American Journal of Educational Research, 2018, Vol. 6, No. 3, 238-257
Available online at http://pubs.sciepub.com/education/6/3/11
DOI:10.12691/education-6-3-11

Development, Standardization and Application of Chemistry Achievement Test Using the One-Parameter Logistic Model (1-Plm) of Item Response Theory (Irt)

Oku Kingsley, Iweka Fidelis*

Department of Educational Psychology, Guidance & Counselling, Faculty of Education, University of Port Harcourt, Nigeria
*Corresponding author: iwekafidel@gmail.com

Abstract This study was centered on the Development, Standardization and Application of Chemistry Achievement Test using the One-Parameter Logistic Model of Item Response Theory. Eight research questions and one hypothesis guided the study. The hypothesis was tested at 0.05 level of significance. The research design was instrumentation research. The Researcher developed an instrument titled "OKUKINS Chemistry Achievement Test (OKUKINS CAT) consisting of 80 items. A proportional stratified random sampling technique was used to obtain a sample of 200 students from a population of 19,082 Senior Secondary School Two (SS II) students from Delta Central Senatorial District of Delta State. The sample was further divided into two equal parts (set A and set B) representing sample A and B respectively while the original sample represent sample O. The OKUKINS CAT yielded favourable statistics under the One-Parameter Logistic Model (1-PLM) with regards to the difficulty (b) parameters and ability estimates. Analyses done using the Xcalibre 4.2 software confirmed suitable indices of b parameters as prescribed by the Xcalibre manual. Using sample O, b parameters ranged from -2.417 to +2.834. Using sample A, b parameters ranged from -2.115 to +2.293. Using sample B, the b parameters ranged from -2.138 to +2.960. The validity of the instrument was estimated using the fit statistics. The results indicated perfect fit of One-Parameter Logistic Model (1-PLM). Using sample O, 74 items out of the 80 items of the OKUKINS CAT fitted the 1-PLM. Using sample A, 79 items out of the 80 items of the OKUKINS CAT fitted the 1-PLM. Using sample B, all the 80 items which constitute the OKUKINS CAT fitted the 1-PLM. The study also revealed that gender has no significant influence on academic achievement in chemistry. Based on the findings, it was recommended among others that, all examination bodies such as JAMB, WAEC, NECO and NABTEB should re-visit and review their test items using IRT models since this has been recommended by the International Testing Commission.

Keywords: *ability estimates, difficulty parameter, standard errors of measurement, item characteristic curves, item response theory, one-parameter logistic model*

Cite This Article: Oku Kingsley, and Iweka Fidelis, "Development, Standardization and Application of Chemistry Achievement Test Using the One-Parameter Logistic Model (1-Plm) of Item Response Theory (Irt)." *American Journal of Educational Research*, vol. 6, no. 3 (2018): 238-257. doi: 10.12691/education-6-3-11.

1. Introduction

Learning is a psychological construct and usually very difficult to tell or explain how it takes place or occurs. It can only be inferred from performance but performance is not learning. What this means is that, an examinee can perform very well in an examination but may not have learnt the learning materials most especially if he or she engaged in examination malpractice. In a different dimension, a test when administered may give inconsistent results if it contains misleading, ambiguous and series of defective items. In another dimension, an examinee can fail in examination even if he /she has mastered the various concepts that form the basis of the test. This may be due to the fact that, the examinee is not ready psychologically prior to writing the test. Thus, there is a negative interaction between the affective and cognitive traits [1].

It becomes evident that ability estimation of examinees with a test is a complicated task. Based on this, test constructors usually painstakingly undergo some rigorous processes to be certain that each item in a test is qualified to be incorporated in a test. This is achieved via item analysis. Item analysis is a statistical technique employed by test constructors to improve on the quality of their tests. According to Payne as cited in [2], item analysis is conducted for four general purposes as thus:

1. To select the best available item for the final version of a test.
2. To detect learning difficulties of the class as a whole; identifying general content area or skills that need to be scrutinized by the instructor.
3. To identify structural or content defects in the items and

4. To identify for individual students, area of weakness in need of remediation.

For instance, if an instrument is achievement test, it is expected that, when administered, it should produce consistent scores. These scores depict the performance of the examinees. Academic achievement according to [3] is the performance outcomes that depict the extent to which a person has achieved definite goals that were the focal point of activities in instructional environments, particularly in school, college and university. If an instrument is poorly constructed, when administered, it will produce defective results. Thus, test developers usually undergo test validation to ensure that the results that will be obtained from the administration of a measuring instrument can be used to draw unquestionable inference, interpretation and conclusive prediction of the future occurrence.

According to [4], "psychometrians are mainly concerned about the worth of test items as well as how examinees respond to them when constructing tests". Measurement experts generally use psychometric methods to find out the validity and reliability of test items. Psychometric theory offers two measurement frameworks in analyzing test data: Classical Test Theory (CTT) and Item Response Theory (IRT). Both theories guide test developers to forecast results of psychological tests by recognizing parameters of items difficulties and the ability of examinees.

Classical Test Theory (CTT) has been the basis for psychometric theory for decades. The conceptual basics, assumptions, and expansion of the basic premises of CTT 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" [5] 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.

IRT 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. According to [6], in IRT, the true component is defined on the latent variable of interest rather than on the test, as in the case in CTT. IRT relates item performance to the ability measured by the whole test. The ability is denoted by θ; IRT uses the ability test to ascertain the amount of latent ability an individual possesses. The ability test is an interval with a midpoint of zero and a standard deviation of one. It ranges from plus infinity to minus infinity, but practical consideration is usually from -3 to +3, in some cases from -4 to +4. The relationship between the probability of answering an item and the ability scale is called Item Response Function [7]. The higher the ability, the higher the probability of getting a right response or responding to a higher category. A trace line called Item Characteristic Curve (ICC) or the Category Response Curve (CRC) is used to illustrate the Item Response Function. The Item Characteristic Curve is the fundamental building block of Item Response Theory; all the other constructs of the theory rely on this curve [8]. The Item Characteristic Curve is describe by two technical properties. The first is the item difficulty. Under Item Response Theory, the item difficulty describes where the item functions along the ability scale. It is the point on the ability scale at which the probability of correct response to an item is 0.5. It is designated by b. For instance, an easy item functions among the low-ability testees and difficult items functions among the high ability testees: Thus, difficulty is a location index.

The second technical property is discrimination, which depicts how well an item can distinguish between examinees having abilities below and above the item location. This property basically shows the steepness of the ICC in its middle section. The item discriminates better as the curve become steeper. However, the item discriminates less as the ICC become flatter. This is because, the probability of answering an item correctly at low ability levels is almost equivalent as it is at high ability levels. Using these two descriptors, one can illustrate the general form of the ICC. These descriptors are also used to talk about the technical attributes of an item.

Item Response Theory Framework covers a group of models and the applicability of each model relies on the nature of the test item. Models could be unidimensional, models measuring one construct or multi-dimensional, models assessing more than one construct. Unidimensional models are either dichotomously scored (wrong / right, yes / no) or polytomously scored (e.g Likert Scale). The dichotomous models are used for cognitive tests or measures where there are right and wrong answers. The dichotomous models are expressed by the number of parameter they employ. One-Parameter Logistic Model (1-PLM) or the Rasch Model which employ the b(difficulty) parameter assumes that item discriminates equally and there is no guessing but varies with difficulty. The Two-Parameter Logistic Model (2-PLM) which employs the b (difficulty) and a (discrimination) parameters assume that items differ in difficulty and discrimination and there is no guessing. The Three-Parameter Logistic Model (3-PLM) took the guessing parameter, c into consideration. It assumes that the low ability examinee will answer a difficult item correctly by guessing. The other two models have the b value at the point where the probability of responding to an item correctly is 0.5, but with the Three-Parameter Model, the lower limit of the Item Characteristic Curve is the value c rather than zero. Hence, the b parameter is at the point on the ability scale where the probability of correct response to an item is $(I + c)/2$, while the slope of the Item Characteristic Curve at θ=b is actually a $(1 – c) /4$.

In recent times, attention has been drawn towards the tests used in measuring the presence or absence of identified trait of interest, considering the precision made by the instrument constructed. Several researchers have questioned the effectiveness of the Classical Test Theory technique anchoring on the following limitations among others:

a) CTT estimates of difficulty of item are group dependent.
b) The p and r coefficients also rely on the sample from which they are taken.

The ability scores of the examinees are completely test dependent. These limitations among others are now being complemented by the IRT technique. IRT provides more accurate estimates of ability, several improvements in scaling persons and items, greater flexibility in situations where different groups or test forms are used. Despite these merits, IRT has not been given sufficient coverage in graduate education. Worse still, most psychometricians and educators are not familiar with both the theoretical and practical aspects of IRT as very few empirical works have been done on it. In fact, IRT has not been given the needed attention as CTT in the development of measuring instruction such as chemistry achievement test.

Chemistry is one the science subjects taught in the Junior Secondary Schools together with Physics and Biology as Basic Science. It is taught as a single subject in the Senior Secondary Schools. Kirti (n.d) conceptualized Chemistry as the branch of science that deals with the recognition of the substances of which matter is composed, the examination of their attributes and the ways in which the matter interact, combine, change, and the use of these processes to form new substance. The importance of Chemistry to humanity cannot be over-emphasized. According to Smita as cited in [9], our all three basic needs i.e food, shelter and cloth are made by chemical processes using chemicals and fibres. That is, Chemistry is always present around us. So, Chemistry is important. It was based on this premise [10] concluded that "Chemistry is involved in the manufacture of all manmade objects and things. So, in your life time, everything you touch is tough by chemistry". Despite the importance of Chemistry in our everyday life, available records have showed poor performance among students in the subject in external examinations. [11] researched on students performance in external examinations in Boarding and Day Secondary Schools in Kano Metropolis, Nigeria using the examinations conducted from 2005 -2011 for selected subjects: English, Mathematics, Sciences (including Chemistry), and Home Economics. The results indicated poor performance among students in the selected subjects including Chemistry. According to the researcher, there was no year where the performance level of Day Senior Secondary Schools in Kano Metropolis reached 50% in any subject from 2005 -2011. Most of the measuring instruments used over the years were anchored on CTT. The credibility of these measuring instruments is yet to be extensively reviewed in the measurement community.

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. This study adopted the One-Parameter Logistic Model (1-PLM) of Item Response Theory in the Development and Standardization of Chemistry Achievement Test. The choice of One-Parameter Logistic Model (1-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 cognizance of the content areas.

1.1. Research Questions

The following research questions were answered in this study:

1) What is the item difficulty of each item of the OKUKINS Chemistry Achievement Test and their associated Standard Errors of Measurement?
2) What are the ability estimates of each examinee's and their associated Standard Errors of Measurement using the OKUKINS Chemistry Achievement test?
3) To what extent do the items of the OKUKINS Chemistry Achievement Test fit the One-Parameter Logistic Model (1-PLM)?
4) What are the correlation coefficients of the Standard Errors of Measurement across samples associated with difficulty parameters as an evidence of item-by-item reliability of the OKUKINS Chemistry Achievement Test?
5) What is the correlation coefficient of the Standard Errors of Measurement across samples associated with ability estimates as an evidence of person reliability of the OKUKINS Chemistry Achievement Test?
6) What are the Item Characteristic Curves (ICC) of the items of the OKUKINS Chemistry Achievement Test?
7) How much information does the OKUKINS Chemistry Achievement Test provide over the ability trait range?
8) What difference exists in the academic achievement of male and female students in the OKUKINS Chemistry Achievement Test?

