thirty-one item characteristic curves for four matching groups: Ijaw ($a=0.28$, $b = 0.78$), Yoruba ($a = 0.06$, $b = 0.87$), Hausa ($a=0.90$, $b = -0.91$) and Igbo ($a=0.82$, $b=0.01$).

Looking at figure 31, the four item characteristic curve showed a differential item functioning because the ICCs were not identical. The ICC of Yoruba was flatter than all other ICCs, indicating a very poor discrimination among ability levels. ICCs of Igbo and Hausa cross ICCs of Ijaw at -0.06 and -1.6 ability levels respectively, indicating a non-uniform DIF. This indicates that the ICCs have similar item difficulty parameter and different maximum slope (item discrimination).

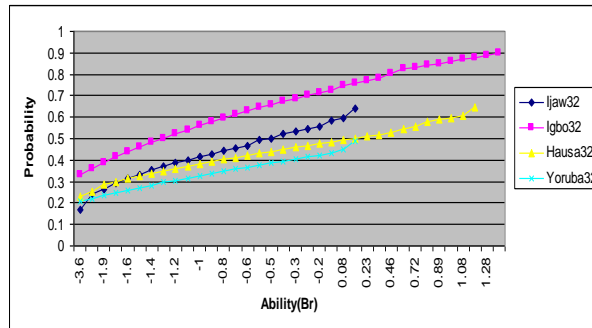


Figure 32: The thirty-two item characteristics curves for four matching groups: Ijaw ($a=0.33$, $b = -0.76$), Yoruba ($a=0.31$, $b = 0.37$), Hausa ($a=0.27$, $b = -0.09$) and Igbo ($a=0.47$, $b = -1.29$).

Item 32 showed a differential item functioning among examining groups. The ICCs shown in figure 32 were not identical especially the ICCs of Igbo and Yoruba when compared with ICC of Ijaw examinees. Hausa examinee ICC seems similar with Ijaw ICC because the area between the curves was small. The item seems easier for Igbo examinees than other groups while looks more difficult for the Yoruba examinees.

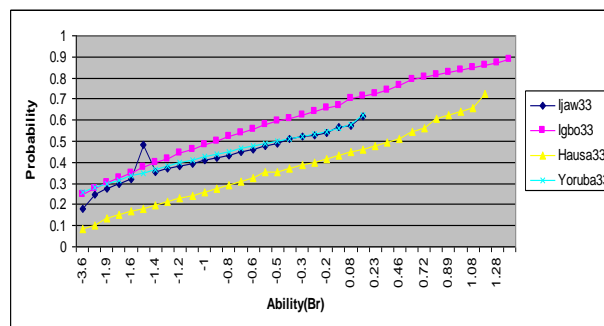


Figure 33: The thirty-three item characteristic curves for four matching groups: Ijaw ($a=0.30$, $b=-0.66$) Yoruba ($a=0.37$, $b=-0.52$), Hausa ($a=0.50$, $b=0.10$) and Igbo ($a=0.51$, $b=-0.90$).

The item characteristic curves for the four ethnic groups showed a peculiar behaviour especially, the Ijaw ICC. The Ijaw examinees required a low ability level of less than -1.4 to have a 50% probability response for correct answers, showing that the item favours the Ijaw examinees at that ability value than the other ethnic groups. ICC Yoruba and ICC Ijaw were seen overlapping each other at the high ability levels. Nevertheless, the ICCs were not identical. This suggested a differential item functioning.

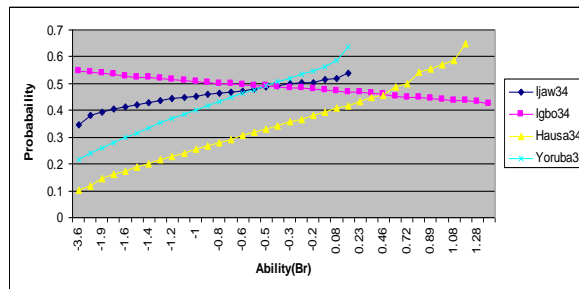


Figure 34: The thirty-four item characteristic curves for four matching groups: Ijaw ($a=0.12$, $b = -0.46$), Yoruba ($a= 0.44$, $b=-0.47$), Hausa ($a=0.42$, $b = 0.38$) and Igbo ($a=-0.08$, $b=-0.76$).

Figure 34 has shown clearly that the item characteristic curves of the comparing groups were not identical hence the item was flagged as showing differential item functioning. It was obvious that ICC of Igbo examinees presented item 34 as favouring the low ability level than the high ability level among the Igbo examinees. There was a very large area between the Hausa and Ijaw ICCs while the ICC of Yoruba crossed the ICC of Ijaw at -0.5 ability value. The item favours the Ijaw at low ability level than the Yoruba examinees. The ICCs also indicate that the item discriminate better among examinees in the Yoruba and Hausa groups.

