A Three Model Parameter Of Chemistry Achievement Test Among Secondary School Students In Ibadan Metropolis, Nigeria

Taiwo Adebukola Kabir

Abstract- The purpose of the study was to investigate the item discrimination /difficulty of each item in chemistry achievement test and to find out whether the guessing ability of each student has relationship in answering chemistry achievement test. The study adopted a descriptive survey design. A sample of 250 participants was randomly selected across twenty-five public secondary schools in Ibadan metropolis. Data were collected using"Secondary School Students chemistry Achievement Test (SSSCAT)" developed by the researcher. Data were analyzed using statistical package of the Social Sciences (SPSS24), polychoric correlation coefficient and JMetrik analysis. The results revealed that the item fitted 3-PL model the most because this model provided the least -2 log likelihood as a result of the model fit assessment conducted and it was discovered that 27 items out of the 30 items of the chemistry achievement test survived under IRT framework. It was recommended that IRT should be used in test development processes as it enhances the validity and reliability of tests which provides quality assurance in measurement and also Using the three parameter model will greatly improve characterization of items, item selection, and in turn lead to improved measures which are the aim of test developers.

Index Terms- Item Response Theory, Item Parameters, Chemistry Achievement Test

I. INTRODUCTION

Chemistry is an important school subject and its importance has been widely reported through the applications of principles in scientific and technological development of any nation (Asiyaii, 2005). The importance of Chemistry to humanity cannot be over-emphasized (Anne,2005). According to Smita as cited in Jason (2013), the three basic needs of humans i.e food, shelter and cloth are made by chemical processes using chemicals and fibres. This therefore, means that Chemistry is always present around us and it is involved in the manufacture of all man-made objects and things. Chemistry is required for variety of science and technology degrees, including all medical related fields, engineering, and science majors.

However, in spite of the strategic position occupied by Chemistry in sciences and other disciplines, the performance of students in the subject has been observed not to be encouraging over the years. Studies (Nbinna, 2012, Ojerinde, 2013) have shown that the performance of Nigerian students in the subject at the secondary school remains a dismal failure. The statistics available shows that about 50% of candidates, who write the May/June West African Senior Secondary School Certificate Examination (WASSCE) each year, scored below 50%. Despite efforts made at all levels of Government across nation most especially Nigeria to encourage students to study science and despite the students interest in science, there is still poor performance, especially in Chemistry. In Chemistry, balancing chemical equations and taking precautions during experiment is a challenge to most students in Nigeria.

It has been observed that the general level of performance of Nigerian students in chemistry subject is still not good enough. The reasons for this poor performance were due to several factors as identified by some authors. These factors include school location (Ndukwu and Odinko, 2002; Odinko, 2002), practical class (Kwale SMASSE, 2005; Bhagwan, 2005; Twoli, 2006), attitude and performance (Hussain, Ali, Khan, Ramzan & Qadeer, 2011; Olatunde, 2009), teaching strategies (Brenner, 2004; Twoli, 2006), assessment (Beck & Earl, 2002; Black, 2002; Brenner, 2004), government interventions (KESSP, 2005; Audinos, Lairez & Makwati 2003; Adams & Chapman, 2002) and students characteristics (SMASSE, 2004; Twoli 2006).

Considering the importance of chemistry to humanity, the need to develop an instrument which will reliably and validly measure students' ability in chemistry cannot be over emphasized so as to improve the performance in chemistry and increase the success rate for the courses requiring chemistry. In this case, today's secondary schools chemistry achievement test without any doubt requires field test to try out new items so as to know how difficult the item is or to spot out any inconsistency in the item in which in turn the empirical data generated can be used to inform which items can be used in operational form and which items needs to be changed or abandon when setting questions in actual operation.

This study adopts the three-Parameter Logistic Model (3-PLM) of Item Response Theory in the Development and Standardization of Chemistry Achievement Test. The choice of three-Parameter Logistic Model (3-PLM) for this study was based on the fact that Chemistry content as indicated in the scheme of work graduate in difficulty i.e. from simple to complex. So, if item discrimination is kept constant, an instrument should be able to measure the learning progress of students in chemistry, taken notice of the content areas.

Achievement tests are considered very important as a measuring instrument to be used in the evaluation process and a compatible test must be taken into consideration with the

psychological ability of the students. In order to reduce the low performance in achievement test in chemistry, the guessing ability of the students must be taken into account. Despite the fact that the item response theory are been used by test developers, attention has not been drawn towards the achievement tests used in secondary schools in measuring the presence or absence of identified trait of interest, considering the precision made by the instrument constructed most especially chemistry achievement test. In addition, there are few studies conducted in Nigeria that specifically looked at three parameter model of chemistry achievement test in relation to know connection between the latent ability a student has in answering an item. Thus, this study, was carried out to investigate the three parameter model in answering an item in achievement test in secondary school chemistry in Ibadan metropolis of Oyo state.

II.   LITERATURE REVIEW

Concept of Achievement Test

Achievement test has been used in different disciplines to assess or measure the cognitive capabilities of an individual on a particular instruction such an individual has been previously exposed to. It is a kind of test designed to assess current or present level of knowledge, skills, mastery or competency after a specific course of instruction (academic area). This kind of test is often used as an indicator of previous learning and is often used to predict future performance (lecture note on measurement and evaluation, 2017).