1.2. Hypothesis

The following hypothesis was tested at 0.05 level of significance

1) There is no significant difference in the academic achievement of male and female students in the OKUKINS Chemistry Achievement Test.

2. Methodology

This study adopted the instrumentation research design since it deals with development and standardization of a measuring instrument. The population of the study is all the Nineteen Thousand and Eighty-Two (19,082) SS II students in the One Hundred and Eighty-five (185) secondary schools in the eight Local Government Areas in Delta Central Senatorial District of Delta State. Research evidence had shown that, in IRT, sample size has direct relationship with the precision in which item parameters can be estimated. According to [12], inadequate sample size can lead to increased estimation error with unfavourable implications for analysis of item and test data. Suen as cited in [2] recommended a sample size of 200 examinees for One-Parameter Logistic Model

(1-PLM). The present researcher used a sample size of 200 SS II students selected from 4 sampled Local Government Areas in Delta Central Senatorial District of Delta State using proportional stratified random sampling technique. The sample was further divided into two equal parts (set A and B) each consisting of 100 students. The original sample was designated as sample O. Set A and set B were designated as sample A and sample B respectively. Equal numbers of male and female students were selected for the study.

The instrument for this study was titled "OKUKINS Chemistry Achievement Test (OKUKINS CAT). The researcher used a table of specification to generate 100 items across the six cognitive domains: knowledge, comprehension, application, analysis, synthesis and evaluation. The content areas considered were:

- Introduction to Chemistry
- Particulate Nature of Matter
- Symbols, Formulae and Equations
- Chemical Combination
- Carbon and its Compounds
- Gas Laws
- Water
- Acids, Bases and Salts
- Standard Separation Techniques

The items generated were administered to 100 students who were not part of the sample used for the investigation. After administration, item analysis was performed on each item using Marginal Maximum Likelihood Estimation Method via Xcalibre 4.2 software anchoring on One-Parameter Logistic Model (I-PLM). The result of the item analysis showed that 99 items of the OKUKINS CAT fitted the One-Parameter Model (1-PLM). One item has F flag with z Residual statistically significant, i.e. it does not fit the model. Out of the 99 items that fitted the 1-PLM, 19 items had flags as outlined below:

a) One item has Lb flag which indicate that the b parameter was lower than the minimum acceptable value.
b) 18 items had K flags which indicate that, the items keyed alternative did not have the highest correlation with the total score. The remaining 80 items which did not have flags either for F, K, Lb or Hb were selected to form the OKUKINS Chemistry Achievement Test (OKUKINS CAT).

The validity of the instrument was established using face, content and fit statistics. In IRT, the internal validity of a test is assessed in terms of the statistical fit of each model. According to [13], if the fit statistic of an item is acceptable, the item is valid. During the item analysis, 99 items fitted the One-Parameter Logistic Model. Out of the 99 items that fitted the One-Parameter Logistic Model (1-PLM), 80 items were selected to form the OKUKINS CAT. Data analysis using three different samples (sample O, sample A and sample B) showed that, using sample O, 74 items out of the 80 items of the OKUKINS CAT fitted the 1-PLM; using sample A, 79 items out of the 80 items of the OKUKINS CAT fitted the 1-PLM and using sample B, all the 80 items of the OKUKINS CAT fitted the One-Parameter Logistic Model. Thus, the OKUKINS CAT is valid internally. The reliability of the instrument was estimated using correlation evidence of standard errors of measurement associated with difficulty parameters and ability estimates. Item-by-item reliability of the instrument was established through correlation evidence using the Standard Errors of Measurement associated with the difficulty parameters. Coefficients of 0.946, 0.721 and 0.604 were obtained. In the same dimension, person reliability of the instrument was estimated through correlation evidence using the Standard Errors of Measurement associated with the ability estimates. Coefficient of 0.595 was obtained.

The validated of OKUKINS Chemistry Achievement Test was administered to the subjects in the locale under investigation by the researcher and eight research assistants. The duration of the test was 1hour 30mins after which the answer scripts were retrieved from the examinees. Data obtained were analyzed using descriptive statistics of mean and standard deviation, independent sample t-test, pearson product moment correlation, factor analysis and marginal maximum likelihood Estimation Technique via Xcalibre 4.2 software anchoring on One-Parameter Logistic Model of IRT.

3. Results

After data analysis, the result obtained for research questions 1-8 and hypothesis 1 were summarized and presented in tables and figures as indicated below.

Table 1 shows the difficulty parameters of the 80 items which constitute the OKUKINS Chemistry Achievement Test and their associated Standard Errors of Measurement (SEM) using sample O. The b parameters ranged from -2.417 to +2.834. Item 54 has the lowest difficulty parameter of -2.417, very easy item while item 74 has the highest difficulty parameter of +2.834, very difficult item. Table 1 also showed the Standard Errors of Measurement (SEM) associated with the b parameters. The SEM of the items of the OKUKINS CAT ranged from 0.148 to 0.344 with mean and standard deviation of 0.167 and 0.030 respectively.

Table 2 indicates the b values of the 80 items which form the OKUKINS Chemistry Achievement Test and their associated Standard Errors of Measurement using sample A and sample B. As shown in Table 2, in sample A, the difficulty parameters of the items of the OKUKINS Chemistry Achievement Test ranged from -2.115 to +2.293. The Standard Errors of Measurement associated with the b parameters using sample A (SEM_A) ranged from 0.210 to 0.382 with mean of 0.235 and standard deviation of 0.037. Also presented in Table 2 for sample B, the difficulty parameters of the items of the OKUKINS CAT ranged from -2.138 to +2.960. The Standard Errors of Measurement associated with the b parameters using sample B (SEM_B) ranged from 0.208 to 0.558 with mean of 0.240 and standard deviation of 0. 049.

Table 3 shows the ability estimates of each examinee's and their associated Standard Errors of Measurement using sample O. The ability of the examinees ranged from -2.276 to +2.163. Also, the Standard Errors of Measurement associated with the ability estimates ranged from 0.2457 to 0.3128 with mean of 0.258 and standard deviation of 0.016. Classification of the score of each examinee's using sample O, revealed that 39 students representing 19.5% have high scores, that is, scores above the item location while 161 students representing 80.5%

have low scores, that is, scores below the item location. This implies that, the pass rate of the OKUKINS Chemistry Achievement Test is 19.5% using sample O anchoring on a cut score of 0.500.

Table 1. Difficulty parameter of each item of the OKUKINS Chemistry Achievement Test and their associated Standard Errors of Measurement using sample O

Item ID	b	SEM_O	Item ID	B	SEM_O	Item ID	B	SEM_O
1	-0.306	0.148	28	0.894	0.176	55	-1.597	0.167
2	-1.464	0.163	29	0.054	0.152	56	0.797	0.176
3	-1.379	0.160	30	0.961	0.179	57	-1.123	0.154
4	0.242	0.156	31	0.214	0.155	58	-0.128	0.150
5	-0.408	0.148	32	-0.941	0.151	59	-1.340	0.159
6	-1.099	0.154	33	1.208	0.191	60	-0.180	0.149
7	-0.128	0.150	34	1.135	0.188	61	1.568	0.214
8	-1.287	0.158	35	0.734	0.170	62	-0.077	0.150
9	1.759	0.228	36	0.241	0.156	63	-0.051	0.151
10	-0.281	0.148	37	-0.787	0.149	64	0.893	0.176
11	0.705	0.169	38	1.851	0.235	65	-0.433	0.148
12	0.409	0.160	39	1.171	0.190	66	-0.534	0.148
13	0.467	0.161	40	1.134	0.188	67	-1.203	0.156
14	0.798	0.172	41	0.134	0.153	68	-0.231	0.149
15	-0.256	0.149	42	-0.307	0.148	69	0.241	0.156
16	0.161	0.154	43	0.612	0.166	70	2.182	0.265
17	0.553	0.164	44	-0.660	0.148	71	-0.711	0.149
18	0.161	0.154	45	-0.889	0.150	72	-0.559	0.148
19	0.269	0.156	46	-0.458	0.148	73	-0.609	0.148
20	0.583	0.165	47	-0.915	0.151	74	2.834	0.344
21	0.269	0.156	48	-1.071	0.153	75	-0.509	0.148
22	0.380	0.159	49	-1.230	0.156	76	-1.718	0.172
23	1.615	0.217	50	-1.044	0.153	77	-1.908	0.180
24	0.409	0.160	51	-0.534	0.148	78	-0.154	0.149
25	-0.102	0.150	52	-0.787	0.149	79	0.702	0.169
26	0.437	0.160	53	1.568	0.214	80	1.282	0.196
27	0.187	0.154	54	-2.417	0.210		**Mean**	0.167
							SD	0.030

Table 2. Difficulty parameter of each item of the OKUKINS Chemistry Achievement Test and their associated Standard Errors of Measurement using sample A and Sample B

SAMPLE A						SAMPLE B					
Item ID	**B**	**SEM_A**	**Item ID**	**B**	**SEM_A**	**Item ID**	**b**	**SEM_B**	**Item ID**	**B**	**SEM_B**
1	-0.341	0.210	41	0.081	0.214	1	0.024	0.217	41	0.397	0.231
2	-1.366	0.228	42	-0.520	0.210	2	-1.122	0.213	42	-0.712	0.208
3	-1.473	0.234	43	0.801	0.240	3	-1.077	0.212	43	1.122	0.278
4	0.223	0.218	44	-0.565	0.210	4	0.172	0.222	44	-0.612	0.208
5	-0.476	0.210	45	-0.939	0.215	5	0.174	0.222	45	-0.662	0.208
6	-1.241	0.224	46	-0.153	0.211	6	-1.604	0.229	46	-0.153	0.213
7	0.130	0.215	47	-1.035	0.218	7	0.007	0.217	47	-0.863	0.209
8	-1.342	0.228	48	-0.987	0.217	8	-1.277	0.217	48	-0.662	0.208
9	1.791	0.318	49	-1.443	0.232	9	1.846	0.355	49	-1.169	0.214
10	-0.428	0.210	50	-0.987	0.217	10	-0.358	0.209	50	-1.015	0.211
11	0.637	0.232	51	-0.474	0.210	11	0.644	0.244	51	-0.813	0.209
12	0.530	0.228	52	-0.939	0.215	12	1.053	0.272	52	-1.015	0.211
13	0.376	0.222	53	1.605	0.299	13	0.644	0.244	53	1.121	0.277
14	0.747	0.237	54	-2.115	0.275	14	1.204	0.285	54	-2.138	0.260
15	-0.291	0.210	55	-1.498	0.235	15	0.115	0.220	55	-1.325	0.218
16	0.129	0.215	56	0.917	0.246	16	-0.100	0.214	56	0.974	0.266

