European Journal of Statistics and Probability

Vol.4, No.3, pp.20-60, July 2016

Published by European Centre for Research Training and Development UK (www.eajournals.org)

PSYCHOMETRIC ADVENT OF ADVANCED PROGRESSIVE MATRICES – SMART VERSION (APM-SV) FOR USE IN NIGERIA

Kpolovie, P.J* and Emekene C.O.

Academic Planning, Research & Control Unit; Vice-Chancellor's Office; University of Port Harcourt; PMB 5323; Choba; Port Harcourt; Nigeria.

ABSTRACT: *The Raven's Advanced Progressive Matrices (APM) test is a leading global non-verbal measure of mental ability, helping to identify individuals with advanced observation and clear thinking skills who can handle rigorous study programmes as well as the complexity and ambiguity of the modern workplace. APM scale is largely employed by researchers and practitioners in the field of psychometrics, education, medicine and the social sciences. A sample of 3100 participants in Nigeria was randomly drawn to answer nine research questions. Triangulation research design, adopting item response theory (IRT) guided the study. The study developed an abridged form of the APM dubbed Advanced Progressive Matrices-Smart Version (APM-SV). Results revealed that all 15 items of the APM-SV test yield favourable statistics under 3-Parameter Logistic IRT Model with regards to item discrimination, difficulty and guessing. Item Response Function showed preponderance of APM-SV's reliability of 0.92. The APM-SV showed perfect fit, is bias-free and very suitable for use in Nigeria. APM-SV scale strongly and positively correlated well with other measures of fluid ability such as the APM scale itself, CFIT, Digit Span scale, and Bennett Mechanical Comprehension Test (BMCT).*

Keywords: Advanced Progressive Matrices, Item Response Theory, X-Calibre, Test Bias, 3-Parameter Logistic Model, Nigeria

INTRODUCTION

Psychometrics is the field of study concerned with the theory and technique of psychological measurements, which includes the measurement of knowledge, abilities, attitudes, personality traits and educational measurements. Michell (1999) says psychometrics can be defined as the branch of psychology concerned with the design and use of psychological tests and the application of statistical and mathematical techniques to psychological testing. Raven's Progressive Matrices tests developed by John C. Raven in 1936 and first published in 1938 are examples of psychological testing tools. Raven's tests exist in three different forms that are progressively more difficult in contents intended for different populations. They are the Standard Progressive Matrices (SPM), the Coloured Progressive Matrices (CPM) and the Advanced Progressive Matrices (APM). APM scale published in 1947 is the most difficult of the three and it is the main instrument of this study. APM test is a leading global non-verbal measure of mental ability, helping to identify individuals with advanced observation, high-level imagination including the domain of duty and clear thinking skills who can handle rigorous study programmes as well as the complexity and ambiguity of the modern workplace. APM test offers information about someone's capacity for analysing and solving problems, abstract reasoning, logical reasoning, quick recognition of differences and similarities, intellectual capacity and the ability to learn. The APM scale assesses the ability or capacity to detect a certain order or structure in a chaos or chaotic situation and the ability to find meaning of apparently randomly compiled elements. It reduces cultural biases with a nonverbal

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

approach. It is very suitable for individuals whose native language is not English (Kpolovie, 2016b; Carlson, Geisinger & Jonson, 2014; Raven, Raven & Court, 2012).