In addition, Achievement test scores are often used in an educational system to determine the level of instruction for which a student is prepared. High achievement scores usually indicate a mastery of grade-level material, and the readiness for advanced instruction. Low achievement scores can indicate the need for remediation or repeating a course grade. The most common type of achievement test is a standardized test developed to measure skills and knowledge learned in a given grade level, usually through planned instruction, such as training or classroom instruction Achievement tests are often contrasted with tests that measure aptitude, a more general and stable cognitive trait (Hawaii Department of Education, 1999).

Concept of A Model Parameter

A parameter generally, is any characteristic that can help in defining or classifying a particular system (meaning an event, project, object, situation, etc.). That is, a parameter is an element of a system that is useful, or critical, when identifying the system, or when evaluating its performance, status, condition, etc (Kpolovie, 2010, 2011; Stephanie, 2013, Georgiev, 2008). Parameter has more specific meanings within various disciplines, including mathematics, computing and computer programming, engineering, statistics , logic and linguistics . Within and across these fields, careful distinction must be maintained of the different usages of the term parameter and of other terms often associated with it, such as argument, property, axiom, variable , function, attribute, etc. Statistically, a parameter is a quantity, such as a mean, that is calculated from data and describes a population (wikipedia).

A model parameter is a configuration variable that is internal to the model and whose value can be estimated from data.

They are required by the model when making predictions. Their values define the skill of the model of researchers problem. They are estimated or learned from data and are often not set manually by the practitioner. They are often saved as part of the learned model. Parameters are key to machine learning algorithms and they are the part of the model that is learned from historical training data. In classical machine learning literature, we may think of the model as the hypothesis and the parameters as the tailoring of the hypothesis to a specific set of data. Often model parameters are estimated using an optimization algorithm, which is a type of efficient search through possible parameter values (Kpolove,2016).

Concept of Item Response Theory

Item response theory is generally claimed as an improvement over classical test theory (CTT). For tasks that can be accomplished using Classical test theory, item response theory generally brings greater flexibility and provides more sophisticated information. Some applications, such as computerized adaptive testing, are enabled by item response theory and cannot reasonably be performed using only classical test theory. Another advantage of item response theory over classical test theory is that the more sophisticated information item response theory provides allows a researcher to improve the reliability of an assessment .Among other things, the purpose of item response theory is to provide a framework for evaluating how well assessments work, and how well individual items on assessments work. The most common application of item response theory is in education, where psychometricians use it for developing and designing exams, maintaining banks of items for exams, and equating the difficulties of items for successive versions of exams (for example, to allow comparisons between results over time), (Hambleton and Jones,1993, Ojerinde, 2013).

According to, Thorpe and Favia (2012) describe item response theory (IRT) as a collection of measurement models that try to explain the connection between observed item responses on a scale and an underlying construct. IRT models are mathematical equations that specifically describe the association between subjects' levels on a latent variable and the probability of a particular response to an item, using a non-linear monotonic function (Hays *et al.*, 2009). As a measurement theory IRT aids to develop an educational measurement scale that has ratio scale, sample independent attributes and students' ability reported on both item and total instrument levels (Ani, 2014). IRT makes it possible to estimate and interpret item statistics referred to as parameters. In IRT, parameters of the persons are invariant across items, and parameters of the items are invariant in different populations of persons. Thus it brings greater flexibility and provides more sophisticated information which allows for the improvement of the reliability of an assessment (Ani, 2014).

Item parameters in IRT are estimated directly using three logistic models instead of proportions (difficulty or threshold) and item-scale correlations (discrimination) (Adedoyin and Mokobi, 2013). In one-parameter logistic (1PL) model, items are only described by a single parameter in terms of location or difficulty (b) while the slopes (discrimination) are held constant. The two-parameter logistic (2PL) model not only checks for the item difficulty but also assess each item's level of discrimination between high and low ability students while the three-parameter

logistic (3PL) model adds a third item parameter which is called pseudo-guessing parameter that reflects the probability that an examinee with a very low trait level will correctly answer an item solely by guessing. This according to Ani (2014) implies that students can through guessing answer an item correctly in an achievement test.

Item response theory (IRT) (also known as latent trait theory, strong true score theory, or modern mental test theory) is a paradigm that is conceptual framework for the design, analysis, and scoring of tests , questionnaires, and similar instruments measuring abilities, attitudes, or other variables. It is a theory of testing based on the relationship between individuals' performances on a test item and the test takers' levels of performance on an overall measure of the ability that item was designed to measure. Several different statistical models are used to represent both item and test taker characteristics. Unlike simpler alternatives for creating scales and evaluating questionnaire responses, it does not assume that each item is equally difficult. This distinguishes item response theory from, for instance, Likert scaling, in which " All items are assumed to be replications of each other or in other words items are considered to be parallel instruments" . In comparism, item response theory treats the difficulty of each item (the item characteristic curves, or ICCs) as information to be incorporated in scaling items.It is based on the application of related mathematical models to testing data (Birbanum,1968).