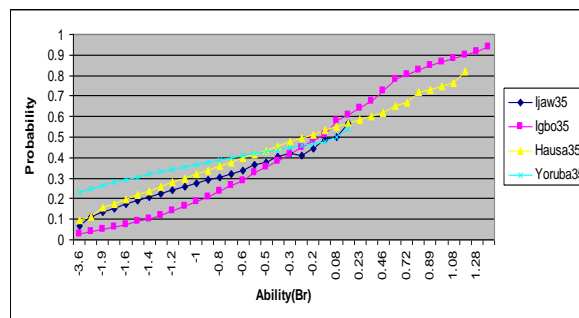


Figure 35: The thirty-five item characteristic curves for four matching groups: Ijaw ($a = 0.44$, $b = -0.07$), Yoruba ($a= 0.32$, $b= 0.01$), Hausa ($a = 0.57$, $b = -0.35$) and Igbo ($a = 1.00$, $b= -0.11$). Figure 35 showed item characteristic curves that exhibit very small area between them. The ICC of Igbo showed a better discrimination than others. For Yoruba examinees the item seems more difficult showing that the ICCs were not the same or identical hence showed some form of differential item functioning.

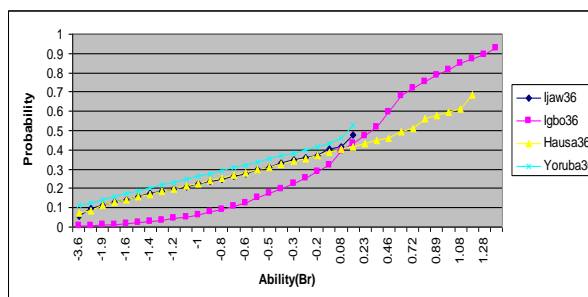


Figure 36: The thirty–six item characteristic curves for four matching groups; Ijaw (a = 0.41, b = 0.41), Yoruba (a = 0.52, b = 0.17), Hausa (a = 0.50, b = 0.32) and Igbo (a = 1.25, b = 0.28).

Item 36 showed similar item characteristic curves for Ijaw, Yoruba and Hausa but different ICC for Igbo examinees. No apparent DIF was flagged for the item because the ICCs, especially ICC for Yoruba and Hausa, were identical however the item seems easier Yoruba group than the other ethnic group.

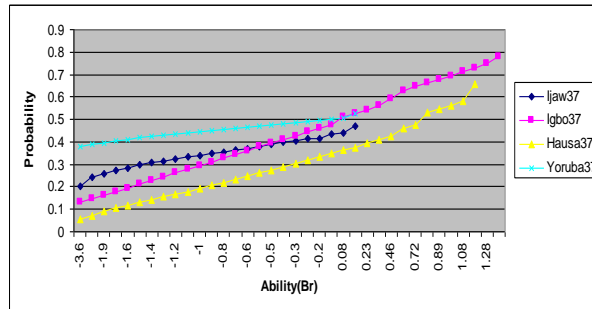


Figure 37: The thirty–seven item characteristic curves for four matching groups: Ijaw (a = 0.19, b = 0.68), Yoruba (a = 0.14, b = -0.11), Hausa (a = 0.52, b = 0.49) and Igbo (a = 0.51, b = 0.03).

Figure 37 above showed item characteristic curves that were not identical because the item discriminated similarly as better Hausa and Igbo examinees while for Yoruba and Ijaw discrimination was poor. It was observed that the item was easier for the Yoruba examinees than any other group. The ICC for Igbo crosses uniform DIF.

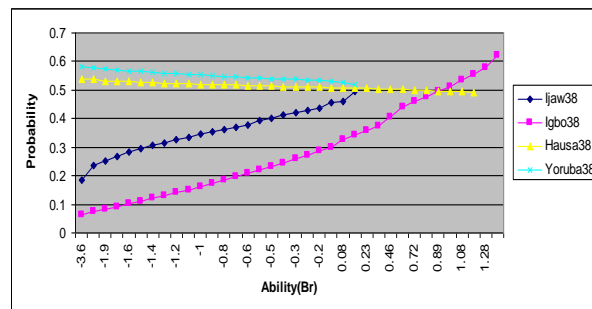


Figure 38: The thirty–eight item characteristic curves for four matching groups: Ijaw (a = 0.22, b = 0.34) Yoruba (a = -0.06, b = 1.08), Hausa (a = -0.03, b = 0.49) and Igbo (a = 0.51, b = 0.91).

The item characteristic curves shown in figure 38 was observed indicating a differential item functioning because the ICCs were not identical ICC for Hausa and Yoruba showed that the low ability group endorsed the correct answer more than the high ability level, while ICC for Ijaw and Igbo showed a fair discrimination. Nevertheless, they appear not identical.

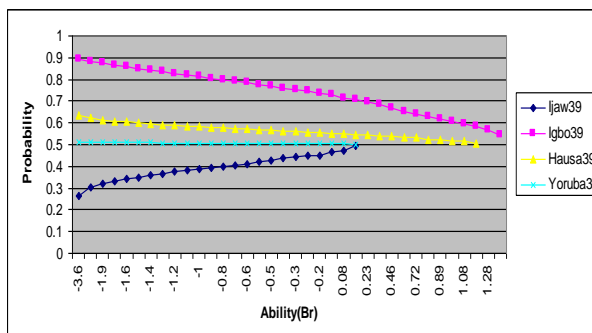


Figure 39: The thirty-nine item characteristic curves for four matching groups: Ijaw (a = 0.15, b = 0.41), Yoruba (a = -0.01, b = 0.68), Hausa (a = -0.08, b = 1.32) and Igbo (a = -0.31, b = 1.81).

The item characteristics curves in figure 39 showed that the item 39 discriminated similarly among the Igbo; Yoruba and Hausa examinees. This made their ICC similar. However, when compared to Ijaw ICC it was obvious to identify un-identical ICCs. Hence, the item showed a differential item functioning.