When administered untimed, the Advanced Progressive Matrices differentiates between people at the high end of intellectual ability. When administered under timed conditions, the APM can be used to assess intellectual efficiency - quick and accurate high-level intellectual work and the ability to be sharp and quick at decision making (Raven, 1962; Raven, Raven & Court, 2012; 1998). Items on all forms of Raven's Progressive Matrices ask the examinee to identify the missing component in a series of figural patterns. Grouped in sets, the items graduates in the difficulty index from very easy items to very difficult items. Therefore the items require increasingly greater skills in encoding, analysing, recognizing patterns and identifying the right answers. The Raven's APM produces a single raw score as well as percentile rank to indicate the candidate's educative ability or the ability to make sense of complex situations, compared to a norm group (Raven, Raven and Court, 2012). Evers (2011) presented the report of a large survey conducted in nineteen European countries by several members of the International Test Commission (ITC) at the 12th European Congress of Psychology that held in Istanbul in the month of July 2011. He asserted that "the Raven's Matrices are in the fourth position among the ten most used tests in Europe." The report further stated that "among them the Advanced Progressive Matrices are widely employed for assessing fluid ability in adolescents and adults." Evers (2011) also reported that "the Raven's Advanced Progressive Matrices (APM) scale has been recommended as a useful measure for identifying academic potential;" and that the APM is in high demand as an instrument of choice among researchers in America, Europe and Asia because of its utility value in psychological research works. Meanwhile, the instrument is hardly known let alone effectively employed in psychological research works in Africa, particularly Nigeria. The use of the APM in Nigeria will enhance the identification, placement, acceleration and enrichment of the gifted/talented students in Nigerian and ability selection and placement within and outside the education management climes. An abridged form of the APM dubbed APM-SV which can serve as a quick measure of fluid ability and a reliable alternative to the full form of APM will serve a veritable and useful purpose in the measurement of intelligence in Nigeria. The review of literature of this study focused on conceptual review, theoretical framework, related empirical studies and summary of literature review.

The problem of this study can be categorized into three. First, it has been suggested that "one of the constraints in the use of the Advanced Progressive Matrices (APM), especially when included in a test battery, is its length and the time of administration" (Evers, 2011). In an earlier investigation of Item Response Theory validation of Advanced Progressive Matrices in Nigeria, Kpolovie and Emekene (2016) found that though the test has high validity and reliability, and is bias-free in Nigeria; there is need for exploration of possibility of psychometrically arriving at a shorter version of Advanced Progressive Matrices in Nigeria. As a way out, the current study is aimed at modifying the APM scale to create a short form named by these researchers as the Advanced Progressive Matrices – Smart Version (APM-SV). Secondly, the International Test Commission recommended that IRT be used for the proper description and evaluation of existing and widely used psychological instruments (Muñiz, 2011). The short-form of the APM scale is yet to be examined in Nigeria with IRT. Therefore IRT will be applied on the APM-SV scale. Thirdly, it is feared that with the type of cultural diversity in Nigeria, a test may not suitably measure mental ability without bias (Gilovich, Griffin & Kahneman, 2002). This investigation is therefore also aimed at ascertaining whether the APM-SV could actually be biased in measuring the attribute in Nigeria or not.

The purpose of this study therefore is to create a more user friendly version of the APM dubbed Advanced Progressive Matrices – Smart Version (APM-SV) by the researchers and to solve the identified and categorized problems, using multiple perspectives that satisfactorily establish the reliability and validity of APM-SV in addition to empirically determining whether the APM-SV can suitably be used in Nigeria without biases. Consequently, nine research questions as follows were posed and answered in this study because they all help in inferring test reliability, validity and bias in IRT.

1. Which is the most suitable Item Response Theory (IRT) Parameter Logistics Model (PLM) for Raven's Advanced Progressive Matrices-Smart Version (APM-SV) in Nigeria?
2. What is the Overall Model Fit of APM-SV using Nigerian validation sample?
3. What is the person separation reliability of APM-SV that can be inferred from the contribution of each of the items to the Test Response Function (TRF)?
4. What is the Item Response Function (item-by item) evidence of reliability of APM-SV in Nigeria?
5. What is the evidence of unidimensionality, if any, of APM-SV in Nigeria?
6. What is the b-parameter index (item difficulty parameter) for each APM-SV item in Nigeria?
7. What is the range of Differential Item Function (discriminatory index) popularly referred to in IRT as a-parameter of the APM-SV in Nigeria?
8. What is the c-parameter (probability of guessing) for each APM-SV item as evidence of bias culturally (ethnic group, school type/age, and sex)?
9. What are the Correlations between APM-SV scale and other measures of mental ability such as working memory measures, Mechanical Reasoning ability, Mathematical Reasoning ability, Probabilistic Reasoning ability, APM and CFIT?