Furthermore, Item response theory is based on the idea that the probability of a correct or keyed response to an item is a mathematical function of person and item parameters . The person parameter is construed as (usually) a single latent trait or dimension. Examples include general intelligence or the strength of an attitude. Parameters on which items are characterized include their difficulty (known as "location" for their location on the difficulty range); discrimination (slope or correlation), representing how steeply the rate of success of individuals varies with their ability; and a pseudo-guessing parameter, characterizing the (lower) asymptote at which even the least able persons will score due to guessing (for instance, 25% for pure chance on a multiple choice item with four possible responses). In the same manner, Item response theory can be used to measure human behaviour in online social networks. The views expressed by different people can be aggregated to be studied using item response theory. Its use in classifying information as misinformation or true information has also been evaluated (Hambleton etal,1995).

The conceptual basics, assumptions, and expansion of the basic premises of classical test theory have paved way for the construction of some outstanding psychometrically sound scales in the measurement practices of educational bodies in Africa; This is owing to the simplicity of interpretation which can usually be applied to testees achievement and aptitude test performance. Despite the popularity of classical item statistics as a fundamental part of standardized testing and measurement framework, it is fraught with many shortcomings. Among these are the values of standard item parameters that are not invariant across groups of test takers that vary in ability. The invariance characteristic of item parameters demands that the estimate of the parameter of an item across two or more groups of population that vary in ability must be the same (Emekene,2016).

Item response theory is a model which expresses the probability of the relationship between an individual's answering an item and the underlying latent trait often called ability measured by the instrument. In IRT, the true component is defined on the latent variable of interest rather than on the test, as in the case in classical test theory. Item response theory entails three assumptions (Bomo,2015).
1. A unidimensional trait
2. Local independence of items;
3. The response of a person to an item can be modeled by a mathematical item response function (IRF).

Guessing parameter

In educational testing literature, guessing is presumed to occur when a test taker does not absolutely know the correct response but still tries to arrive at the right answer (Hutchinson, 1991; Maris, 1995; San Martin et al., 2006). There are several ways to conceptualize the process for problem solving and guessing and they revolve around the question of whether the guessing process(GP) comes before or after the problem solving process(PSP).What is commonly found in the literature is the presumption that the guessing process is based on knowledge1 that is insufficient to complete the problem solving process successfully. In this conceptualization, the degree of incompleteness of knowledge would be associated with a test-taker's proficiency being measured, so the GP becomes the interaction between test taker and item. Lord (1974) noted that c-parameter estimates were often smaller than the value that would result if a test taker guessed completely at random—probably because low-proficiency test takers were likely to exhibit a pattern of choosing attractive but incorrect choices. Taking this line of conceptualization a step further, San Martin et al. (2006) came up with the one-parameter logistic model with ability-based guessing, or1PL-AG model, where the interaction between a person's proficiency and guessing was taken into account. Interpreting the c parameter as an interaction between examinee and item rather than as one of item properties is problematic, however, because a- and/or b- parameters cannot be viewed purely as item properties— a- and b-parameter estimation is inseparable from c-parameter estimation.

In theory, the item parameters of the 3PLM are independent of one another and independent of a person's proficiency in the mathematical forms of the response models. But, when it comes to the parameter estimation procedure and the maximum likelihood algorithm attempts to find IRF best fitting to response data, the effect of person's proficiency on the c parameter estimates would influence the other item parameter estimates, as well.

III. RESEARCH QUESTIONS

1. To what extent do the items of chemistry test obey item unidimensionality assumption of IRT framework?
2. To what extent do the items of chemistry achievement test obey item local dependence assumption of IRT framework?
3. What is the item difficulty and item discrimination parameters estimated by IRT?

4. What is the best available items for the final operation form in chemistry achievement test on the basis of criteria set for IRT framework?

Research Design

A survey design was adopted for this study. Survey design was on relative elements of the population with common attributes which were chosen with a view to representing the entire population. Moreover, the outcome of the study and selective group was adequate and sufficient which was used as a basis for generalization. The survey method paved way for the researcher to make use of questionnaire. Descriptive analysis was used to analyze the data generated from demographic characteristics of respondents while correlation analysis was used in order to capture in a preliminary manner, the nature of relationship between the variables. This design was considered appropriate for the elicited responses from the respondents which enhanced the systematic investigation of the three parameter model of chemistry achievement test in secondary school in Ibadan metropolis, Oyo State, Nigeria.

Population

The population of this study consisted of all public secondary school students in Ibadan Metropolis of Oyo State.

Sample and Sampling Techniques

The sample for the study comprised two hundred and fifty (250) public secondary school participants selected by stratified and simple random sampling procedures. There are 5 local government areas in Ibadan metropolis, Nigeria. Five public secondary schools were selected from each of the 5 local government areas. This makes a total of 25 public secondary schools randomly selected in Ibadan Metrolis of Oyo State. Ten Chemistry students were stratified from each selected public secondary schools.

Instrument

The research instrument that was used in the study was "Secondary School Students Chemistry Achievement Test (SSSCAT)". The instrument was a 30-items multiple choice chemistry achievement test, which was developed by the researcher. The instrument contains 30-4 options multiple-choice test which was developed using SS1 Chemistry Syllabus and the instrument was dichotomous. The number of options was maintained at four to give the probability of selection of a key (correct answer) by mere guess as one-fourth or 25 per cent. By so doing, this has reduced the burden (difficulties) on the researcher in providing options for items in the study. A test blue print and table of specification showing topics and number of items per cognitive level was prepared by the researcher. The table of specification was used as guide to develop a specific number of multiple choice test items to cover the various levels of Bloom's taxonomy of cognitive objectives.