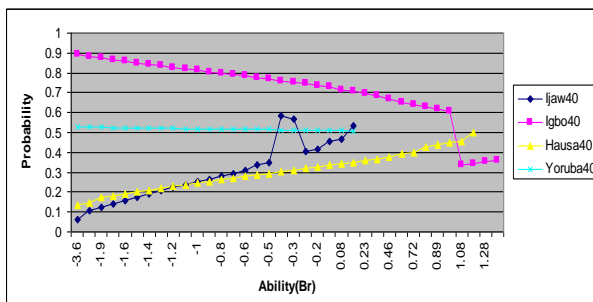


Figure 40: The forty item characteristic curves for four matching groups: Ijaw (a = 0.43, b = 0.12), Yoruba (a = -0.02, b = 1.20), Hausa (a = 0.28, b 1.22) and Igbo (a =-0.31, b = 1.81).

Figure 40 above showed an item that shows differential item functioning because all the ICCs displayed were not identical. The ICC of Yoruba was flatter than other ICCs. It was observed that ICC for Ijaw had a sharp rise at -0.5 ability level and fall sharply at -0.3 ability levels, then rise slowly in the higher ability levels. While the ICC of Igbo indicated that the low ability level endorsed the correct answer to the item more than the high ability level and experience, a sharp fall at 0.89 ability level and their rose slowly. The ICC of Hausa appear having better discrimination among ability groups though it crosses ICC for Ijaw at -1.2 ability level.

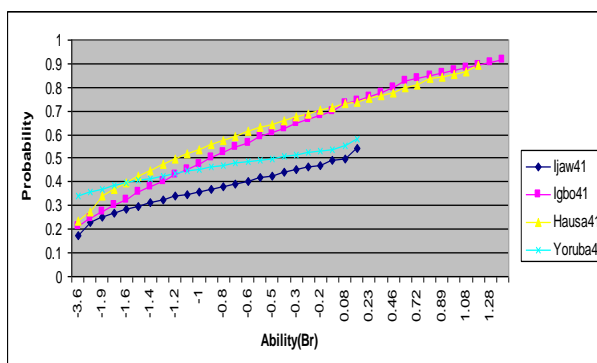


Figure 41: The forty–one item characteristic curves for four matching groups: Ijaw (a = 0.26, b = -0.07), Yoruba (a =0.23, b =-0.52), Hausa (a = 0.50, b =-1.31) and Igbo (a = 0.60, b =-0.90).

The item characteristic curves shown in figure 41 above were observed with a small area between them. The item 41 was more difficult to the Ijaw examinees than the other ethnic groups, especially that of Igbo and Hausa examinees, because the Igbo examinees require a higher ability value to endorse 50% probability for correct answer. Nevertheless, Yoruba ICC and Ijaw ICC seem identical while the ICCs of Igbo and Hausa were not identical with Ijaw ICC. Therefore, the item was flagged as a DIF item.

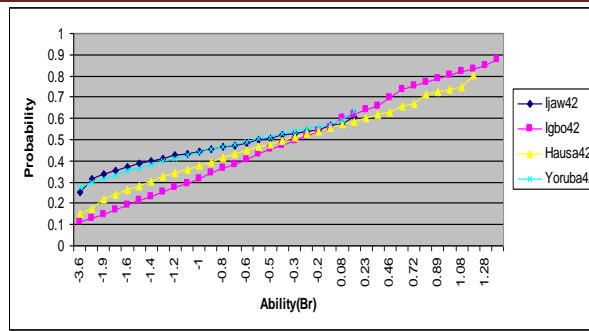


Figure 42: The forty–two item characteristic curves for four matching groups: Ijaw ($a = 0.23$, $b = -0.81$) Yoruba ($a = 0.36$, $b = -0.61$), Hausa ($a = 0.47$, $b = -0.51$) and Igbo ($a = 0.65$, $b = -0.29$).

Item 42 was endorsed similarly by all ethnic group examinees. The ICCs were showing very small area in between the focal and the reference groups. In fact, Yoruba ICC was overlapping on the Ijaw ICC while all ICCs progress on the ability continuum and met at -02 ability values, showing the ICC to be identical. Hence, there was no clear cut differential item functioning.

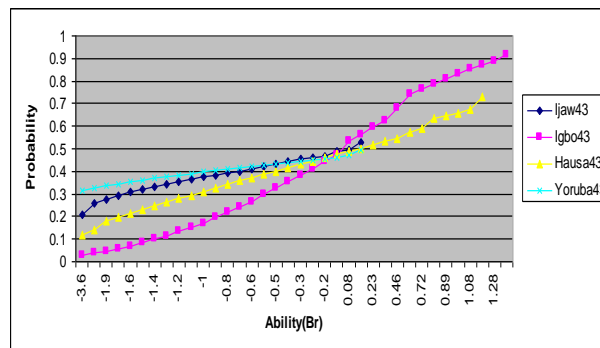


Figure 43: The forty–three item characteristic curves for four matching groups: Ijaw ($a = 0.22$, $b = 0.01$) Yoruba ($a = 0.18$, $b = 0.37$), Hausa ($a = 0.45$, $b = 0.09$) and Igbo ($a = 0.95$, $b = -0.01$).