SAMPLE A						SAMPLE B					
Item ID	**B**	**SEM_A**	**Item ID**	**B**	**SEM_A**	**Item ID**	**b**	**SEM_B**	**Item ID**	**B**	**SEM_B**
17	0.477	0.226	57	-1.285	0.226	17	0.006	0.217	57	-0.964	0.210
18	-0.106	0.211	58	-0.337	0.210	18	0.340	0.228	58	-0.612	0.208
19	0.082	0.214	59	-1.553	0.238	19	0.060	0.218	59	-1.378	0.220
20	0.582	0.230	60	0.081	0.214	20	0.838	0.256	60	0.005	0.217
21	0.129	0.215	61	1.688	0.307	21	0.282	0.226	61	0.974	0.266
22	0.375	0.222	62	-0.337	0.210	22	1.124	0.278	62	-0.359	0.209
23	1.781	0.317	63	0.274	0.219	23	1.449	0.308	63	0.339	0.228
24	0.325	0.221	64	0.744	0.237	24	0.579	0.240	64	0.836	0.256
25	-0.245	0.210	65	-0.291	0.210	25	-0.308	0.210	65	-1.378	0.220
26	0.129	0.215	66	-0.519	0.210	26	0.517	0.237	66	-0.153	0.213
27	-0.013	0.213	67	-1.285	0.226	27	-0.047	0.215	67	-0.863	0.209
28	0.801	0.240	68	-0.153	0.211	28	1.123	0.278	68	-0.257	0.211
29	0.374	0.222	69	0.177	0.217	29	-0.359	0.209	69	0.835	0.256
30	1.104	0.258	70	2.293	0.382	30	0.905	0.261	70	2.960	0.558
31	0.275	0.219	71	-0.704	0.211	31	0.397	0.231	71	-1.066	0.212
32	-0.891	0.214	72	-0.199	0.210	32	-1.066	0212	72	-1.486	0.224
33	1.778	0.316	73	-0.565	0.210	33	1.122	0.278	73	-0.913	0.210
34	0.801	0.240	74	2.289	0.381	34	1.537	0.318	74	-0.612	0.208
35	0.978	0.250	75	-0.519	0.210	35	1.048	0.271	75	-0.410	0.209
36	0.225	0.218	76	-1.667	0.244	36	0.225	0.224	76	-1.432	0.222
37	-0.658	0.211	77	-1.667	0.244	37	-0.763	0.208	77	-1.541	0.226
38	1.966	0.338	78	-0.337	0.210	38	1.830	0.353	78	-0.461	0.208
39	1.235	0.267	79	0.743	0.237	39	1.361	0.299	79	0.516	0.237
40	1.103	0.258	80	1.167	0.262	40	1.199	0.284	80	1.198	0.284
				Mean	**0.235**					**Mean**	**0.240**
				SD	**0.037**					**SD**	**0.049**

Table 3. Ability estimates of each examinee's and their associated Standard Errors of Measurement using sample O

Sequence ID	Score	Theta	SEM	Classification	Sequence ID	Score	Theta	SEM	Classification
1	64	1.6456	0.3017	High	101	69	2.1635	0.3461	High
2	60	1.308	0.2806	High	102	22	-1.1667	0.2708	Low
3	57	1.0815	0.2695	High	103	22	-1.1668	0.2708	Low
4	56	1.0096	0.2665	High	104	48	0.4792	0.2507	High
5	18	-1.4771	0.2874	Low	105	18	-1.4773	0.2874	Low
6	23	-1.0943	0.2676	Low	106	23	-1.0943	0.2676	Low
7	14	-1.8349	0.3128	Low	107	31	-0.5612	0.251	Low
8	41	0.0499	0.2459	High	108	32	-0.4987	0.2498	Low
9	45	0.2931	0.2477	High	109	30	-0.6246	0.2523	Low
10	32	-0.4986	0.2498	Low	110	21	-1.2411	0.2743	Low
11	32	-0.4986	0.2498	Low	111	38	-0.1313	0.2459	Low
12	46	0.3547	0.2485	High	112	33	-0.4365	0.2488	Low
13	52	0.7364	0.2569	High	113	35	-0.3135	0.2472	Low
14	56	1.0097	0.2665	High	114	29	-0.6886	0.2539	Low
15	38	-0.1313	0.2459	Low	115	31	-0.5613	0.251	Low
16	41	0.0499	0.2459	High	116	34	-0.3748	0.2479	Low
17	40	-0.0105	0.2457	High	117	34	-0.3748	0.2479	Low
18	42	0.1104	0.2461	High	118	25	-0.9542	0.262	Low
19	44	0.2319	0.247	High	119	26	-0.8863	0.2597	Low
20	48	0.4792	0.2507	High	120	20	-1.3172	0.2782	Low
21	41	0.0499	0.2459	High	121	26	-0.8862	0.2597	Low
22	40	-0.0105	0.2457	Low	122	18	-1.4771	0.2874	Low
23	39	-0.0709	0.2458	Low	123	24	-1.0235	0.2647	Low
24	39	-0.0709	0.2457	Low	124	10	-2.2762	0.3552	Low
25	42	0.1104	0.2461	High	125	29	-0.6887	0.2539	Low
26	43	0.171	0.2465	High	126	24	-1.0235	0.2647	Low
27	31	-0.5613	0.251	Low	127	23	-1.0943	0.2676	Low
28	36	-0.2526	0.2466	Low	128	25	-0.9543	0.262	Low
29	35	-0.3135	0.2472	Low	129	26	-0.8862	0.2597	Low

Sequence ID	Score	Theta	SEM	Classification	Sequence ID	Score	Theta	SEM	Classification
30	38	-0.1313	0.2459	Low	130	24	-1.0236	0.2647	Low
31	34	-0.3748	0.2479	Low	131	27	-0.8194	0.2575	Low
32	31	-0.5613	0.251	Low	132	28	-0.7535	0.2556	Low
33	33	-0.4365	0.2488	Low	133	39	-0.0709	0.2457	Low
34	32	-0.4986	0.2498	Low	134	21	-1.2409	0.2743	Low
35	32	-0.4986	0.2498	Low	135	19	-1.3959	0.2826	Low
36	35	-0.3135	0.2472	Low	136	23	-1.0943	0.2676	Low
37	35	-0.3135	0.2472	Low	137	22	-1.1667	0.2708	Low
38	31	-0.5613	0.251	Low	138	19	-1.3959	0.2826	Low
39	35	-0.3135	0.2472	Low	139	28	-0.7536	0.2556	Low
40	34	-0.3748	0.2479	Low	140	17	-1.561	0.2927	Low
41	37	-0.1919	0.2462	Low	141	32	-0.4986	0.2498	Low
42	35	-0.3135	0.2472	Low	142	32	-0.4986	0.2498	Low
43	36	-0.2526	0.2466	Low	143	22	-1.1667	0.2708	Low
44	35	-0.3135	0.2472	Low	144	38	-0.1313	0.2459	Low
45	38	-0.1313	0.2459	Low	145	23	-1.0942	0.2675	Low
46	32	-0.4986	0.2498	Low	146	28	-0.7535	0.2556	Low
47	35	-0.3135	0.2472	Low	147	49	0.5424	0.2519	High
48	27	-0.8193	0.2575	Low	148	35	-0.3135	0.2472	Low
49	24	-1.0235	0.2646	Low	149	35	-0.3135	0.2472	Low
50	25	-0.9541	0.262	Low	150	37	-0.1918	0.2462	Low
51	65	1.7386	0.3085	High	151	33	-0.4365	0.2488	Low
52	65	1.7386	0.3085	High	152	48	0.4792	0.2507	High
53	56	1.0097	0.2665	High	153	46	0.3547	0.2485	High
54	38	-0.1313	0.2459	Low	154	42	0.1104	0.2461	High
55	38	-0.1314	0.2459	Low	155	32	-0.4986	0.2498	Low
56	41	0.0499	0.2459	High	156	30	-0.6247	0.2524	Low
57	41	0.0499	0.2459	High	157	28	-0.7535	0.2556	Low
58	58	1.155	0.2729	High	158	31	-0.5612	0.251	Low
59	52	0.7363	0.2569	High	159	36	-0.2525	0.2466	Low
60	39	-0.0709	0.2457	Low	160	32	-0.4986	0.2498	Low
61	31	-0.5612	0.251	Low	161	31	-0.5612	0.251	Low
62	43	0.171	0.2465	High	162	32	-0.4986	0.2498	Low
63	40	-0.0105	0.2457	Low	163	31	-0.5612	0.251	Low
64	35	-0.3136	0.2472	Low	164	23	-1.0943	0.2676	Low
65	34	-0.3748	0.2479	Low	165	37	-0.1918	0.2462	Low
66	30	-0.6246	0.2523	Low	166	23	-1.0942	0.2675	Low
67	34	-0.3748	0.2479	Low	167	28	-0.7535	0.2556	Low
68	23	-1.0942	0.2675	Low	168	47	0.4167	0.2495	High
69	35	-0.3135	0.2472	Low	169	37	-0.1919	0.2462	Low
70	31	-0.5612	0.251	Low	170	30	-0.6245	0.2523	Low
71	37	-0.1918	0.2462	Low	171	35	-0.3135	0.2472	Low
72	39	-0.0709	0.2458	Low	172	34	-0.3748	0.2479	Low
73	34	-0.3748	0.2479	Low	173	34	-0.3748	0.2479	Low
74	42	0.1104	0.2461	High	174	31	-0.5613	0.251	Low
75	31	-0.5613	0.251	Low	175	38	-0.1313	0.2459	Low
76	38	-0.1313	0.2459	Low	176	33	-0.4364	0.2488	Low
77	30	-0.6245	0.2523	Low	177	31	-0.5612	0.251	Low
78	26	-0.8863	0.2597	Low	178	43	0.1711	0.2465	High
79	25	-0.9544	0.2621	Low	179	43	0.1711	0.2465	High
80	24	-1.0234	0.2646	Low	180	38	-0.1313	0.2459	Low
81	32	-0.4986	0.2498	Low	181	62	1.4707	0.2901	High
82	31	-0.5612	0.251	Low	182	26	-0.8862	0.2597	Low
83	23	-1.0943	0.2676	Low	183	18	-1.477	0.2873	Low
84	37	-0.1918	0.2462	Low	184	34	-0.3747	0.2479	Low
85	23	-1.0942	0.2675	Low	185	33	-0.4364	0.2488	Low
86	28	-0.7535	0.2556	Low	186	35	-0.3135	0.2472	Low
87	47	0.4167	0.2495	High	187	30	-0.6246	0.2523	Low
88	29	-0.6887	0.2539	Low	188	32	-0.4985	0.2498	Low

Sequence ID	Score	Theta	SEM	Classification	Sequence ID	Score	Theta	SEM	Classification
89	25	-0.9542	0.262	Low	189	35	-0.3135	0.2472	Low
90	23	-1.0942	0.2675	Low	190	34	-0.3748	0.2479	Low
91	23	-1.0943	0.2676	Low	191	30	-0.6245	0.2523	Low
92	23	-1.0943	0.2676	Low	192	38	-0.1313	0.2459	Low
93	29	-0.6886	0.2539	Low	193	23	-1.0944	0.2676	Low
94	23	-1.0942	0.2675	Low	194	27	-0.8192	0.2575	Low
95	22	-1.1667	0.2708	Low	195	21	-1.2409	0.2743	Low
96	29	-0.6887	0.2539	Low	196	19	-1.3959	0.2826	Low
97	25	-0.9542	0.262	Low	197	23	-1.0943	0.2676	Low
98	23	-1.0944	0.2676	Low	198	22	-1.1667	0.2708	Low
99	42	0.1104	0.2461	High	199	19	-1.3959	0.2826	Low
100	44	0.2319	0.247	High	200	28	-0.7536	0.2556	Low
							Mean	**0.258**	
							SD	**0.016**	

Table 4. Ability estimates of each examinee's and their associated Standard Errors of Measurement using Sample A and Sample B