The instrument was subjected to both content and face validation in order to estimate the psychometric properties of the instruments. The instruments was subjected to the scrutiny by the experts in Science and Technology Education Department and Institute of Education, University of Ibadan for content and face validity before final administration to the respondents.

The reliability of the items was established by conducting a pilot study on students in the schools that were not included in the sample. The complete instrument was tested through Kuder Richardson method and the reliability was ensured with coefficient of 0.84 before the final copy was produced. The result of reliability coefficients of each factor was used after the necessary modification of the research instrument before final administration to research.

Data Analysis

The instrument administered was retrieved, counted and crosschecked to determine attrition rate then coded before analysis. Demographic variables of respondents was analyzed by percentage distribution and frequency counts. Statistical analysis was carried out using polychoric correlation coefficient and JMetrik analysis.

IV. RESULTS

Research Question One

To what extent do the items of chemistry test obey Item unidimensionality assumption of IRT framework?

To answer this research question, factor analysis was to be computed using SPSS version 24, but it was impossible to conduct a factor analysis in the initial because of high number of redundant items which generated factorability error. To resolve this an item-total correlation was done to remove redundant items; which discovered 20 redundant items. This implies that Chemistry achievement test did not satisfy the criteria for item unidimensionality. To further justify the number of factor loading extracted a screen graph was plotted.
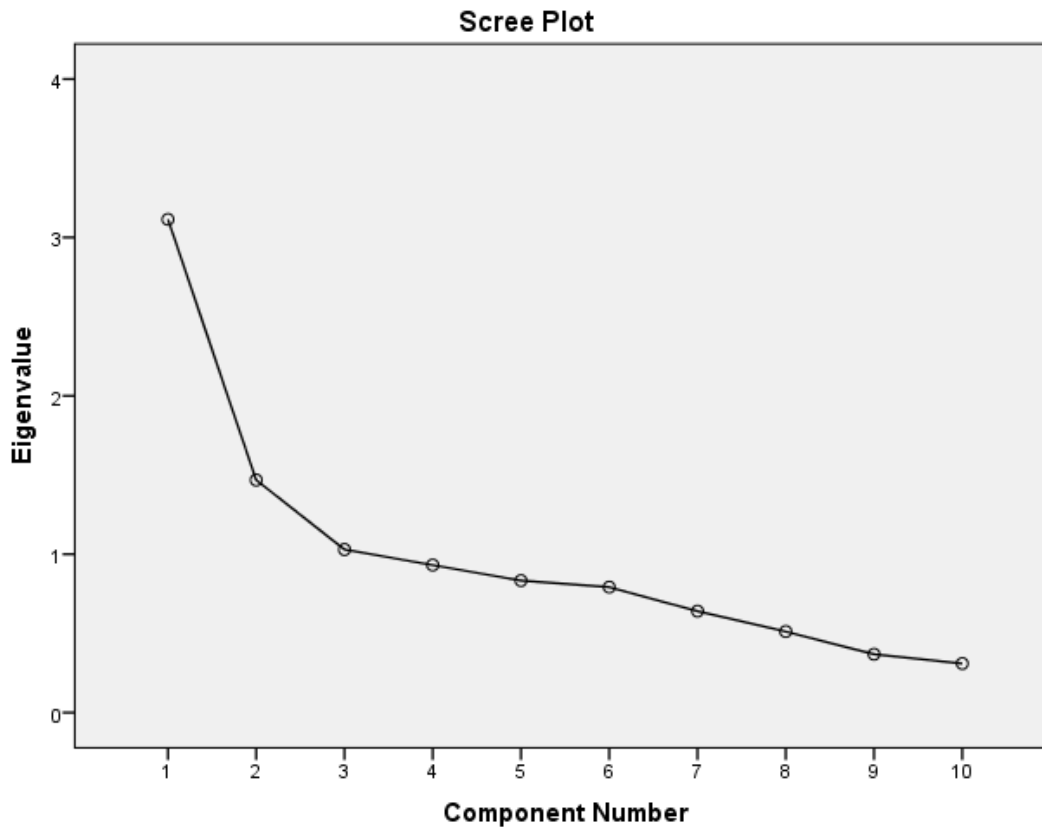
Scree Plot

Figure 1 showing the screen plot revealed that based on the first elbow (first line brake), items of the Chemistry achievement test cannot load on one factor but on two factor having the eigenvalue above 1.0 in view. The scree-plot suggested that Chemistry achievement test items loaded on two component which did not satisfy the criteria for unidimensionality criteria of the item response theory (IRT).

Research Question Two
To what extent do the items of chemistry test obey Item local dependence assumption of IRT framework?
To answer the second question on IRT criteria for local dependence. A polychoric correlation was done using LISREL version 8.80.

Table 1: Polychoric Correlation of 30items of the Chemistry achievement test

Items	1	2	3	4	5	+	27	28	29	30
1	1.00									
2	.00	1.00								
3	.00	.360**	1.00							
4	.00	.110	.018	1.00						
5	.00	.170*	.095	.341**	1.00					
+						+				
27	.00	.434**	.225**	.087	-.057		1.00			
28	.00	.282**	.028	-.124	.005		.391**	1.00		
29	.00	.285**	.123	-.157*	-.168*		.307**	.494**	1.00	
30	.00	.453**	.195*	-.006	.010		.593**	.351**	.411**	1.00

Table 1 reveals the polychoric relationship that exist among the 30 items of Chemistry achievement test. From Item 1, there are 49 items correlation coefficients; that is, it correlated with item 2, item 3, item 4 and so on up to item 30. Table 2 further presents the summary of frequencies of the observed polychoric correlations among the 30 items.