Figure 43 has shown four item characteristic curves for the four ethnic groups. The graphs showed similar behaviour at the low ability level and because fused at -0.2 ability level. The item was more difficult for the Yoruba examinees while easier for the Hausa examinees. Looking at them critically, the ICC of the Ijaw examinees was similar to the Yoruba ICC but not identical with the Hausa and Igbo ICCs.

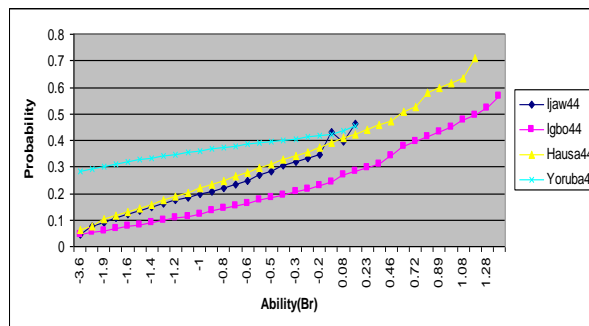


Figure 44: The forty–four item characteristic curves for four matching groups: Ijaw ($a = 0.44$, $b = 0.50$), Yoruba ($a = 0.18$, $b = 0.87$), Hausa ($a = 0.54$, $b = 0.26$) and Igbo ($a = 0.53$, $b = 1.19$).

Looking at figure 44, it was obvious that the ICC for Ijaw examinees was not identical with the ICCs of the other ethnic groups. The ICCs indicated an item that show-case a differential item functioning. Item 44 which is shown in the

figure was observed difficult for all groups but more difficult to the Igbo examinees because the item required a higher ability (1.19) for the Igbo examinees to endorse 0.50, probability for correct answer.

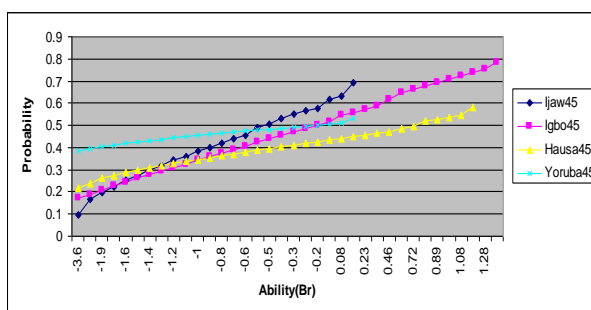


Figure 45: The forty-five item characteristic curves for four matching groups: Ijaw ($a = 0.46$, $b = -0.76$), Yoruba ($a = 0.14$, $b = -0.22$), Hausa ($a = 0.24$, $b = 0.43$) and Igbo ($a = 0.46$, $b = -0.15$).

The item characteristic curves shown in figure 45 were not identical with the Ijaw ICC. This, therefore, suggested that the endorsement of the test item differ among the matching group, especially between the Ijaw examinees with the other ethnic group. Nevertheless, the test item was an easy one for the Ijaw examinees while it was more difficult for the Hausa examinees. The item discriminated similarly among the Igbo and Ijaw examinees.

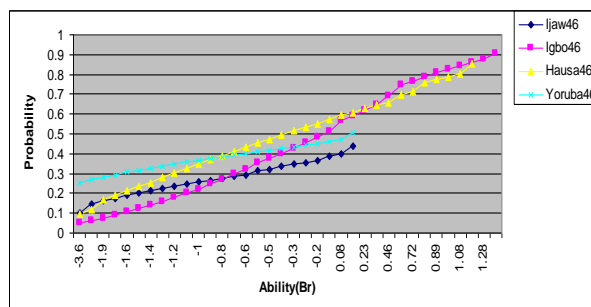


Figure 46: The forty-six item characteristic curves for four matching groups: Ijaw ($a = -0.29$, $b = 0.780$), Yoruba ($a = 0.26$, $b = 0.24$), Hausa ($a = 0.60$, $b = -0.51$) and Igbo ($a = 0.84$, $b = -0.11$).

Figure 46 above showed that ICCs for Ijaw and Yoruba examinees appeared identical while the ICC for Ijaw was not identical with Igbo and Hausa ICCs. Looking at the parameters measured, the test item was easier and better discriminated for the Igbo and Hausa examinees while the item was more difficult for the Ijaw group. The item endorsement showed some form of differential item functioning.

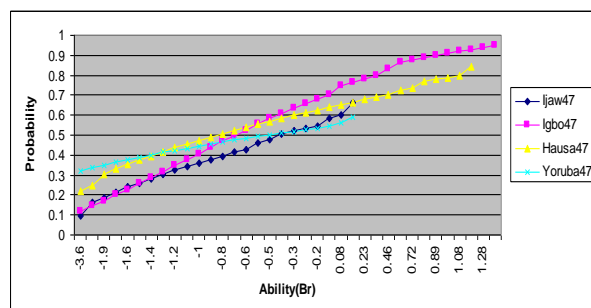


Figure 47: The forty-seven characteristic curves for four matching groups: Ijaw ($a = 0.44$, $b = -0.61$), Yoruba ($a = 0.27$, $b = -0.52$), Hausa ($a = 0.44$, $b = -0.99$) and Igbo ($a = 0.80$, $b = -0.71$).

It was obvious that the area between the item characteristic curve for Ijaw examinees and the three ethnic groups' ICCs was very small. A critical look on the graph suggest that the ICCs were not identical, especially that of Hausa and

Igbo. The item was cheap to all groups and discriminated better in Igbo group. Therefore, the item was flagged a differentially functioning item.