METHODOLOGY

Multiple triangulation research design which Kpolovie (2010) described as "the highest and most comprehensive and all-embracing form of triangulation research" was employed in this study. This design was used because "it allows for multi-method approach in studying psychometric properties of an instrument and some aspects of human behaviour. It helps to map out or explain more fully, the richness and complexity of a psychometrical instrument and/or human characteristics by studying it from more than one stand point" (Kpolovie, 2010). This research design allowed for application of various methods including the IRT logistic models (1-PLM, 2-PLM, 3-PLM), DIF, TIF, Factor analyses using the data reduction option, Test of Fitness of Good Statistics, among many other statistical procedures. The study was carried out in Nigeria. The population of the study comprised all the university undergraduates (1,794,989) and all the senior secondary school students in Nigeria (4,758,739); making a total population of 6,553,728 (FRN National Population Commission, 2014; Federal Ministry of Education, 2014; NEEDS Assessment of Nigerian Universities, 2013). A total sample of 3,100 (1,500 undergraduates and 1,600 secondary school students) was randomly drawn, using disproportional stratified random sampling technique (Kpolovie, 2011) as participants in the study. The sample had males and females between the age range of 12 and 40 years; spread over four main cultural groups (Hausa, Igbo, Yoruba and Minorities) that adequately cover four geopolitical zones in Nigeria.

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

The main instrument of this study was the APM scale. Eight research assistants were engaged, trained to administer the tests and accompanied the researcher to the four geopolitical zones to conduct the tests to the various participants. The exercise was carried out in two parts. A total of 2100 (1,000 undergraduates and 1,100 secondary school students) took part in the first exercise while 1000 (500 undergraduates and 500 secondary school students) took part in the second exercise. The APM scale with 36 items was the only instrument used in the first round of field work. The APM scale comes in two sets. Set 1 and Set 2. Set 1 contains 12 items while Set 2 contains 36 items. The Set 1 items were used as practice test. The set 2 which is the main scale of this study was the real test. Each item has eight options from which the participant is expected to select one option. The options selected by the participants to each item on the test were subjected to IRT analysis using the X-Calibre 4.2 software. The X-Calibre 4.2 IRT analysis was performed using the three different Parameter Logistic Models, that is, the 1-Parameter Logistic Model (1-PLM), the 2-Parameter Logistic Model (2-PLM) and the 3-Parameter Logistic Model (3-PLM). The other statistics performed using the X-Calibre software includes: The Differential Information Function (DIF) for group comparisons: gender (male and female), age as inferred from the school categories (university undergraduates within the age range of 16-40 years old and senior secondary school students within the age range of 12-20 years old) ethnicity (Hausa, Igbo, Yoruba and Minorities), and, Test Information Function, Item Information Function, Item-by-Item Analysis that included the Item characteristic curve also known as Item Response Function. SPSS was used to perform dimension reduction analysis, reliability analysis, correlation analysis, and Factor Analysis. After collating, scoring and analysing scores generated from the APM test, the APM was modified to create the APM-SV scale on the basis items with highest loadings and which best fitted into the IRT 3-Parameter Logistic Model. Thereafter the APM-SV scale was administered to a new set of sampled respondents in a second round of outing. In all a total of 500 university undergraduates and 500 senior secondary school students participated in the second exercise of tests administration. During the second outing the APM-SV was the main instrument. Six other instruments that included the APM, Culture Fair Intelligence Test (CFIT), Digit Span scale of the WAIS-R, Heuristics and biases literature tasks, Mathematics Ability Test and Bennett Mechanical Comprehension Test were used. The researchers used the following statistical packages: X-Caliber 4.2 and SPSS version 22. These statistical packages were employed in accordance with the statistical triangulation demands of Multiple Triangulation research design (Kpolovie, 2016; 2015; Verguts & De Boeck, 2002). Ololube, Kpolovie and Makewa (2015); Ojerinde, Popoola, Ojo and Onyeneho (2012); and Ojerinde, Popoola, Ojo and Ariyo (2014) have equally called for use of these three statistical packages for analysis of data in an investigation of this nature. Furthermore, Guyer & Thompson (2011) posited that "Item response theory (IRT) presents a powerful psychometric paradigm for developing, delivering, analysing, and scoring assessments, and that in order to utilize IRT with the aim of obtaining accurate results, assessment data must be calibrated with sophisticated software designed for that purpose." Similar calls have also been made by Muniz (2009), Orluwene (2012), Raven, Raven and Court (1993), Vigneau and Bors (2015), and Joint Admissions and Matriculation Board JAMB (2016).