For instance item 2, there were 47 correlation coefficients in that it correlates with items 2, 3, 4, and so on up to items 30.

Among the 47 correlation coefficients, 11 have values that are less than or equal to 0.099; 17 have values that are less than or equal to 0.199, and have also values that are less than or equal to 0.299 and so on. The table presents the polychoric correlations between item 2 and 3, 4, 5, …..30. The tables explains the inter-correlation between other preceding items and the correlations with the subsequent items. In all there were 900 correlations coefficients.

Table 2: Summary of Polychoric correlation coefficients among the 30 items

Item	≤0.099	≤0.199	≤0.299	≤0.399	≤0.499	≥0.500	No of Correlation
1-30	--	--	--	--	--	--	--
2-30	11	17	10	4	4	1	27
3-30	21	12	9	3	1	--	26
4-30	15	16	6	5	3	--	25
5-30	13	21	6	2	--	--	22
6-30	10	18	8	4	--	--	20
7-30	17	16	9	1	--	--	23
8-30	20	9	7	3	1	1	21
9-30	19	10	3	4	1	--	37
10-30	19	9	5	3	1	--	37
11-30	15	14	4	2	-	--	35
12-30	13	12	7	1	1	--	34
13-30	14	10	3	5	2	--	34
14-30	17	12	5	1	-	--	35
15-30	13	13	5	2	-	--	33
16-30	11	13	5	1	-	--	30
17-30	14	12	4	-	2	--	32
18-30	13	11	5	1	-	--	30
19-30	17	8	4	1	1	--	31
20-30	17	7	3	1	1	--	29
21-30	13	7	2	2	4	--	28
22-30	11	6	6	1	1	--	25
23-30	12	5	6	4	1	--	28
24-30	11	8	4	2	-	--	25
25-30	10	8	4	1	-	--	23
26-30	9	8	3	1	-	--	21
27-30	12	6	4	1	-	--	23
28-30	7	6	3	1	-	--	17
29-30	6	8	3	2	1	--	20
30-30	6	7	6	-	-	--	19
Total	120	100	109	73	43	2	900
Percentage	39.92%	35%	15.75%	6.1%	3.6%	0.16%	100%

Out of 900 correlations among the 30 items, 120 (representing 39.92%) have polychoric correlation coefficient equal or less than 0.099; 100 (representing 35%) have polychoric correlation coefficient equal or less than 0.199; 109 (representing 15.75%) have polychoric correlation coefficient equal or less than 0.299; 73(representing 6.1%) have polychoric correlation coefficient equal or less than 0.399; 43 (representing 3.6%) have polychoric correlation coefficient equal or less than 0.499; 2 (representing 0.16%) have polychoric correlation coefficient equal or less than 0.500. Considering the rule of thumb associated with correlation effect size (when r= 0.3-0.44 it is called moderate correlation and when r= 0.5 upward is large correlation) the percentage of significant correlations ranging between 0.3-0.5 (i.e

6.1%-0.16%) is very small. Therefore Chemistry achievement test displayed poor local dependence since the items of the scale refuse to display strong correlation value within itself.

Research question 3: What is the item difficulty and item discrimination parameters estimated by IRT?

To answer the question 30-items on Chemistry achievement test were subjected to analysis using JMetrik. As a result of the model fit assessment conducted, the item fitted 3-PL model the most because this model provided the least -2 log likelihood. The 3PL for IRT respectively. Table 4.4 presents these statistics. The IRT, *b* represents the difficulty parameter, **a** represents the discriminating parameter and **c** represents the guessing factor.

Table 3: Item Parameter Estimate Summary for IRT models on 30-items of Chemistry achievement test.

Item No	IRT		
	C	A	B
1	1.00	1.01	-0.54

2	0.07	1.76	-0.42
3	0.15	1.85	0.47
4	0.26	0.47	-0.01
5	0.50	0.68	-1.63
6	0.50	0.58	-1.73
7	0.01	**0.82**	**29.54**
8	0.25	1.77	-0.39
9	0.20	2.19	1.81
10	0.04	1.35	5.96
11	0.50	**0.91**	**-2.19**
12	0.12	2.03	0.67
13	0.50	-0.73	-0.33
14	0.50	0.60	-1.84
15	0.50	0.85	-1.95
16	0.24	0.07	-2.95
17	**0.01**	**0.82**	**29.54**
18	0.50	0.14	-0.98
19	0.23	0.49	-0.72
20	0.29	1.86	1.21
21	0.79	0.72	0.20
22	0.21	2.29	2.10
23	0.16	2.38	2.13
24	0.25	0.25	-0.43
25	0.22	0.42	-1.11
26	0.25	0.23	2.15
27	0.50	0.78	-0.88
28	0.50	2.11	2.23
29	0.25	1.64	-1.00
30	0.77	1.76	-0.12

Items on **bold** are abnormal items

Table 4.4 reveals the IRT framework five items (7, 11, 17) displayed extreme values on difficulty and the discrimination index