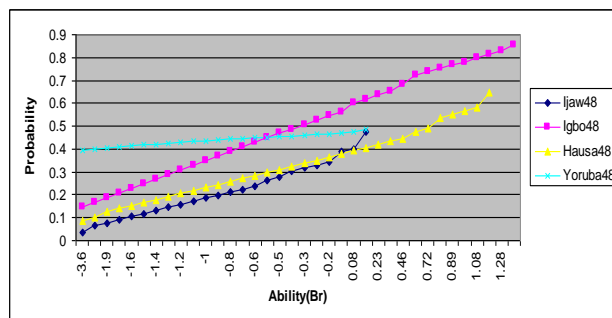


Figure 48: The forty–eight item characteristic curves for four matching groups; Ijaw ($a = 0.49$, $b = 0.41$) Yoruba ($a = 0.09$, $b = 0.67$), Hausa ($a = 0.45$, $b = 0.43$) and Igbo ($a = 0.57$, $b = -0.34$).

Figure 48 has shown four item characteristic curves which are appearing differentially, especially the ICC for Igbo and Ijaw. The ICC for Hausa was closely similar with the Ijaw ICC while the Yoruba ICC met that of the Ijaw at 0.08 ability value. Nevertheless the item discriminate fairly more among the Igbo examinees than others but also indicated that the test item was easier for them. This item showed differential item functioning.

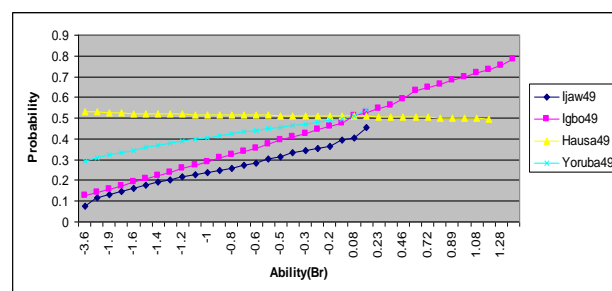


Figure 49: The forty–nine item characteristic curves for four matching groups: Ijaw ($a = 0.35$, $b = 0.58$) Yoruba ($a = 0.25$, $b = -0.11$), Hausa ($a = -0.02$, $b = 0.89$) and Igbo ($a = 0.52$, $b = 0.03$).

There was a clear observation that the item characteristic curves in figure 49 were not identical and hence showed differential item functioning. The Hausa ICC was flatter indicating that the item discriminated poorly among the examinees. The item discriminated better among the Igbo examinees. The test item was easier for the Hausa examinees but more difficult for the Ijaw examinees.

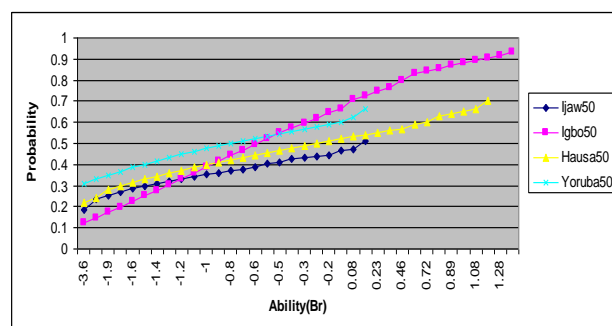


Figure 50: The fifty item characteristic curves for four matching groups: Ijaw ($a = 0.23$, $b = 0.19$), Yoruba ($a = 0.35$, $b = -0.83$), Hausa ($a = 0.32$, $b = -0.37$) and Igbo ($a = 0.74$, $b = -0.62$).

Item 50 shown in figure 50 indicated as examinees performing differentially. The ICCs were not identical, though, the area between them was small. Nevertheless, the test item was easier for the Yoruba group and was more

difficult for the Ijaw group. The item discriminated better among the Igbo group than the other ethnic groups. This was a differential item functioning item.

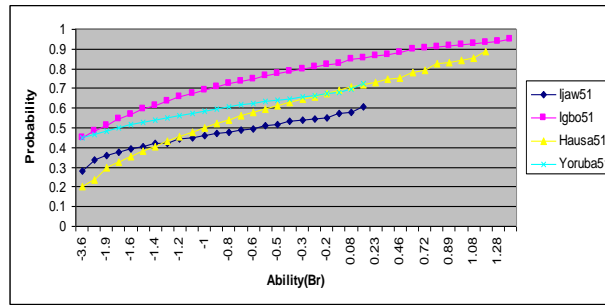


Figure 51: The fifty-one item characteristic curves for four matching groups: Ijaw $a = 0.21$, $b = -0.94$), Yoruba ($a = 0.28$, $b = -1.73$), Hausa ($a=0.52$, $b = -1.13$) and Igbo ($a=0.50$, $b =-1.93$).

A visual inspection of figure 51 showed that the four item characteristics curves were progressing on the ability continuum similarly. The item was easy for all the ethnic groups’ examinees. However, the item discriminated better among the Hausa and Igbo examinees. Nevertheless, it can be deduced that there was obvious differential functioning among comparing ethnic groups.