RESULTS

Modification of APM scale and Construction of the Advanced Progressive Matrices-Smart Version (APM-SV) scale

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

In order to construct an abridged version of the APM known as the Advanced Progressive Matrices – Smart Version, the APM test with 36 items was administered to 2100 (1,000 undergraduates and 1,100 secondary school students). The results were collated, graded and analyzed. Of importance were the Pearson point-biserial correlation (*r-pbis*) and Item Difficulty for the scores of the Raven's Advanced Progressive Matrices (APM) performed.

Table 1: 3-PLM Item Parameters for All Calibrated Items

Seq.	Item ID	P	R	a	b	c	Flag(s)
1	1	0.746	0.159	0.640	-2.595	0.499	
2	2	0.760	0.158	0.756	-2.574	0.189	
3	3	0.758	0.045	0.663	-2.557	0.331	
4	4	0.754	0.060	0.772	-2.546	0.448	
5	5	0.735	0.089	0.607	-2.436	0.252	
6	6	0.765	0.042	0.746	-2.394	0.252	
7	7	0.721	0.158	0.834	-2.234	0.249	
8	8	0.750	0.044	0.674	-2.170	0.253	
9	9	0.734	0.035	0.550	-2.092	0.253	
10	10	0.753	0.085	0.752	-2.076	0.251	
11	11	0.730	0.049	0.608	-2.049	0.253	
12	12	0.734	0.108	0.677	-2.001	0.252	
13	13	0.730	0.129	0.760	-1.919	0.251	
14	14	0.797	0.104	0.704	-1.891	0.255	
15	15	0.714	0.136	0.748	-1.505	0.252	
16	16	0.740	0.090	0.703	-1.347	0.251	
17	17	0.740	0.155	0.803	-1.145	0.251	
18	18	0.743	0.124	0.772	-1.076	0.250	
19	19	0.733	0.093	0.590	-1.041	0.253	
20	20	0.709	0.116	0.599	-1.007	0.253	
21	21	0.719	0.207	0.616	0.085	0.252	
22	22	0.685	0.128	0.557	0.105	0.256	
23	23	0.716	0.232	0.707	0.112	0.252	
24	24	0.709	0.199	0.717	0.313	0.254	
25	25	0.662	0.232	0.636	0.425	0.253	
26	26	0.716	0.189	0.512	0.523	0.255	
27	27	0.717	0.221	0.536	0.645	0.255	
28	28	0.746	0.290	0.645	0.775	0.253	
29	29	0.726	0.273	0.599	0.838	0.254	
30	30	0.790	0.220	0.494	0.850	0.256	
31	31	0.016	0.283	0.784	0.993	0.617	
32	32	0.633	0.264	0.431	1.076	0.258	
33	33	0.607	0.269	0.367	1.086	0.260	
34	34	0.530	0.239	0.313	1.505	0.263	
35	35	0.500	0.271	0.338	2.060	0.263	
36	36	0.177	0.076	1.180	2.133	0.449	