SAMPLE A					SAMPLE B				
Sequence ID	**Score_A**	**Theta_A**	**SEM_A**	**Classification**	**Sequence ID**	**Score_B**	**Theta_B**	**SEM_B**	**Classification**
1	64	1.6516	0.3023	High	1	70	2.29	0.3582	High
2	60	1.3124	0.2814	High	2	22	-1.1752	0.2708	Low
3	57	1.0844	0.2704	High	3	22	-1.1752	0.2708	Low
4	56	1.0121	0.2673	High	4	49	0.5462	0.2529	High
5	18	-1.4818	0.287	Low	5	17	-1.5686	0.292	Low
6	23	-1.0996	0.2675	Low	6	24	-1.0319	0.2649	Low
7	15	-1.7435	0.3047	Low	7	32	-0.5044	0.2507	Low
8	41	0.0469	0.2464	High	8	31	-0.5676	0.2519	Low
9	45	0.2912	0.2484	High	9	30	-0.6313	0.2531	Low
10	32	-0.5034	0.2501	Low	10	22	-1.1753	0.2708	Low
11	32	-0.5035	0.2501	Low	11	38	-0.1339	0.2471	Low
12	46	0.3531	0.2492	High	12	33	-0.4418	0.2498	Low
13	52	0.7371	0.2577	High	13	35	-0.3178	0.2483	Low
14	56	1.0121	0.2673	High	14	30	-0.6313	0.2531	Low
15	38	-0.1351	0.2464	Low	15	32	-0.5045	0.2508	Low
16	41	0.0469	0.2464	High	16	35	-0.3178	0.2483	Low
17	40	-0.0138	0.2462	High	17	34	-0.3796	0.249	Low
18	42	0.1076	0.2467	High	18	26	-0.8942	0.2601	Low
19	44	0.2297	0.2477	High	19	25	-0.9624	0.2624	Low
20	48	0.4784	0.2514	High	20	21	-1.2494	0.2742	Low
21	41	0.0469	0.2464	High	21	26	-0.8942	0.2601	Low
22	40	-0.0138	0.2463	Low	22	18	-1.4849	0.2868	Low
23	39	-0.0744	0.2462	Low	23	24	-1.0319	0.2649	Low
24	39	-0.0744	0.2462	Low	24	10	-2.2783	0.3535	Low
25	42	0.1076	0.2467	High	25	28	-0.761	0.2562	Low
26	43	0.1686	0.2471	High	26	25	-0.9624	0.2624	Low
27	31	-0.5662	0.2512	Low	27	23	-1.1028	0.2677	Low
28	36	-0.2568	0.247	Low	28	25	-0.9624	0.2624	Low
29	35	-0.3179	0.2476	Low	29	26	-0.8942	0.2601	Low
30	38	-0.1351	0.2464	Low	30	25	-0.9624	0.2624	Low
31	26	-0.8914	0.2597	Low	31	26	-0.8941	0.2601	Low
32	28	-0.7588	0.2557	Low	32	28	-0.761	0.2562	Low
33	39	-0.0744	0.2462	Low	33	39	-0.0728	0.247	Low
34	21	-1.2461	0.2741	Low	34	22	-1.1751	0.2707	Low
35	19	-1.4008	0.2822	Low	35	19	-1.404	0.2822	Low
36	23	-1.0996	0.2675	Low	36	23	-1.1028	0.2677	Low
37	22	-1.172	0.2706	Low	37	22	-1.1752	0.2708	Low
38	20	-1.3225	0.278	Low	38	19	-1.404	0.2822	Low
39	28	-0.7589	0.2557	Low	39	28	-0.761	0.2562	Low
40	17	-1.5655	0.2922	Low	40	17	-1.5686	0.292	Low

SAMPLE A					SAMPLE B				
Sequence ID	**Score_A**	**Theta_A**	**SEM_A**	**Classification**	**Sequence ID**	**Score_B**	**Theta_B**	**SEM_B**	**Classification**
41	32	-0.5034	0.2501	Low	41	33	-0.4418	0.2498	Low
42	32	-0.5034	0.2501	Low	42	32	-0.5044	0.2507	Low
43	22	-1.172	0.2706	Low	43	22	-1.1752	0.2708	Low
44	38	-0.1351	0.2464	Low	44	38	-0.1338	0.2471	Low
45	24	-1.0289	0.2646	Low	45	23	-1.1027	0.2677	Low
46	28	-0.7588	0.2557	Low	46	28	-0.761	0.2562	Low
47	49	0.5419	0.2527	High	47	50	0.6105	0.2543	High
48	35	-0.3179	0.2476	Low	48	36	-0.2563	0.2478	Low
49	35	-0.3179	0.2476	Low	49	36	-0.2563	0.2478	Low
50	37	-0.1958	0.2466	Low	50	38	-0.1339	0.2471	Low
51	33	-0.4412	0.2491	Low	51	34	-0.3796	0.249	Low
52	48	0.4784	0.2514	High	52	49	0.5462	0.2529	High
53	46	0.3531	0.2492	High	53	47	0.4195	0.2506	High
54	42	0.1077	0.2467	High	54	43	0.1716	0.2477	High
55	32	-0.5034	0.2501	Low	55	33	-0.4418	0.2498	Low
56	29	-0.6939	0.254	Low	56	29	-0.6958	0.2546	Low
57	29	-0.6938	0.254	Low	57	28	-0.7609	0.2562	Low
58	31	-0.5662	0.2512	Low	58	31	-0.5675	0.2519	Low
59	36	-0.2567	0.247	Low	59	37	-0.1949	0.2474	Low
60	32	-0.5034	0.2501	Low	60	33	-0.4418	0.2498	Low
61	31	-0.5662	0.2512	Low	61	32	-0.5043	0.2507	Low
62	32	-0.5034	0.2501	Low	62	33	-0.4418	0.2498	Low
63	31	-0.5662	0.2512	Low	63	31	-0.5675	0.2519	Low
64	23	-1.0996	0.2675	Low	64	23	-1.1028	0.2677	Low
65	37	-0.1958	0.2466	Low	65	37	-0.195	0.2474	Low
66	24	-1.0289	0.2646	Low	66	23	-1.1027	0.2677	Low
67	28	-0.7588	0.2557	Low	67	28	-0.761	0.2562	Low
68	47	0.4155	0.2502	High	68	48	0.4826	0.2517	High
69	37	-0.1959	0.2466	Low	69	38	-0.1339	0.2471	Low
70	30	-0.6296	0.2525	Low	70	31	-0.5675	0.2519	Low
71	35	-0.3178	0.2476	Low	71	36	-0.2562	0.2478	Low
72	34	-0.3793	0.2482	Low	72	34	-0.3796	0.249	Low
73	34	-0.3793	0.2482	Low	73	35	-0.3178	0.2483	Low
74	31	-0.5662	0.2512	Low	74	32	-0.5044	0.2507	Low
75	38	-0.1351	0.2464	Low	75	39	-0.0728	0.247	Low
76	33	-0.4411	0.2491	Low	76	34	-0.3795	0.249	Low
77	31	-0.5662	0.2512	Low	77	32	-0.5043	0.2507	Low
78	43	0.1686	0.2471	High	78	44	0.2331	0.2482	High
79	43	0.1686	0.2471	High	79	43	0.1716	0.2477	High
80	38	-0.1351	0.2464	Low	80	39	-0.0728	0.247	Low
81	62	1.4759	0.2908	High	81	63	1.5625	0.2951	High
82	25	-0.9594	0.262	Low	82	25	-0.9623	0.2623	Low
83	19	-1.4009	0.2823	Low	83	18	-1.4848	0.2868	Low
84	34	-0.3793	0.2482	Low	84	35	-0.3177	0.2483	Low
85	33	-0.4411	0.2491	Low	85	34	-0.3795	0.249	Low
86	35	-0.3179	0.2476	Low	86	36	-0.2562	0.2478	Low
87	30	-0.6296	0.2525	Low	87	31	-0.5675	0.2519	Low
88	32	-0.5034	0.2501	Low	88	33	-0.4417	0.2498	Low
89	35	-0.3178	0.2476	Low	89	35	-0.3177	0.2483	Low
90	34	-0.3793	0.2482	Low	90	35	-0.3178	0.2483	Low
91	30	-0.6296	0.2525	Low	91	31	-0.5675	0.2519	Low
92	38	-0.135	0.2464	Low	92	39	-0.0728	0.247	Low
93	22	-1.1718	0.2706	Low	93	22	-1.1751	0.2707	Low
94	27	-0.8245	0.2576	Low	94	28	-0.7609	0.2562	Low
95	21	-1.2461	0.2741	Low	95	22	-1.1751	0.2707	Low
96	19	-1.4008	0.2822	Low	96	19	-1.404	0.2822	Low
97	23	-1.0996	0.2675	Low	97	23	-1.1028	0.2677	Low

SAMPLE A					SAMPLE B				
Sequence ID	Score_A	Theta_A	SEM_A	Classification	Sequence ID	Score_B	Theta_B	SEM_B	Classification
98	22	-1.172	0.2706	Low	98	22	-1.1752	0.2708	Low
99	20	-1.3225	0.278	Low	99	19	-1.404	0.2822	Low
100	28	-0.7589	0.2557	Low	100	28	-0.761	0.2562	Low
		Mean	0.257				**Mean**	**0.260**	
		SD	0.014				**SD**	**0.018**	

Table 4 depicts the ability estimates of each examinee's and their associated Standard Errors of Measurement using the OKUKINS CAT for Sample A and Sample B. For sample A, ability estimate of each examinee's ranged from -1.743 to +1.652. The Standard Errors of Measurement for sample A associated with the ability estimates (SEM_Atheta) ranged from 0.246 to 0.3047 with mean and standard deviation of 0.257 and 0.014 respectively. For Sample B, ability estimates of each examinee's ranged from -2.278 to +2.290. Also, the Standard Errors of Measurement for Sample B associated with the ability estimates (SEM_Btheta) ranged from 0.247 to 0.3582 with mean of 0.260 and standard deviation of 0.018. Classification of the examinees scores using Sample A showed that, 25 students representing 25% have high scores, that is, scores above the item location while 75 students representing 75% have low scores, that is, scores below the item location. This means that, the pass rate of the OKUKINS Chemistry Achievement Test is 25.0% using Sample A anchoring on a cut score of 0.500. In the same dimension, classification of the examinees scores using Sample B, revealed that, 10 students representing 10% have high scores, that is, scores above the item location while 90 students representing 90% have low scores, that is, scores below the item location. This implies that, the pass rate of the OKUKINS chemistry Achievement Test (OKUKINS CAT) is 10.0% using Sample B anchoring on cut score of 0.500.