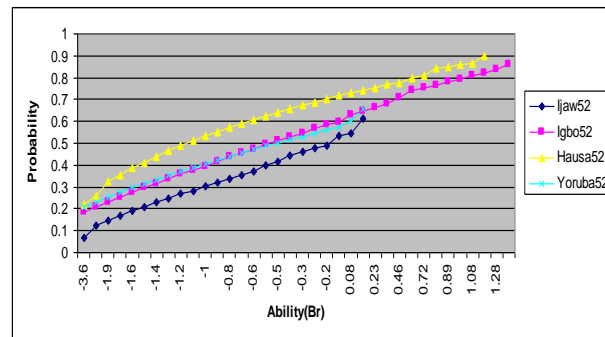


Figure 52: The fifty-two item characteristic curves for four matching group: Ijaw ($a= 0.46$, $b =-0.30$) Yoruba ($a=0.47$, $b =-0.52$), Hausa ($a = 0.52$, $b =-1.27$) and Igbo ($a = 0.53$, $b =-0.52$).

The figure 52 above clearly showed ICC of Ijaw and ICC of Yoruba identical. Also, the ICCs of Hausa and Igbo look identical with the Ijaw ICC through the item discriminated better among their examinees than the Ijaw examinees. The test item was an easy one for all groups but seems easier for the Hausa examinees.

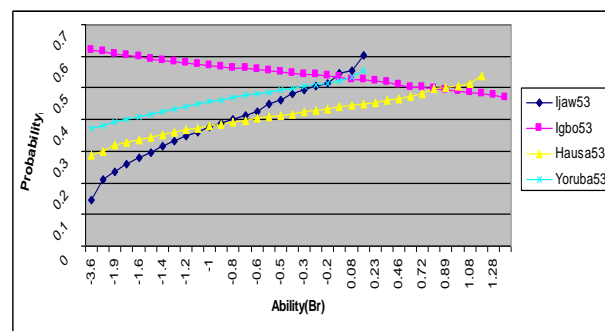


Figure 53: The fifty-three item characteristic curves for four matching group: Ijaw ($a = 0.33$, $b = -0.46$), Yoruba ($a = 0.18$, $b =-0.42$), Hausa ($a = 0.16$, $b =- 0.68$) and Igbo ($a=-0.10$, $b =0.72$).

It was obvious that figure 53 showed item characteristics curves that were not identical. This item showed a clear DIF among comparing ethnic groups. The Igbo ICC showed that, the item was earlier for the low ability group than the

high ability levels, indicating poor item discrimination. Meanwhile, the item discriminated better among the Ijaw examinees.

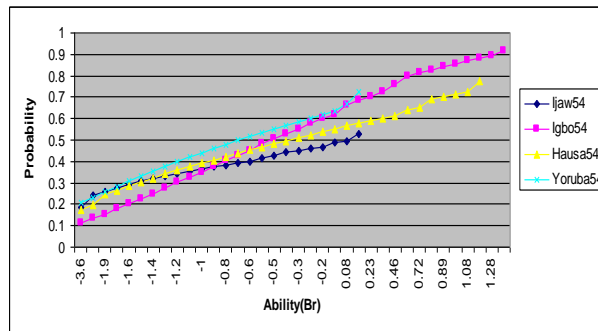


Figure 54: The fifty-four item characteristic curves for four matching groups. IJAW ($a=0.24$, $b = -0.01$) Yoruba ($a=0.55$, $b= -0.74$), Hausa ($a=0.42$, $b=-0.51$) and Igbo ($a=0.72$, $b =-0.48$).

Item 54 displayed a differential item functioning among comparing groups. Though, the area between identical curves was very small but they were not identical. The item discriminated better among the Igbo examinees than the other ethnic groups. It was flagged a DIF item.

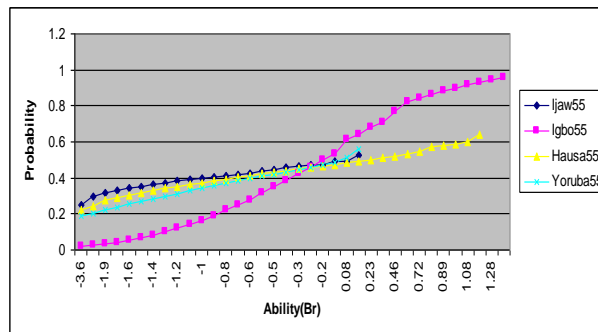


Figure 55: The fifty-five item characteristic curves for four matching groups: Ijaw ($a=0.18$, $b = -0.01$) Yoruba ($a=0.41$, $b =-0.06$), Hausa ($a=0.27$, $b=-0.00$) and Igbo ($a=1.15$, $b =-0.15$).

In figure 55, the ICC curves for the Ijaw, Yoruba, Hausa and Igbo groups were not the same, the four curves met at -0.3 ability value. The area between the Ijaw ICC and that of Yoruba and Hausa was smaller than that of Igbo. The test item was a bit easier for the Igbo group than the Ijaw group. Looking at the four ICC curves, the item was better discriminated among the Igbo examinees.

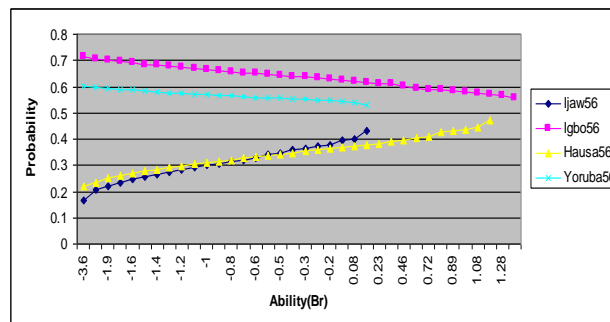


Figure 56: The fifty-six item characteristic curves for four matching groups: Ijaw ($a=0.20$, $b =1.12$), Yoruba ($a=-0.07$, $b =1.34$), Hausa ($a=0.17$, $b = 1.64$) and Igbo ($a=-0.11$, $b = 2.68$).