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

The result is shown in **Table 1**. The selection of items included in the APM-SV scale followed the same pattern used by Arthur and Day with some slight modifications as indicated below. Arthur and Day (1994) published 12-item version dubbed Advanced Progressive Matrices-Short Form (APM-SF). Arthur and Day used items 1, 4, 8, 11, 15, 18, 21, 23, 25, 30, 31, and 35 from the 36-items of set 2 of the APM based on a set of three decision rules which they summed up as follows:

1. Dividing the APM into 12 sections with each section containing 3-items, based on difficulty.
2. Taking the item with the highest item-total correlation from each section. (Analysis was done using the CTT model).
3. And in the case of a tie, including the item that resulted in the largest drop in internal consistency if it was excluded from the full test.

In the case of this present study, the rules of selection were modified with one additional rule added as follows:

i. The APM was divided into 6 sections with each section containing 6-items, based on difficulty. The sections are: Section 1: items 1-6, Section 2: items 7-12, Section 3: items 13-18, Section 4: items 19-24, Section 5: items 25-30 and Section 6: items 31-36.

ii. The two items with the highest Pearson point-biserial correlation (*r-pbis*) were selected from each section: the following items were selected: **Section 1**: items 1 and 2. **Section 2:** items 7 and 12. **Section 3:** items 15 and 17. **Section 4:** items 21 and 23. **Section 5:** items 28 and 29. **Section 6:** items 31 and 35.

iii. One additional item with the highest Pearson point-biserial correlation (*r-pbis*) next to the two already selected in that section was chosen from **Sections 2, 4 and 6 leaving out sections 1, 3 and 5.** Based on this rule, the following items were selected: **Section 2** - item 10, **Section 4** - item 24, **Section 6** - item 33.

iv. And in the case of a tie, including the item that resulted in the largest drop in internal consistency provided it was not excluded from the full test. (Fortunately in this present study no item was excluded from the analysis as shown in Table 4.8 of the main report of this work. Therefore rule 4 did not apply in the case of this study).

Following the above rules, thus all selected items were based on their order of increasing difficulty and fair contributions to the item total correlation. ***Now the APM-SV has 15 items. The items are: 1, 2, 7, 10, 12, 15, 17, 21, 23, 24, 28, 29, 31, 33, and 35 from the 36-items of set 2 of the APM scale.*** In order to do a thorough work on this newly modified APM scale dubbed APM-SV scale, it was administered to a new set of randomly selected participants. A total of 1000 participants comprising 500 undergraduates and 500 senior secondary school students took the APM-SV test. Similar to the administration of the full APM scale (by full APM scale, the researchers mean APM scale with 36 items as against any of the short forms with less number of items), Set 1 items (set 1 contains 12 items) were used as practice test since they were new examinees totally different from those that took the APM test during the first exercise. The APM-SV scale now served as the main scale of the second part of this study. Each of the selected 15 items has eight response options from

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

which the participant is expected to select one correct option. The options selected by the participants to each item on the test were subjected to IRT analysis using the X-Calibre 4.2 software. The X-Calibre 4.2 IRT analysis was performed using the three different Parameter Logistic Models, that is, the 1- Parameter Logistic Model (1-PLM), the 2-Parameter Logistic Model (2-PLM) and the 3-Parameter Logistic Model (3-PLM). The other statistics performed using the X-Calibre software includes: The Differential Information Function (DIF), Test Information Function, Item-by-Item Analysis that included the Item characteristic curve, Group comparisons for gender, ethnicity, school categories and age. The SPSS was used to perform dimension reduction analysis, reliability analysis, correlation analysis, etc. Now the answers to nine research questions are hereby presented:

Q1: Which is the most suitable Item Response Theory (IRT) Parameter Logistics Model (PLM) for Raven's Advanced Progressive Matrices-Smart Version (APM-SV) in Nigeria?