Table 5. Item parameter calibration using sample O

Item ID	P	R	a	b	Z Resid	P	Flag(s)
1	0.465	0.376	1.000	-0.306	1.080	0.280	
2	0.685	0.417	1.000	-1.464	1.602	0.109	
3	0.670	0.355	1.000	-1.379	1.049	0.294	
4	0.360	0.129	1.000	0.242	1.660	0.097	
5	0.485	0.320	1.000	-0.408	0.528	0.598	
6	0.620	0.021	1.000	-1.099	1.838	0.066	K
7	0.430	0.262	1.000	-0.128	0.961	0.337	
8	0.655	0.177	1.000	-1.287	1.035	0.301	
9	0.140	0.245	1.000	1.759	1.486	0.137	
10	0.460	0.169	1.000	-0.281	0.735	0.463	
11	0.280	0.084	1.000	0.705	2.119	0.034	F
12	0.330	0.255	1.000	0.409	1.108	0.268	
13	0.320	0.365	1.000	0.467	0.484	0.628	
14	0.265	0.359	1.000	0.798	0.960	0.337	
15	0.455	0.351	1.000	-0.256	0.490	0.624	
16	0.375	0.277	1.000	0.161	0.626	0.531	
17	0.305	-0.063	1.000	0.553	2.410	0.016	K, F
18	0.375	0.390	1.000	0.161	0.487	0.626	
19	0.355	0.333	1.000	0.269	0.755	0.450	
20	0.300	0.282	1.000	0.583	0.677	0.499	
21	0.355	0.194	1.000	0.269	0.735	0.462	
22	0.335	0.225	1.000	0.380	0.912	0.362	
23	0.155	0.104	1.000	1.615	1.417	0.156	K
24	0.330	0.352	1.000	0.409	0.379	0.705	
25	0.425	0.236	1.000	-0.102	0.936	0.349	
26	0.325	0.329	1.000	0.437	0.679	0.497	
27	0.370	0.169	1.000	0.187	0.827	0.408	
28	0.250	0.238	1.000	0.894	0.930	0.352	
29	0.395	0.016	1.000	0.054	1.153	0.249	K
30	0.240	0.119	1.000	0.961	1.727	0.084	
31	0.365	0.259	1.000	0.214	0.861	0.389	
32	0.590	0.168	1.000	-0.941	0.814	0.416	
33	0.205	0.157	1.000	1.208	0.968	0.333	
34	0.215	0.219	1.000	1.135	0.912	0.362	
35	0.275	0.231	1.000	0.734	0.611	0.541	
36	0.360	0.095	1.000	0.241	1.624	0.104	

Item ID	P	R	a	b	Z Resid	P	Flag(s)
37	0.560	0.333	1.000	-0.787	0.697	0.486	
38	0.130	0.064	1.000	1.851	1.656	0.098	
39	0.210	0.124	1.000	1.171	0.876	0.381	
40	0.215	0.205	1.000	1.134	1.153	0.249	
41	0.380	0.262	1.000	0.134	0.845	0.398	
42	0.465	0.269	1.000	-0.307	1.067	0.286	
43	0.295	0.185	1.000	0.612	0.693	0.488	
44	0.535	0.348	1.000	-0.660	0.634	0.526	
45	0.580	0.340	1.000	-0.889	0.575	0.565	
46	0.495	0.376	1.000	-0.458	0.535	0.593	
47	0.585	0.386	1.000	-0.915	1.096	0.273	
48	0.615	0.314	1.000	-1.071	0.704	0.481	
49	0.645	0.420	1.000	-1.230	1.267	0.205	
50	0.610	0.306	1.000	-1.044	0.873	0.383	
51	0.510	0.058	1.000	-0.534	1.479	0.139	
52	0.560	0.383	1.000	-0.787	0.722	0.470	
53	0.160	-0.097	1.000	1.568	2.056	0.040	K, F
54	0.830	0.343	1.000	-2.417	1.276	0.202	
55	0.710	0.330	1.000	-1.597	0.974	0.330	
56	0.265	0.161	1.000	0.797	0.765	0.444	
57	0.625	0.408	1.000	-1.123	1.155	0.248	
58	0.430	0.276	1.000	-0.128	0.620	0.535	
59	0.665	0.224	1.000	-1.340	0.902	0.367	
60	0.440	0.287	1.000	-0.180	0.706	0.480	
61	0.160	-0.169	1.000	1.568	2.340	0.019	K, F
62	0.420	0.025	1.000	-0.077	1.504	0.133	
63	0.415	0.275	1.000	-0.051	0.589	0.556	
64	0.250	0.240	1.000	0.893	0.843	0.399	
65	0.490	0.034	1.000	-0.433	1.803	0.071	K
66	0.510	0.393	1.000	-0.534	1.076	0.282	
67	0.640	0.357	1.000	-1.203	0.991	0.321	
68	0.450	0.348	1.000	-0.231	0.686	0.493	
69	0.360	0.335	1.000	0.241	0.586	0.558	
70	0.100	0.313	1.000	2.182	0.793	0.428	
71	0.545	0.261	1.000	-0.711	0.886	0.376	
72	0.515	0.031	1.000	-0.559	1.415	0.157	
73	0.525	0.196	1.000	-0.609	1.033	0.302	
74	0.055	-0.186	1.000	2.834	2.484	0.013	K, F
75	0.505	0.407	1.000	-0.509	1.249	0.212	
76	0.730	0.299	1.000	-1.718	1.248	0.212	
77	0.760	0.395	1.000	-1.908	1.811	0.070	
78	0.435	0.196	1.000	-0.154	0.895	0.371	
79	0.280	-0.153	1.000	0.702	2.585	0.010	K, F
80	0.195	0.140	1.000	1.282	1.004	0.316	

Table 5 shows the item parameters for all calibrated items depicting the fit statistic of each item using the Z Residual. Four (4) items (item 6, 23, 29 and 65) have K flags, which indicate that the items keyed alternatives did not have the highest correlation with the total score. One item (item 11) had F flag which shows that the fit statistic was significant, thus, item 11 did not fit the One-Parameter Logistic Model (Z Resid = 2.119, P = 0.034). Five (5) items (item 17, 53, 61, 74 and 79) had both K and F flags. This means that, the items keyed alternatives did not have the highest correlation with the total score. Also, the items fit statistics were statistically significant, thus, the items did not fit the One-Parameter Logistic Model (item 17: Z Resid = 2.410, P = 0.016); item 53: Z Resid = 2.056, P =0.040; item 61: z Resid = 2.340, P =0.019; item 74: Z Resid = 2.484, P =0.013; item 79: Z Resid = 2.585, P =0.010). On the whole, using sample O, 74 items of the OKUKINS Chemistry Achievement Test (OKUKINS CAT) fitted the One- Parameter Logistic Model.

Table 6 depicts the item parameters for all calibrated item showing the fit statistic of each item using the Z Residual. Fifteen (15) items have K flags which indicate that the items keyed alternative did not have that highest correlation with the total score (item 4, 6, 17, 29, 30, 33, 36, 38, 51, 53, 61, 62, 65, 72 and 79). One (1) item (item 74) had K and F flag. This implies that, item 74 keyed alternative did not have the highest correlation with the total score. Also, the item fit statistic is statistically significant (Z Resid =2.481, P = 0.000), thus, item 74 did not fit the One -Parameter Logistic Model using sample A.

The result shows that, 79 items out of the 80 items of the OKUKINS CAT fitted the One-Parameter Logistic Model using sample A.

Table 7 shows the item parameters for all calibrated items indicating the fit statistic of each item using the Z Residual. Fifteen (15) items had K flags (item 6, 11, 17, 23, 33, 35, 36, 38, 51, 53, 60, 61, 62, 69 and 79). This implies that, the items keyed alternative did not have the highest correlation with the total score. None of the items were flagged either for F, Lb or Hb, thus, all the 80 items of the OKUKINS Chemistry Achievement Test fitted the One-Parameter Logistic Model using sample B.

A cursory examination of the results in Table 5, Table 6 and Table 7 reveals that the items of the OKUKINS ACT fitted the I- PLM more perfectly using a sample size of 100 examinees than when a sample size of 200 examinees was used.

Table 6. Item parameter calibration using sample A

Item ID	P	R	a	b	Z Resid	P	Flag(s)
1	0.490	0.425	1.000	-0.341	0.638	0.523	
2	0.700	0.424	1.000	-1.366	1.214	0.225	
3	0.720	0.278	1.000	-1.473	0.638	0.523	
4	0.370	0.033	1.000	0.223	1.443	0.149	K
5	0.520	0.291	1.000	-0.476	0.583	0.560	
6	0.680	-0.088	1.000	-1.241	1.601	0.109	K
7	0.390	0.401	1.000	0.130	0.601	0.548	
8	0.700	0.248	1.000	-1.342	0.840	0.401	
9	0.120	0.252	1.000	1.791	1.145	0.252	
10	0.510	0.025	1.000	-0.428	1.183	0.237	
11	0.290	0.271	1.000	0.637	1.207	0.227	
12	0.310	0.338	1.000	0.530	0.464	0.643	
13	0.340	0.364	1.000	0.376	0.520	0.603	
14	0.270	0.355	1.000	0.747	0.967	0.334	
15	0.480	0.294	1.000	-0.291	0.325	0.745	
16	0.390	0.318	1.000	0.129	0.587	0.557	
17	0.320	-0.134	1.000	0.477	1.593	0.111	K
18	0.440	0.391	1.000	-0.106	0.452	0.651	
19	0.400	0.277	1.000	0.082	0.676	0.499	
20	0.300	0.325	1.000	0.582	0.399	0.690	
21	0.390	0.281	1.000	0.129	0.489	0.625	
22	0.340	0.396	1.000	0.375	0.860	0.390	
23	0.120	0.088	1.000	1.781	1.119	0.263	
24	0.350	0.365	1.000	0.325	0.467	0.641	
25	0.470	0.341	1.000	-0.245	0.526	0.599	
26	0.390	0.362	1.000	0.129	0.720	0.472	
27	0.420	0.242	1.000	-0.013	0.970	0.332	
28	0.260	0.345	1.000	0.801	0.710	0.478	
29	0.340	-0.036	1.000	0.374	1.064	0.287	K
30	0.210	0.062	1.000	1.104	1.494	0.135	K
31	0.360	0.256	1.000	0.275	0.778	0.437	
32	0.610	0.142	1.000	-0.891	0.508	0.611	
33	0.120	-0.048	1.000	1.778	1.221	0.222	K
34	0.260	0.249	1.000	0.801	0.508	0.612	
35	0.230	0.344	1.000	0.978	1.136	0.256	
36	0.370	0.003	1.000	0.225	1.530	0.126	K
37	0.560	0.358	1.000	-0.658	0.650	0.515	
38	0.100	-0.005	1.000	1.966	0.952	0.341	K
39	0.190	0.218	1.000	1.235	0.395	0.693	
40	0.210	0.274	1.000	1.103	0.518	0.605	
41	0.400	0.211	1.000	0.081	0.823	0.411	
42	0.530	0.359	1.000	-0.520	0.829	0.407	
43	0.260	0.334	1.000	0.801	0.909	0.364	
44	0.540	0.451	1.000	-0.565	0.805	0.421	
45	0.620	0.311	1.000	-0.939	0.229	0.819	
46	0.450	0.338	1.000	-0.153	0.311	0.756	
47	0.640	0.469	1.000	-1.035	0.943	0.346	
48	0.630	0.340	1.000	-0.987	0.598	0.550	
49	0.720	0.481	1.000	-1.443	1.335	0.182	
50	0.630	0.371	1.000	-0.987	0.883	0.377	
51	0.520	-0.083	1.000	-0.474	1.846	0.065	K
52	0.620	0.388	1.000	-0.939	0.537	0.591	