Research Question 4:
On the basis of the criteria set for IRT framework (Model fit and test item function assessment) which and how many of the items survived under IRT framework?
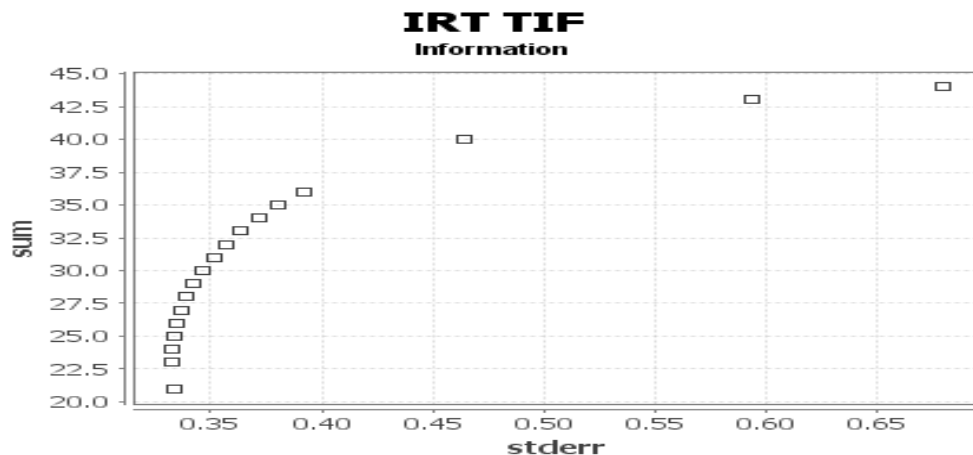
To answer the research question, table 4.4 was reproduced in table 4.5 (see the preceding sections). Looking at the right hand column of the table where IRT statistics are presented, the items that survived the set criteria are rated as being good (see remark column).

Table 4: Item Parameters estimates and Survived Items

Item No	IRT		
	A	**B**	**Remark**
1	1.01	-0.54	Good
2	1.76	-0.42	Good
3	1.85	0.47	Good
4	0.47	-0.01	Good
5	0.68	-1.63	Good
6	0.58	-1.73	Good
7	**0.82**	**29.54**	**Poor**

8	1.77	-0.39	Good
9	2.19	1.81	Good
10	1.35	5.96	Good
11	**0.91**	**12.19**	**Poor**
12	2.03	0.67	Good
13	-0.73	-0.33	Good
14	0.60	-1.84	Good
15	0.85	-1.95	Good
16	0.07	-2.95	Good
17	**0.82**	**29.54**	**Poor**
18	0.14	-0.98	Good
19	0.49	-0.72	Good
20	1.86	1.21	Good
21	0.72	0.20	Good
22	2.29	2.10	Good
23	2.38	2.13	Good
24	0.25	-0.43	Good
25	0.42	-1.11	Good
26	0.23	2.15	Good
27	0.78	-0.88	Good
28	2.11	2.23	Good
29	1.64	-1.00	Good
30	1.76	-0.12	Good

Column 3 of the table 4 presents the assessment of items based on the information each of them contributes to the overall information supplied by the whole test. This assessment requires looking at the test information function (TIF) produced for the 30 items.

From the test information function, items whose difficulty level fall between -0.9 and 3.0 were good and selected while items whose difficulty parameter fall outside of the range were poor and were removed from the items. There were 3 items in this category. The items include: 7, 11, 17. Therefore 25 items survived the IRT framework criteria out of 30 items of Chemistry achievement test.

V. DISCUSSION OF FINDINGS

Research question one examined the extent do the items of chemistry achievement test obey the unidimensionality assumption of IRT framework. The result showed that chemistry achievement test did not satisfy the criteria for item unidimensionality. This is an implication that chemistry test was not driven by one factor; meaning that it was not testing one thing. This can go a long way in affecting the IRT framework because the first assumption is bridged. The result of this study partially

corroborated Wise and Sundre (2003) who found similar reliability coefficients for different degrees of motivation filtering. Conversely, findings from Wise,Wise and Bhola (2006) showed a simultaneous decrease in the internal consistency coefficients and the corresponding standard error of measurement.

Research question two examined to what extent the items of chemistry test obeys item local dependence assumption of IRT. The result revealed that from item 1, there are 49 items correlation coefficients; that is, it correlated with item 2, item 3, item 4 and so on up to item 30. Considering the inter -correlation between other preceding items and the correlations with subsequent items, there were 900 correlations coefficients. In the 900 correlations , the following are the polychoric correlational coefficient equal or less than 0.099 (for 120 correlations), 0.199 (f0r 100correlations), 0.299 (for 109 correlations), 0.399 (f0r 73 correlations), 0.499 (for 43 correlations), 0.500 (for 2 correlations). considering the rule of thumb associated with correlation effect size, the percentage of significant correlations ranged between 0.3-0.5 (6.1%-0.16%) and it is very small, meaning that chemistry achievement test displayed poor local dependence since the items of the scale refuse to display strong correlation value within itself and this suggest that chemistry achievement test to some extent satisfy IRT framework. The result of the findings partially corroborated with that of Wingersky, Barton and Lord (1982).