Item 56 was easier for the Igbo group than the Ijaw group. In fact, the low ability group in the Igbo and Yoruba groups have a high probability of correct response than the low ability Ijaw and Hausa groups. The item never discriminated well in the Igbo and Yoruba groups. The item characteristic curves were not identical hence the item was flagged DIF.

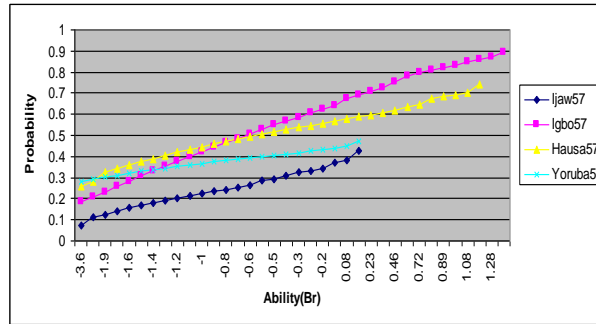


Figure 57: The fifty-seven item characteristic curves for four matching groups: Ijaw ($a = 0.34, b = -0.78$), Yoruba ($a=0.20, b = 0.60$) Hausa ($a=0.32, b = -0.73$) and Igbo ($a=0.58, b = -0.66$).

In figure 57, the ICC curves for the four ethnic groups were not the same. The ICC curves for Igbo, Yoruba and Hausa have a higher probability of correct response at the lower ability level than the Ijaw group. In fact, the test item was easier for the three groups than the Ijaw group. That is, the test item was more difficult for the Ijaw item showed DIF among matching groups.

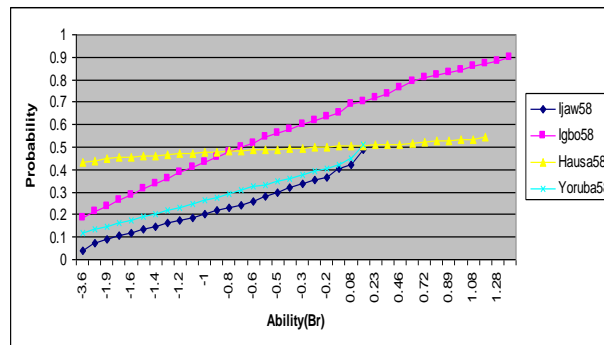


Figure 58: The fifty-eight item characteristics curves for four matching groups: Ijaw ($a=0.47, b=0.34$), Yoruba ($a=0.49, b=0.24$), Hausa ($a=0.07, b=-0.33$) and Igbo ($a=0.59, b=-0.71$).

For item 58, the four ICC areas are not identical, though that of Yoruba and Ijaw were a bit similar. Nevertheless, the test item was bit easier for Igbo, and Yoruba but difficult for Ijaw group. Through the item discriminate fairly in Ijaw group than Hausa group while more fairly in Igbo and Yoruba. The item was a difficult item.

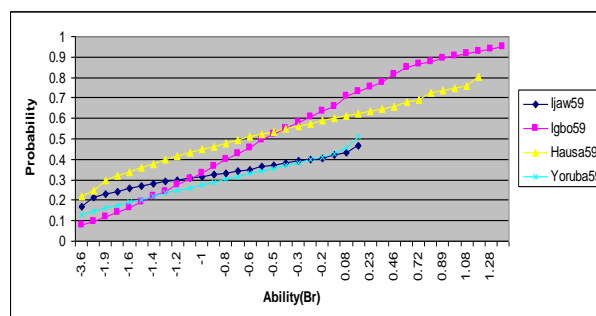


Figure 59: The fifty-nine item characteristics curve for four matching group: Ijaw ($a=0.22$, $b=0.68$), Yoruba ($a=0.46$, $b=0.24$), Hausa ($a=0.40$, $b=-0.82$) and Igbo ($a=0.88$, $b=-0.52$).

In figure 59, the item characteristic curves for Ijaw and Yoruba appeared identical while Ijaw ICC differs distinctly from ICC for Hausa and Igbo. This item show differential functioning. Nevertheless, the test item was a bit easier for the Hausa and Igbo group then Ijaw and Yoruba groups. The ICC for Igbo crossed the ICC for Ijaw indicating a non-uniform different.

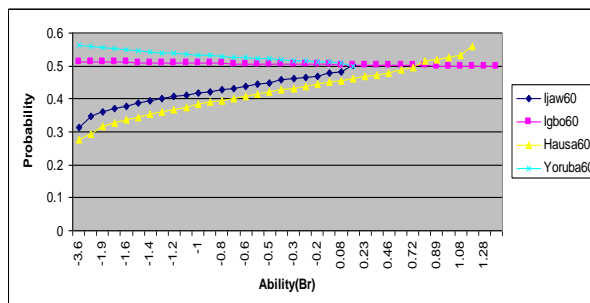


Figure 60: The sixty item characteristic curves for four matching groups: Ijaw ($a=0.12$, $b=0.26$), Yoruba ($a=-0.06$, $b=0.30$), Hausa ($a=0.18$, $b=0.43$), and Igbo ($a=-0.01$, $b=0.98$).