Item ID	P	R	a	b	Z Resid	P	Flag(s)
53	0.140	-0.026	1.000	1.605	0.895	0.371	K
54	0.830	0.372	1.000	-2.115	0.657	0.511	
55	0.730	0.397	1.000	-1.498	1.112	0.266	
56	0.240	0.288	1.000	0.917	0.386	0.700	
57	0.690	0.458	1.000	-1.285	0.870	0.384	
58	0.490	0.230	1.000	-0.337	1.301	0.193	
59	0.740	0.212	1.000	-1.553	0.633	0.526	
60	0.400	0.365	1.000	0.081	0.986	0.324	
61	0.130	-0.214	1.000	1.688	1.644	0.100	K
62	0.490	-0.079	1.000	-0.337	1.692	0.091	K
63	0.360	0.281	1.000	0.274	0.557	0.577	
64	0.270	0.157	1.000	0.744	0.872	0.383	
65	0.480	-0.032	1.000	-0.291	1.338	0.181	K
66	0.530	0.317	1.000	-0.519	0.808	0.419	
67	0.690	0.460	1.000	-1.285	1.211	0.226	
68	0.450	0.265	1.000	-0.153	0.728	0.466	
69	0.380	0.354	1.000	0.177	0.909	0.363	
70	0.070	0.220	1.000	2.293	0.451	0.652	
71	0.570	0.279	1.000	-0.704	0.742	0.458	
72	0.460	-0.013	1.000	-0.199	1.332	0.183	K
73	0.540	0.061	1.000	-0.565	1.821	0.069	
74	0.070	-0.365	1.000	2.289	2.481	0.013	K, F
75	0.530	0.410	1.000	-0.519	0.791	0.429	
76	0.760	0.181	1.000	-1.667	0.587	0.557	
77	0.760	0.463	1.000	-1.667	1.511	0.131	
78	0.490	0.161	1.000	-0.337	0.630	0.529	
79	0.270	-0.156	1.000	0.743	1.512	0.131	K
80	0.200	0.256	1.000	1.167	0.336	0.737	

Table 7. Item parameter calibration using sample B

Item ID	P	R	a	b	Z Resid	P	Flag(s)
1	0.370	0.322	1.000	0.024	0.535	0.593	
2	0.590	0.439	1.000	-1.122	1.118	0.264	
3	0.580	0.373	1.000	-1.077	0.956	0.339	
4	0.340	0.110	1.000	0.172	1.308	0.191	
5	0.340	0.161	1.000	0.174	0.643	0.520	
6	0.680	0.015	1.000	-1.604	1.183	0.237	K
7	0.370	0.168	1.000	0.007	0.913	0.361	
8	0.620	0.248	1.000	-1.277	0.569	0.569	
9	0.110	0.169	1.000	1.846	1.332	0.183	
10	0.440	0.168	1.000	-0.358	0.625	0.532	
11	0.260	0.017	1.000	0.644	1.762	0.078	K
12	0.200	0.166	1.000	1.053	1.330	0.184	
13	0.260	0.287	1.000	0.644	0.492	0.623	
14	0.180	0.295	1.000	1.204	0.705	0.481	
15	0.350	0.168	1.000	0.115	0.451	0.652	
16	0.390	0.309	1.000	-0.100	0.608	0.543	
17	0.370	-0.057	1.000	0.006	1.803	0.071	K
18	0.310	0.242	1.000	0.340	0.433	0.665	
19	0.360	0.242	1.000	0.060	0.637	0.524	
20	0.230	0.235	1.000	0.838	0.693	0.488	
21	0.320	0.065	1.000	0.282	1.188	0.235	
22	0.190	0.193	1.000	1.124	1.052	0.293	
23	0.150	-0.040	1.000	1.449	1.514	0.130	K
24	0.270	0.394	1.000	0.579	0.455	0.649	
25	0.430	0.235	1.000	-0.308	0.503	0.615	
26	0.280	0.337	1.000	0.517	0.408	0.683	
27	0.380	0.198	1.000	-0.047	0.415	0.678	
28	0.190	0.196	1.000	1.123	1.241	0.215	
29	0.440	0.048	1.000	-0.359	0.698	0.485	
30	0.220	0.108	1.000	0.905	1.400	0.162	
31	0.300	0.164	1.000	0.397	0.723	0.470	

Item ID	P	R	a	b	Z Resid	P	Flag(s)
32	0.580	0.111	1.000	-1.066	0.653	0.514	
33	0.190	0.009	1.000	1.122	1.065	0.287	K
34	0.140	0.168	1.000	1.537	1.312	0.189	
35	0.200	0.112	1.000	1.048	0.843	0.399	K
36	0.330	-0.048	1.000	0.225	1.646	0.100	K
37	0.520	0.282	1.000	-0.763	0.799	0.424	
38	0.110	-0.101	1.000	1.830	1.811	0.070	K
39	0.160	0.012	1.000	1.361	1.219	0.223	
40	0.180	0.244	1.000	1.199	0.789	0.430	
41	0.300	0.247	1.000	0.397	0.735	0.463	
42	0.510	0.390	1.000	-0.712	0.715	0.475	
43	0.190	0.099	1.000	1.122	0.985	0.324	
44	0.490	0.379	1.000	-0.612	0.647	0.518	
45	0.500	0.266	1.000	-0.662	0.469	0.639	
46	0.400	0.394	1.000	-0.153	0.538	0.591	
47	0.540	0.403	1.000	-0.863	0.851	0.395	
48	0.500	0.359	1.000	-0.662	0.660	0.509	
49	0.600	0.471	1.000	-1.169	1.411	0.158	
50	0.570	0.331	1.000	-1.015	0.814	0.416	
51	0.530	0.036	1.000	-0.813	0.663	0.507	K
52	0.570	0.428	1.000	-1.015	0.894	0.371	
53	0.190	-0.175	1.000	1.121	1.492	0.136	K
54	0.770	0.286	1.000	-2.138	0.751	0.453	
55	0.630	0.300	1.000	-1.325	0.924	0.355	
56	0.210	0.117	1.000	0.974	1.157	0.247	
57	0.560	0.440	1.000	-0.964	1.106	0.269	
58	0.490	0.443	1.000	-0.612	1.357	0.175	
59	0.640	0.284	1.000	-1.378	1.102	0.271	
60	0.370	0.060	1.000	0.005	1.046	0.295	K
61	0.210	-0.171	1.000	0.974	1.893	0.058	K
62	0.440	-0.217	1.000	-0.359	1.807	0.071	K
63	0.310	0.412	1.000	0.339	0.598	0.550	
64	0.230	0.185	1.000	0.836	0.574	0.566	
65	0.640	0.138	1.000	-1.378	1.275	0.202	
66	0.400	0.310	1.000	-0.153	0.404	0.687	
67	0.540	0.343	1.000	-0.863	0.642	0.521	
68	0.420	0.194	1.000	-0.257	0.680	0.496	
69	0.230	0.152	1.000	0.835	0.811	0.418	K
70	0.030	0.330	1.000	2.960	0.727	0.467	
71	0.580	0.351	1.000	-1.066	0.653	0.514	
72	0.660	0.156	1.000	-1.486	0.952	0.341	
73	0.550	0.236	1.000	-0.913	0.552	0.581	
74	0.490	0.443	1.000	-0.612	0.824	0.410	
75	0.450	0.448	1.000	-0.410	1.250	0.211	
76	0.650	0.168	1.000	-1.432	0.548	0.584	
77	0.670	0.373	1.000	-1.541	1.079	0.280	
78	0.460	0.299	1.000	-0.461	0.327	0.744	
79	0.280	-0.061	1.000	0.516	1.782	0.075	K
80	0.180	0.145	1.000	1.198	1.167	0.243	

Table 8. Correlation evidence of item-by-item reliability of the OKUKINS Chemistry Achievement Test

Comparability	No. of items	r	P Value
Sample O VS Sample A	80	0.946	0.000
Sample O VS Sample B	80	0.604	0.000
Sample A VS Sample B	80	0.721	0.000

Table 8 shows the correlation coefficient on the Standard Errors of Measurement across samples associated with difficulty parameters. Relating the Standard Errors of Measurement associated with b parameters in Sample O (SEM_O) with Sample A (SEM_A), a coefficient of 0.946 was obtained. Relating the Standard Errors of Measurement associated with b parameters in Sample O (SEM_O) with Sample B (SEM_B), a coefficient of 0.604 was obtained. Relating the Standard Errors of Measurement associated with b parameters in Sample A (SEM_A) with Sample B

(SEM_B), a coefficient of 0.721 was obtained. P value of 0.000 was obtained from the correlation of SEM_O versus SEM_A, SEM_O versus SEM_B and SEM_A versus SEM_B. The p value of 0.000 is statistically significant at 0.05 alpha level, thus, the item-by-item reliability of the OKUKINS Chemistry Achievement Test is verified through correlation evidence using the Standard Errors of Measurement across samples associated with the difficulty parameter. According to [14], item reliability as used in IRT, is the degree in which item difficulties are differentiated. In this case, the difficulty parameters of the items of the OKUKINS Chemistry Achievement Test are appropriately differentiated.

Figure 1 presents the Test Response Function (TRF) for all calibrated items. The Test Response Function predicts the number of items an examinee will respond correctly as a function of theta. In this case, the TRF predicts 84.6% of the score of each examinee on the OKUKINS Chemistry Achievement Test.

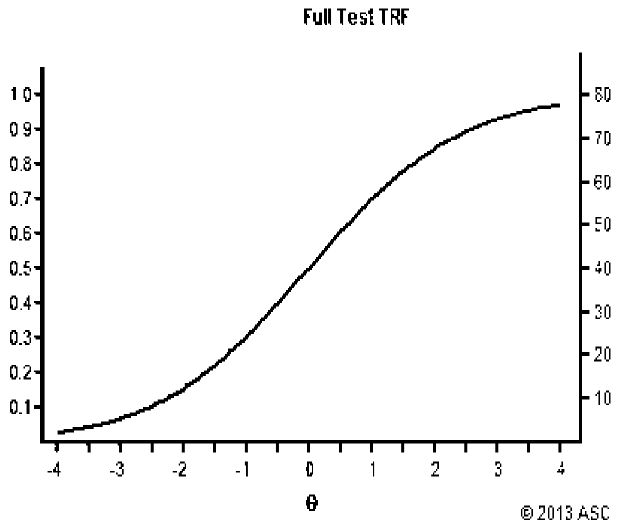

Figure 1. Test Response Function

Table 9. Correlation evidence of person reliability of the OKUKINS Chemistry Achievement Test

Comparability	No. of items	r	P Value
Sample O VS Sample A	80	0.595	0.000

Table 9 indicates the correlation coefficient of the Standard Errors of Measurement across samples associated with ability estimates. Relating the Standard Errors of Measurement associated with ability estimates in Sample A (SEM_A) with that of Sample B (SEM_B), a coefficient of 0.595 was obtained, with probability of 0.000. The P value of 0.000 is statistically significant at 0.05 alpha level. Thus, the Standard Errors of Measurement (SEM) across samples associated with ability estimates are consistent. This is an evidence of person reliability of the OKUKINS Chemistry Achievement Test. Person reliability is the degree in which a measuring instrument differentiates persons in the test outcome. In this case, the OKUKINS CAT differentiates ability appropriately in the test outcome.

4. Item Characteristic Curves of Three Items of the Okukins Cat

Figure 2 to Figure 4 present the Item Characteristics Curves for three items of the OKUKINS CAT. Also presented along side with the ICC are four tables.

1) **Item Information Table:** This table records the information supplied by the control file (classic Data Header) for the item.
2) **Classical Statistics Table:** This table presents the classical statistics for the item.
3) **IRT Parameter:** This table records item parameter estimates for the item.
4) **Option Statistics:** This table provides detailed statistics for each item, which helps to identify issues in items with poor statistics.

In the classical statistics table, the P value and the point-biserial correlations are presented in the first three columns. The P value is the proportion of testees that answered an item in the keyed direction. It ranges from 0 to 1. The S- Rpbis and T- Rpbis are the point- biserial correlations of an item with total score and theta respectively. The last column in the classical statistics table is the Alpha w/o which is the Cronbach's alpha computed with the current item excluded.