Research question three examined the item difficulty and item discrimination parameters estimated by IRT. The result revealed that item difficulty ranges between 0.33-0.6 (3.0%-92%). By implication the range of examinees that got the items correctly are between 33% to 66%. Considering the criteria for IRT difficulty index which ranges from (-00 to +00) (baker, 2001), therefore 18 items are considered to be easy and 10 items are considered to be difficult while two of the items shows extreme values of difficulty. The result of the findings on the IRT discrimination index column. Following the criteria for difficulty index which ranges from 0.2 above. Six of the items do not show good discrimination values with poor discrimination power . The result of this findings corroborated Adedoyin and Mokobi (2013) who reported that the discriminations value for a good test items ranges from 0.5 to 2. The result also corroborated that of Baker (2001) who reported that the difficulty values usually are in the range of -3 to +3.

Research question four stated that on the basis of criteria set for IRT framework which and how many of the items survived under IRT framework. The result showed that based on stated criteria for item response difficulty and discrimination index, 3 items were considered poor. Considering the item response difficulty and discrimination index 3 items were considered poor, these are item (7,11, 17). Therefore 27 items survived the item response framework criteria out of 30 items of Chemistry achievement test. IRT reduces excessive variance in test items through the Item information curse with respect to testees trait. This result partially corroborated previous research which showed that response latency can be useful in measuring both examinee effort in the test (Wise & Kong, 2005) and the effort received by items (Wise, 2006). Additionally, Wise and DeMars (2006) used an effort moderated IRT model based on response latency to consider examinee behavior and were able to better estimate both person and item parameter estimates.

VI. CONCLUSION

Based on the findings obtained from the study, it can be concluded, therefore, that the item of chemistry achievement test cannot be loaded on one factor despite all the efforts to do so. It can also be concluded that the items of chemistry achievement test survived under IRT framework, meaning that it fitted the three model parameter to certain extent. Also based on the findings, it can be concluded that the discrimination and difficulty parameter have a level of comparability.

VII. RECOMMENDATIONS

As a result, the following recommendations are proffered for improving the quality of tests used for educational and psychological researches, namely:

1. IRT should be used in test development processes as it enhances the validity and reliability of tests which provides quality assurance in measurement.
2. Using the three parameter model will greatly improve characterization of items, item selection, and in turn lead to improved measures which are the aim of test developers.
3. Test developers can bank on time-tested IRT item parameter methods to establish item difficulty and item discrimination characteristic of items.
4. Collaboration of government and non-governmental agencies to fund research and also provide a more conducive research enhancing supports and environments for further findings and finally researchers and scholars are called upon to corporatelycohabit research finding(s) with the western research so as to promote adaptation and acceptance across the world. In conclusion, the study has shown a positive significance to criterion related (predictive) validity when subjected to an exploratory factor analysis.

REFERENCES

[1] Ababio, O.Y. (2005). The New School Chemistry for Senior Secondary Schools, Africa Fast Publishers Ltd, Onitsha, Pp 2-5.

[2] Adams, D. (1993). Defining educational quality. IEQ Publication #1: Biennial Report.

[3] Adams, D., & Chapman, D. (2002). The quality of education: Dimensions and strategies. Education in Developing Asia. Vol. 5.

[4] Adedoyin, O.O. and Mokobi, T. (2013) .Using IRT psychometric analysis in examining the quality of junior certificate mathematics multiple choice examination test items. International Journal of Asian Social Science, 3(4): 992 -1011. View at Google Scholar.

[5] Anderson, L.W., Ryan, D.W. & Shapiro, B.J. (1989). The IEA classroom environment study. Oxford: Pergamon Press.

[6] Ani, E.N. (2014).Application of item response theory I the development and validation of multiple choice test in economics. (Master's Thesis). University of Nigeria, Nsukka.

[7] Anne, B. (2005). "Sociable Hyperlinks: An Ethnographic Approach to Connectivity." In: Hine, C. (ed). Virtual Methods: Issues in Social Research on the Internet. Oxford: Berg, 183-197.

[8] Asiyai,R.I.(2005)-Enhancing Chemistry Teaching in Secondary Schools through   Concept Mapping Instructional Strategy. In Nzewi, U.(Ed.)Proceedings of the 46th Annual Conference of Science Teachers

Association of Nigeria.205-209.Ibadan Nigeria: Heinemann Educational Books.

[9] Audinos, B. Lairez, T & Makwati, G. (2003). The role of statistics in improving the quality of classes: The case of Cyprus. Manuscript submitted for publication. Science Education International 13(2) 20-21.

[10] Ayot .H., (1986).Understanding Social Studies. Kenyatta University, Kenya and university of London institute of education.

[11] Baker, F. B. & Kim, S.-H. (2004). Item response theory: parameter estimation techniques.CRC Press.

[12] Baker, F. B. (2001) The Basics of Item Response Theory. USA: ERIC Clearinghouse on Assessment and Evaluation.

[13] Beck & Earl, M. (2002). Key issues in secondary education. Continuum: New York: MPG Book LTD, Cornwall.

[14] Bhagwan, D. (2005). Educational uses of information and communication. Encyclopaedia

[15] Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee'sability. Statistical theories of mental test scores.

[16] Black, P. (2002). Working inside the black box: Assessment for learning in the classroom. London: King"s College.

[17] Bomo, C. A. (2015) Application of Item Response Theory in the development of students' attitude towards Mathematics Tests .Unpublished dissertation, University of Port Harcourt.1-6.

[18] Brenner, K. (2004). Challenge your teaching. American Society for Cell Biology.Retrieved.January 15, 2011from www.nature.com/nsmb/journal/v/full.