Both ethnic groups curves were shifted up horizontally, especially that of Yoruba and Igbo. This was as a result of the low ability level having a higher probability for correct answers than the higher ability group. The test item discriminated partially better for ICC Ijaw and Hausa than others. The item was flagged a DIF item.

After the DIF analyses, using the item characteristics curve method, it was observed that 56 items, (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60) out of the 60-Item Chemistry Achievement Test were flagged as displaying differential item functioning among the comparing ethnic groups. Only 4 items (items -11, 36, 42, & 52) were detected not showing differential item functioning.

4. Discussion

The study findings to this research questions identified 56 items showing DIF in one way or the other. That is to say, the DIF for the item may be due to difference between the focal group ICC and one or more ICC(s) for the reference groups. The ICCs were displayed in a graph (refer figure 1-60). The item characteristic method flagged 56 items biased because the four matching groups ICCs were not appearing identical in all the items concern. However, the study was not able to compute the area index between the focal group ICC and the reference group ICCs. Nevertheless, on visual inspection the researcher was able to identify non identical curves. Abedalaziz (2011), in his work, identified DIF items with visual inspection. Similarly, Adedoyin (2011) used visual inspection to flag the ICC(s) of items that were biased.

Consequently, the study observed in the ICC graphs that they vary in terms of their position and slope. Zumbo (1999) justified this observation that parametric ICCs vary in terms of position on the X-axis, their slope and their intercept with the Y-axis. Zumbo further stated that DIF is assessed by comparing the ICCs of different groups on an item. This study compared ICCs of Ijaw examinees (focal group) with the ICCs of Hausa, Igbo and Yoruba (reference group).

Moreover, Zumbo (1999) declared that “If the ICCs are identical for each group, or very close to identical, it can be said that the item does not display DIF. if, however, the ICCs are significantly different from one another across groups, then the item is said to show DIF (p. 19)”. This assertion strengthens the findings of the study for flagging 4 items non-DIF items and 56 as displaying DIF. These items flagged as showing DIF were also seen as items showing bias across the groups. Adedoyin (2010) stated that if the ICCs for two populations are the same, the item is not biased but if the ICCs are different, the item is biased. Arguably, it has been observed that all DIF items may not be biased items but biased items should always be flagged DIF (Schumacker, 2005; Swaminathan & Rogers, 1990, Zumbo, 2007).

The study was able to identify items that established differential item functioning but were as well flagged as biased items. IRT-based ICC curve method detected 56 items (93.3%) as DIF/biased items

5. Conclusion

From the results, this current study concludes that the IRT- Based Item Characteristic curve method detected 56 items biased because the various ICCs were not identical when compared between the focal group (Ijaw examinees) and the reference group (Ibgo, Hausa & Yoruba examinees). Thus there was ethnic bias in the Chemistry Achievement Test (CAT).

6. Recommendation

We therefore, recommend that IRT-based Item Characteristic Curve method be applied in Nigerian educational testing. Since, in Nigeria IRT methods are still new, workshops, seminars and conferences should be organized for researchers in educational testing.

REFERENCES

- i. *Abedalaziz, N. (2011). Detecting DIF using item characteristics curve approaches. The Internationals Journal of Educational and Psychological Assessment, 8(2), 1 – 15.*
- ii. *Adedoyin, O. O. (2010). Using IRT approach to detect gender biased items in public examinations: A case study from the Botswana junior certificate examination in Mathematics. Educational Research and Review, 5(7), 385-399.*
- iii. *Anastasi, A. & Urbina, S. (1997). Psychological Testing. India: Prentice – Hall.*
- iv. *Brown, A. (2012). Measurement invariance: university of Cambridge, department of psychology*
- v. *Cohen, L.A. (2006). Item bias and differential item functioning. In T.M. Wood & W. Zhu (Eds), Measurement theory and practice in kinesiology.*
- vi. *Orluwene, G.W. (2007). Application of two-parameter Latent trait model in development and validation of chemistry achievement test. Unpublished Ph.D Dissertation, University of Port – Harcourt.*
- vii. *Pedrajita, J.Q. (2009). Using Logistic regression to detect biased test items. The Internationals Journal of Educational and Psychological Assessment 2, 54-73.*
- viii. *Penfield, R.D., Alvarez, K. & Lee, O. (2009). Using a taxonomy of differential step functioning to improve the interpretation of DIF in polytomous items: An illustration. Applied Measurement in Education, 22, 61-78.*
- ix. *Queensoap, M. (2014). Application of differential item functioning in detecting chemistry achievement test in Nigeria. Unpublished doctoral dissertation, University of Port-Harcourt, Nigeria.*
- x. *Schumacker, R.E. (2005). Test bias and differential item functioning. Retrieved October 11, 2010, from a. <http://www.appliedmeasurementassociates.com.pdf>.*
- xi. *Swaminathan, H., & Rogers, H.J. (1990). Detecting differential item functioning using logistic regression procedures. Journal of educational measurement, 27,361-370.*
- xii. *Zumbo, B.D. (1999). A handbook on the theory and methods of differential item functioning (DIF):logistic regression modeling as a unitary frame work for binary and likert-type (ordinal) item scores. Canada: Directorate of Human Resources Research and Evaluation.*
- xiii. *Zumbo, B.D. (2007). Three generations of DIF analyses: considering where it has been, where it is now, and where it is going. Language Assessment Quarterly. 4 (2), 223-233.*