The IRT parameters table presents data for item parameters and fit statistics. In this case, the discriminative parameter is constant for all items i.e. a = 1.0, while the b parameter varies depending on the difficulty of the item. Also presented in the IRT parameter table is the Standard Error (SE) for each item. A large SE for an item parameter (compared to the other items) shows that the item parameter was poorly estimated. The z Residual is also presented in the IRT parameter table, used to determine the fit statistic of a dichotomous item. For dichotomous items, the P value for rejecting the item as poor fit was computed using the z residual with the standard normal distribution as its sampling distribution.

In the option statistics table, the responses of examinees across the options for each item are clearly indicated. [8] used five verbal terms to describe the difficulty of an item: very easy --- easy --- medium –hard --- very hard.

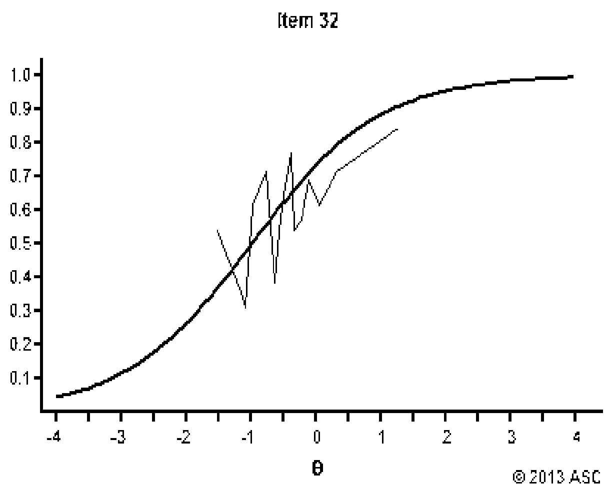

Figure 2. OKUKINS CAT Item 32

According to [14] "Higher b-parameters (> 1.0) indicate that the item is more difficult, a value below -1.0 indicates that the item is very easy". Following the interpretation of Kpolovie and Emekene, it can be inferred from the result that, item 32 is an easy item; item 13 is an item of medium difficulty and item 28 is difficult item.

Table 10.1. OKUKINS CAT Item 32 Item information

Seq.	ID	Model	Key	Scored	Num Options	Domain	Flags
32	32	1PL	A	Yes	5	1	

Table 10.2. OKUKINS CAT Item 32 Classical statistics

N	P	S-Rpbis	T-Rpbis	Alpha w/o
200	0.590	0.168	0.216	0.845

Table 10.3. OKUKINS CAT Item 32 IRT parameters

a	b	a SE	b SE	Chi-sq	Df	p	z Resid	p	InMSQ	InZstd	OutMSQ	OutZstd
1.000	-0.941	0.220	0.151	12.447	14	0.570	0.814	0.416	1.061	0.416	1.068	0.433

Table 10.4. OKUKINS CAT Item 32 Option statistics

Option	N	Prop.	S-Rpbis	T-Rpbis	Mean	SD	
A	118	0.590	0.168	0.216	-0.305	0.714	**KEY**
B	46	0.230	0.009	-0.022	-0.453	0.553	
C	9	0.045	-0.172	-0.183	-0.992	0.470	
D	7	0.035	-0.047	-0.054	-0.617	0.373	
E	20	0.100	-0.140	-0.163	-0.754	0.606	
Omit	0						
Not Admin	0						

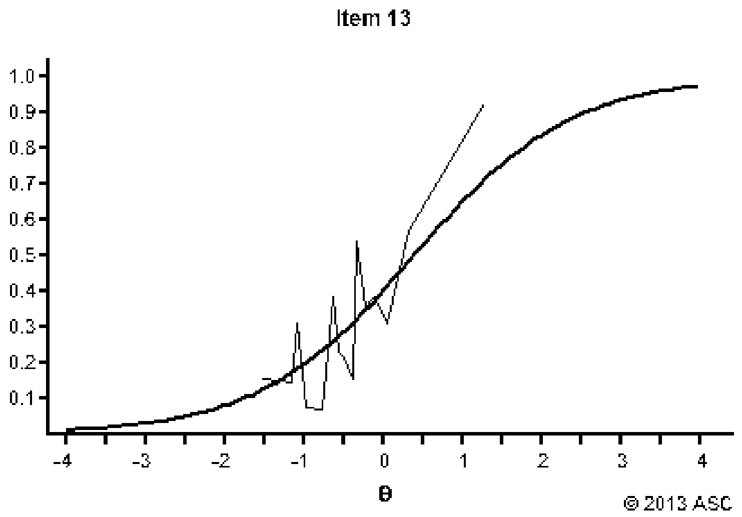

Figure 3. OKUKINS CAT Item 13

Table 11.1. OKUKINS CAT Item 13 information

Seq.	ID	Model	Key	Scored	Num Options	Domain	Flags
13	13	1PL	C	Yes	5	1	

Table 11.2. OKUKINS CAT Item 13 Classical statistics

N	P	S-Rpbis	T-Rpbis	Alpha w/o
200	0.320	0.365	0.402	0.842

Table 11.3. OKUKINS CAT Item 13 IRT parameters

A	b	a SE	b SE	Chi-sq	Df	p	z Resid	P	InMSQ	InZstd	OutMSQ	OutZstd
1.000	0.467	0.160	0.161	15.088	14	0.372	0.484	0.628	0.953	-0.081	0.953	-0.037

Table 11.4. OKUKINS CAT Item 13 Option statistics

Option	N	Prop.	S-Rpbis	T-Rpbis	Mean	SD	
A	32	0.160	0.000	-0.016	-0.450	0.506	
B	53	0.265	-0.211	-0.225	-0.677	0.484	
C	64	0.320	0.365	0.402	-0.033	0.819	**KEY**
D	17	0.085	-0.160	-0.166	-0.791	0.600	
E	34	0.170	-0.088	-0.097	-0.569	0.416	
Omit	0						
Not Admin	0						

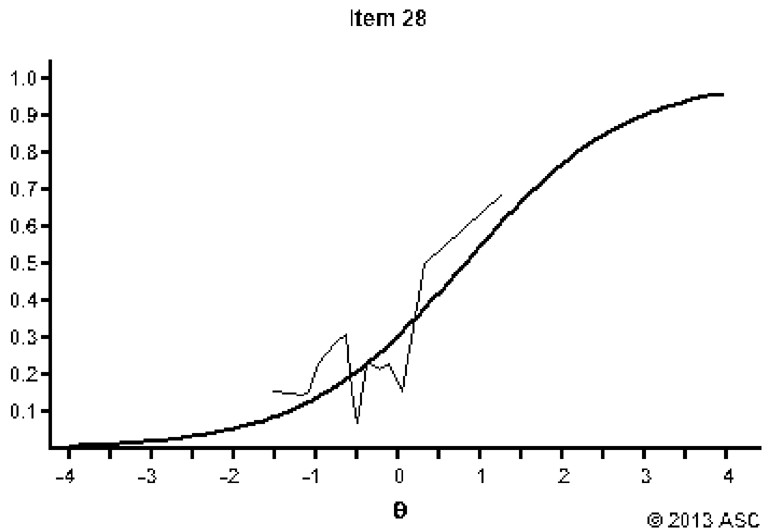

Figure 4. OKUKINS CAT Item 28

Table 12.1. OKUKINS CAT Item 28 information

Seq.	ID	Model	Key	Scored	Num Options	Domain	Flags
28	28	1PL	A	Yes	5	1	

Table 12.2. OKUKINS CAT Item 28 Classical statistics

N	P	S-Rpbis	T-Rpbis	Alpha w/o
200	0.250	0.238	0.277	0.844

Table 12.3. OKUKINS CAT Item 28 IRT parameters

A	b	a SE	b SE	Chi-sq	Df	p	z Resid	p	InMSQ	InZstd	OutMSQ	OutZstd
1.000	0.894	0.133	0.176	9.926	14	0.768	0.930	0.352	1.066	0.300	1.115	0.394

Table 12.4. OKUKINS CAT Item 28 Option statistics

Option	N	Prop.	S-Rpbis	T-Rpbis	Mean	SD	
A	50	0.250	0.238	0.277	-0.104	0.836	**KEY**
B	24	0.120	-0.017	-0.033	-0.485	0.618	
C	44	0.220	0.107	0.094	-0.307	0.655	
D	18	0.090	-0.254	-0.258	-0.977	0.434	
E	64	0.320	-0.149	-0.159	-0.582	0.443	
Omit	0						
Not Admin	0						

Figure 5 displays a graph of the information function for all calibrated items of the OKUKINS CAT. The Test Information Function (TIF) is a graphical representation of how much information the test is providing at each level of theta. In this case, the maximum information was 16.561 at theta = -0.050, thus, the TIF provides satisfactory information over the ability trait range since it takes the shape of a normal distribution curve.

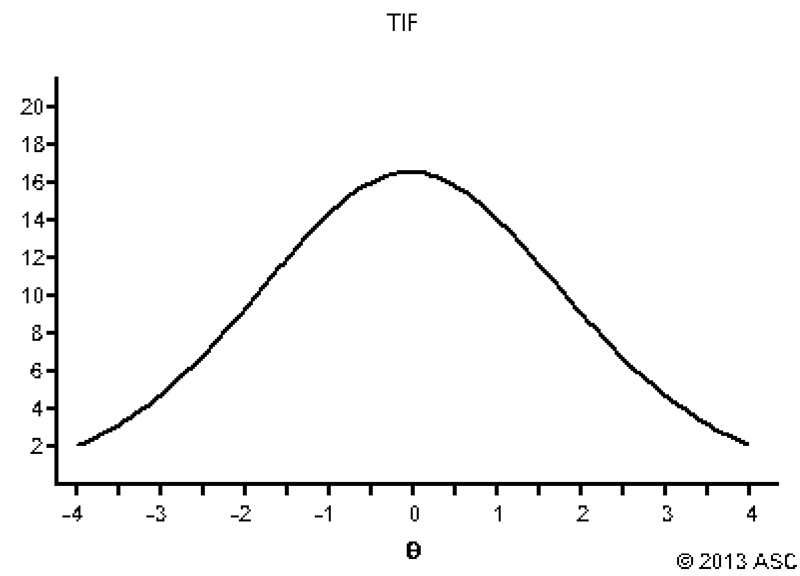

Figure 5. Test Information Function

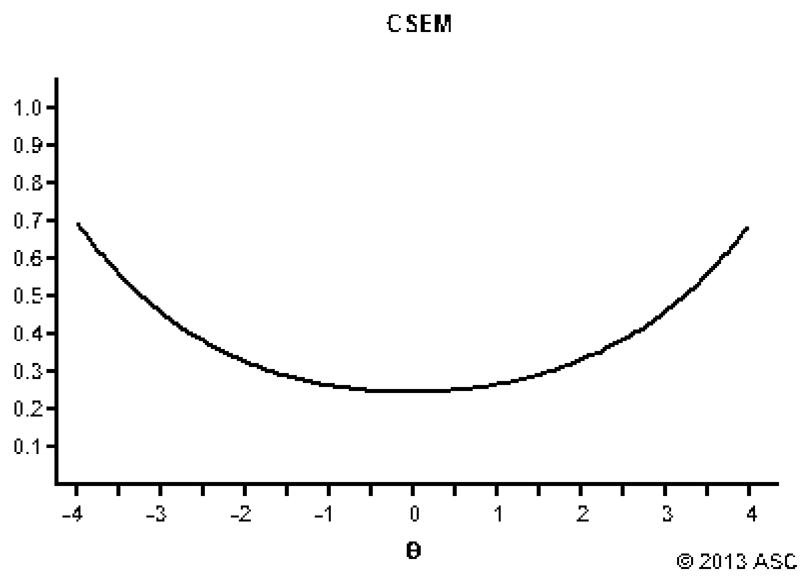

Figure 6. CSEM Function

Figure 6 presents the graph of the Conditional Standard Error of Measurement (CSEM) Function. The CSEM is an inverted function of the TIF. It estimates the amount of error in theta estimation for each level of theta. In this case, the minimum CSEM was 0.246 at theta = -0.050.