[19] Georgiev, N., (2008) Item analysis of C, D and E series form Raven's Standard Progressive Matrices with Item Response Theory two-parameter logistic model Europe's Journal of Psychology (2008) http://www.ejop.org/archives/2008/08/item_analysis_o.html

[20] Haebera, T. (1980). Equating logistic ability scales by weighted least squares method. Japanese Psychological Research, 22(3), 144–149.

[21] Hambeleton, R.K & Jones, R.W (1993).Comparison of Classical Test Theory and Item response Theory and then Application to test development. Educational Measurement: Issues &Practices. 12(3), 38-47

[22] Hambleton R.K. Item response theory: A broad psychometric framework for measurement advances. Psicothema. 1994;6(3):535–56.

[23] Hawaii Department of Education.(1999). Assessment Terminology. Retrieved October 8,2018. Retrieved 2018-10-08.

[24] Hays, R.D., J.B. Bjorner, D.A. Revicki, K.L. Spritzer and D. Cella, 2009.Development of physical and mental health summary scores from the patient-reported outcomes measurementinformation system (PROMIS) global items. Quality of Life Research, 18(7): 873 –880. Viewat Google Scholar |View at Publisher.

[25] Hussain S., Ali R., Khan M.S., Ramzan M., &Qadeer M. Z. (2011). Attitude of secondary school teachers towards teaching profession. International Journal of Academic Research. 3(1), 985-990.

[26] Hutchinson, T. P. (1991). Ability, partial information and guessing: Statistical models applied to multiple-choice test .Rundle Mall, South Australia: Rumsby Scientific Publishing.

[27] Kenya Education Sector Support Programme 2005-2010: Delivering Quality Education and Training to All Kenyans, 25thJuly 2005: Nairobi; Pdf http:www.kessp@education.go.ke.

[28] Kpolovie, P. J. &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), 20-30.

[29] Kpolovie, P. J. (2011), Statistical Techniques for Advanced Research. New Owerri: Springfield Publishers Ltd.

[30] Kpolovie, P. J. (2010). Advanced Research Methods, New Owerri: Springfield Publishers Ltd.

[31] Kwale SMASSE Panel (2004). Gender issues in Mathematics and Science education. A paper presented in cycle 1 of Kwale district INSET- 2004.

[32] Kwale SMASSE Panel (2005).Resources utilisation for teaching/ learning of Mathematics and Science. A paper presented in cycle 2 of Kwale district INSET- 2005.

[33] Lecture note on Measurement and Evaluation. Retrieved 0ctober,11,2018.

[34] Lord, F.M. (1974). Estimation of latent ability and item parameters when there are omitted responses, Psychometrika, 39, 29-51

[35] Maris, E. (1995). Psychometric latent response models. Psychometrika, 60, 523–547.

[36] Nbina, J.B. (2012). Analysis of poor performance of senior secondary students in chemistry in Nigeria. An International Multidisciplinary Journal, vol. 6(4) pp. 324-334. ISNN 2070-0083.

[37] Ndukwu, P.N., School and teacher factors material resources utilization in pre-primary schools in Lagos State. Unpublished Ph.D Thesis. University of Ibadan, Ibadan, 2002. [39]

[38] Odinko, M. N. (2002). Home and school factors as determinants of literacy of skill development among Nigeria pre-primary school children. Unpublished P.hD thesis University of Ibadan, Ibadan.

[39] (17) (PDF)

[40] Ojerinde, D. (2013) Classical Test Theory VS Item Response Theory: An evaluation of the classroom and models for change? Connecting Research in Physics Education with Teacher Education.physics.ohio-state.edu//jossem/ICPE/TOC.html

[41] Olatunde, Y.P. (2009). Students attitude towards Mathematics and academic achievement in some selected secondary schools in Southwestern Nigeria. European Journal of Scientific Research, 36: 336-341.

[42] San Martín, E., delPino, G., & De Boeck, P. (2006). IRT models for ability-based guessing. Applied Psychological Measurement, 30, 183-203.https://doi.org/10.1177/0146621605282773

[43] Thorpe, G.L. and A. Favia, 2012. Data analysis using item response theory methodology: An introduction to selected programs and applications. Psychology Faculty Scholarship: 1-34.

[44] Twoli, W.N. (2006). Teaching secondary school Chemistry. A text book for teachers in developing countries. Nairobi: Nehema Publishers.

[45] Wingersky, M. S., Barton, M. A., & Lord, F. M. (1982).LOGIST User's Guide. Princeton NJ: Educational Testing Service.

[46] Wise, S.L. &Demars, C.E. (2006).An application of item response time: The effort-moderated IRT model. Journal of Educational Measurement, 43(1), 19-38. Doi:10.1111/j.1745-3984.2006.00002.x.

[47] Wise, S.L., &Kong .X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests performance. Applied Measurement in Education, 18(2), 163-183. Doi:10.1207/s15324818ame1802-2.

[48] Wise, S.L. &Sundre, D.L. (2003).Motivation filtering :An exploration of the impact of low examinee motivation on the psychometrics quality of tests. Paper presented to the National Council on in Education .Chicago ,II.

[49] Wise, V.L., Wise, S.L., & .Bhola, D.S. (2006). The generalizability of motivation, anxiety and test performance. Applied measurement in Education, 8(3),227-242.

AUTHORS

First Author – Taiwo Adebukola Kabir
ak.taiwo5@gmail.com