Table 13. t test analysis of the influence of gender on students' academic achievement in Chemistry

Gender	N	Mean	SD	df	t-cal	p-value	Decision
Male	100	34.91	11.83	198	1.598	0.112	Not significant
Female	100	32.48	9.56				

Table 13 shows the mean scores of male and female students' academic achievement in the OKUKINS Chemistry Achievement Test as 34.91 and 32.48 respectively, with respective standard deviation of 11.83 and 9.56. This suggests that, there is a difference between male and female students Performance in chemistry in the locale under investigation. When these values were subjected to independent sample t test analysis, it was revealed that the calculated t which is 1.598 is not statistically significant at 0.05 alpha level. Since the probability (p- value) of 0.112 is greater that the significant level of 0.05, the null hypothesis is therefore accepted. That is, there is no significant difference in the academic achievement of male and female students in the OKUKINS Chemistry Achievement Test.

5. Discussion of the Findings

The result of this study indicated that the items of the OKUKINS CAT fitted perfectly the 1-PLM when a sample size of 100 examinees was used for item calibration than when a sample size of 200 examinees was used. Using a sample size of 200 examinees, 74 items out of the 80 items which form the OKUKINS CAT fitted the One-Parameter Logistic Model (I-PLM). Using a sample size of 100 examinees (set A), 79 items fitted the 1-PLM.

Using a sample size of 100 examinees (set B), all the 80 items of the OKUKINS CAT fitted the One-Parameter Logistic Model. The results are in consonance with that of [15] who reported that the 15 items of the Advanced Progressive Matrices Scale- Smart Version (APM-SV) fitted perfectly the One-Parameter Logistic Model (1-PLM), Two-Parameter Logistic Model (2-PLM) and Three-Parameter Logistic Model (3-PLM). The findings diverges from that of [15] who reported that out of the 40 items of the 2010 Botswana Junior Certificate (JC) Mathematics paper 1, only one item fitted the 1-PLM, eleven (11) items fitted the 2-PLM and 23 items fitted the 3-PLM. Based on the result, the 3-PLM was used to estimate the item parameters.

One of the result of the study showed that, the difficulty parameters of the items of the OKUKINS CAT yielded satisfactory result. The result of item parameters for all calibrated items showed that, using sample O, the b parameters ranged from -2.417 to +2.834. Using sample A, the difficulty parameters ranged from - 2.115 to +2.293. Using sample B, the b parameters ranged from -2.138 to +2.960. According to the Xcalibre manual, difficulty index ranges in theory from negative infinity to positive infinity, but practical consideration is usually from -3.0 (very easy) to + 3.0 (very difficult). In this case, the b parameter across samples graduated from very easy to very difficult. Similar result was found in a related study by [16] who used the 3-PLM to calibrate all the 36 items of the Advanced Progressive Matrices (APM) Scale. The result of item parameters for all calibrated items yielded b parameters ranging from -2.595 to + 2.133.

The results of this study also revealed that, the Test Information Function (TIF) of the OKUKINS CAT provide satisfactory information over the ability trait range since it took the shape of a normal distribution curve. Using sample O, the maximum information was 16.561 at theta = -0.050. Using sample A, the maximum information of the OKUKINS CAT provided was 16.494 at theta = -0.050. Using sample B, the maximum information was 16.395 at theta = -0.050. In a similar study, [16] reported maximum information of 8.258 at theta = 0.000 for the Advance Progressive Matrices (APM) Scale. Also, [14] reported maximum information of 8.090 at theta = - 0.200 for APM –SV. The TIF in the two investigations also took the shape of a normal distribution curve.

It was revealed in this study that, gender has no significant influence on academic achievement in chemistry. The mean scores of the male and females students were 34.91 and 32.48 respectively, with respective standard deviation of 11.83 and 9.56. When these observations were subjected to independent t-test analysis, it was revealed that, no significant difference exist between the academic achievement of the male and female students in the OKUKINS Chemistry Achievement Test.

The overall result showed that, gender has no significant influence on academic achievement. Similar results were found in related studies by Nwachukwu and Inomiesa as cited in [17] that, gender does not really counts in academic achievement in the sciences. That is, there is no significant difference between the performance of male and female students in the sciences. However, [18] and [19] reported contrary results that, there is significant difference between the performance of male and female students in favour of the male students. In a different dimension, Weerakkody and Ediriweera reported that, significant difference exist between the performance of male and female students in favour of the female students.

Over the years, research indicated that, the African woman was never perceived traditionally to be on equal footing with the male counterpart. The corollary deprivation and marginalization of the African woman has been substantiated in the investigation of Otite and Ogionwo as cited in [18]. The pervasive system of patriarchy placed the man on an enviable and revered pedestal and gave him ample latitude and social leverage to Lord it over the African woman, thereby relegating them to the position of subservience. According to Jike and Buadi as cited in [17] “this patriarchally –induced gender dichotomies have been transposed to the education system”.

Nosike as cited in [18] observed that, social customs often account for the popular belief that, girls do not need education since they will marry and raise children other than work at a Job outside their home where educational qualifications are required. In line with the above, most girls do opt for expressive Jobs like cooking, cleaning and child nurturing in tandem with social perception of woman’s role. The result of this study clearly showed that, the subservient position of the African woman as noted by Jike and Buadi and the social customs substantiated by Nosike have been over-taken by events. Owing to sensitization and more enlightenment, the female students have now on equal footing to compete with their male counterparts in the academia.

6. Contributions to Knowledge

Based on the findings, the following are contributions to existing knowledge

1. Through the study an instrument to measure students’ true ability had been developed.
2. An instrument that will aid tailor testing is brought to the open.
3. Through the study, the workability of IRT in developing a test is determined.

7. Conclusion

Based on the findings, the following conclusions are made:

1) The OKUKINS CAT fitted the One-Parameter Logistic Model of IRT.
2) The difficulty parameters of the items of the OKUKINS CAT ranged from -2.417 (very easy) to +2.834 (very difficult)using sample O. Using sample A the difficulty parameters ranged from -2.115 (very easy) to +2.293 (very difficult) using sample B, the difficulty parameters ranged from -2.138 (very easy) to +2.960 (very difficult). Thus, the range if difficulty parameters of the items of the OKUKINS CAT fall within the range of difficulty parameters prescribed by the Xcalibre manual i.e +3.0 (very easy) to -3.0 (very difficult).

3) The ability estimate of the OKUKINS CAT ranged from -2.276 to +2.163 using sample O; from -1.743 to + 1.652 using sample A; from -2.278 to +2.290 using sample B, all yielding favourable statistics.
4) The standard errors of measurement associated with difficulty parameters ranged from 0.148 to 0.344 using sample O, from 0.210 to 0.382 using sample A; from 0.208 to 0.558 using sample B, all producing favourable indices.

The standard errors of measurement associated with ability estimates ranged from 0.2457 to 0.3128 using sample O, from 0.246 to 0.3047 using sample A; from 0.247 to 0.3582 using sample B, all yielding favourable statistics.

8. Recommendations

Based on the findings of this study the following recommendations are made:

1. IRT software such as the Xcaliber 4.2 should be made available in every institution of higher learning considering its uniqueness in data analysis.
2. All examination bodies such as JAMB, WAEC, NECO and NABTEB should re-visit and review their test items using the IRT models since this has been recommended by the International Testing Commission.
3. Effort should be made by psychometricians in Nigeria to promote the use of IRT models during test construction.
4. The female students should be encourage to opt for science courses in the institution of higher learning, since the result of this study had disclosed that gender does not influence performance in the sciences.
5. Non-Governmental Organizations (NGOs) interested in measuring the basic ability of students in Chemistry for the sake of scholarship and concerned parents and guardians who want to assess the learning progress of their children and wards in Chemistry should opt for the OKUKINS Chemistry Achievement Test.

References

[1] Oku, K. (2015) *Measures for achieving academic excellence*. Paper presented at the lesser chapter All Saints' Cathedral (Anglican Communion), Ughelli Delta State, during the AYF youth week celebration.

[2] Iweka, F. (2014) *Comprehensive guide to test construction and administration*. Omoku: Chifas Nigeria.

[3] Steinmagr, R, Meibner, A, Weidinger, A.F,& Wirthwein, L. (2015). *Academic achievement*. Retrieved from http://www. Oxford bibliographies.com/view / document / obo -9780199756810/ obo -9780199756810 -0108. Xml.

[4] Magno, C. (2009). *Demonstrating the difference between classical test theory and item response theory using derived test data. The international journal of educational and psychological assessment,* 1(1), 1-11.

[5] Ojerinde, D. (2013) Classical Test Theory VS Item Response Theory: An evaluation *of the comparability of item analysis results.* Retrieved from http://ui.edu.ng/sites/default/files/PROF%20OJERINDE%27S%20LECTURE%20(Autosaved).pdf.

[6] Kpolovie, P.J. (2014) *Test measurement and evaluation in education*. Owerr i : Springfield.

[7] 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.

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

[9] Anne, M.H. (2015) *Why is chemistry important?* Retrieved from https://www.thoughtco.com/why-is-chemistry-important-604144.

[10] Jason, S. (2013) *The importance of chemistry in everyday life.* Retrieved from https://sciencezoneja.wordpress.com/2013/12/24/the-importance-of-chemistry-in-everyday-life/.

[11] Ogunbanwo, R. A. (2014) *Analysis of students' performance in West African Senior Certificate Examination in boarding and day secondary schools in Kano metropolis, Nigeria.* Master Thesis, Ahmadu Bello University, Zaria.

[12] Wheadon, (2014) *Classification: Accuracy and consistency under item response theory models using the package classify. Journal of statistical software*, 56, issue 10.

[13] Dayalata, A. & Obinne, E. (2013) Test item validity: Item response theory(IRT) perspective Nigeria. Research journal in organizational psychology and education studies, 2(1),

[14] 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.

[15] Omobola, O.A & Adedoyin, J. A (2015) *Assessing the comparability between Classical Test Theory (CTT) and Item Response Theory (IRT) models in meters. Herald journal of education and general studies* 2(3), 107 -114.

[16] Emekene, C.O. (2017) *Psychometric analysis of the advanced progressive matrices scale for use in Nigeria*. Ph.D Thesis, University of Port Harcourt.

[17] Oku, K. (2008) *Comparison of academic achievement of sandwich and regular postgraduate students of the University of Port Harcourt*. Unpublished master thesis, University of Port Harcourt.

[18] Ileh, L. N. (2017) *Development and validation of Mathematics achievement test for continuous assessment in junior secondary schools in Delta State.* Unpublished thesis, Delta State University, Abraka.

[19] Akpochafo, W. P.(2003) *Gender related differences and academic achievement in junior secondary school social studies. Journal of educational research and development,* 2(1), 35-42.