Research Question 1

Table 2: 1-PLM Item Parameters for All Calibrated Items for APM-SV

Old No	Seq.	Item ID	P	R	a	Flag(s)
1	1	1	0.562	0.030	1.000	
2	2	2	0.471	0.199	1.000	
7	3	3	0.373	0.078	1.000	
10	4	4	0.321	0.075	1.000	
12	5	5	0.314	0.161	1.000	
15	6	6	0.307	0.123	1.000	
17	7	7	0.245	0.207	1.000	
21	8	8	0.208	0.079	1.000	
23	9	9	0.192	0.083	1.000	
24	10	10	0.158	0.130	1.000	
28	11	11	0.116	0.110	1.000	
29	12	12	0.066	0.134	1.000	
31	13	13	0.015	0.026	1.000	
33	14	14	0.016	0.130	1.000	
35	15	15	0.027	0.163	1.000	

As shown in **Table 2**, no item out of the 15 items of APM-SV was flagged under the 1-PLM. This is an indication of a perfect fit under the 1-Parameter Logistic IRT Model.

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

Table 3: 2-PLM Item Parameters for All Calibrated Items for APM-SV

Seq.	Item ID	P	R	a	b	Flag(s)
1	1	0.546	0.050	0.415	-0.322	
2	2	0.560	0.158	0.656	-0.524	
3	3	0.558	0.045	0.520	-0.549	
4	4	0.454	0.060	0.590	-1.002	
5	5	0.335	0.089	0.504	-0.442	
6	6	0.365	0.042	0.599	-1.264	
7	7	0.221	0.158	0.682	-0.305	
8	8	0.350	0.044	0.541	0.291	
9	9	0.234	0.035	0.450	0.371	
10	10	0.153	0.085	0.603	0.589	
11	11	0.130	0.049	0.484	0.316	
12	12	0.134	0.108	0.547	0.316	
13	13	0.130	0.129	0.615	1.406	
14	14	0.197	0.104	0.523	0.024	
15	15	0.114	0.136	0.566	0.086	

As shown in **Table 3**, no item out of the 15 items of APM-SV was flagged under the 1-PLM. This is an indication of a perfect fit under the 2-Parameter Logistic IRT Model. Consequently the result indicates a perfect fit since all the items fitted reasonably well under the 2-Parameter Logistic IRT Model.

Table 4: 3-PLM Item Parameters for All Calibrated Items for APM-SV

Seq.	Item ID	P	R	a	b	c	Flag(s)
1	1	0.546	0.159	0.640	-0.322	0.399	
2	2	0.560	0.158	0.756	-0.524	0.089	
3	3	0.358	0.045	0.663	-0.549	0.133	
4	4	0.254	0.060	0.772	-1.002	0.248	
5	5	0.335	0.089	0.607	-0.442	0.152	
6	6	0.365	0.042	0.746	-1.264	0.152	
7	7	0.221	0.158	0.834	-0.305	0.149	
8	8	0.250	0.044	0.674	0.291	0.053	
9	9	0.134	0.035	0.550	0.371	0.053	
10	10	0.153	0.085	0.752	0.589	0.051	
11	11	0.130	0.049	0.608	1.316	0.053	
12	12	0.134	0.108	0.677	1.316	0.052	
13	13	0.030	0.129	0.760	2.406	0.151	
14	14	0.097	0.104	0.704	0.6 24	0.055	
15	15	0.014	0.136	0.748	1.086	0.052	

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

As shown in **Table 4,** no item out of the 15 items of the APM-SV under the 3-PLM was also flagged either for F, K, La/b/c or Ha/b/c. Consequently the result indicates a perfect fit since all the items fit the 3-Parameter Logistic IRT Model. Thus within the framework of IRT, the 1-Parameter Logistic Model, 2-Parameter Logistic Model and the 3-Parameter Logistic Model all showed perfect fit statistics. Therefore the 1, 2 and 3-Parameter Logistic IRT Models are all suitable for examining the Advance Progressive Matrices-Smart Version (APM-SV) scale.

Research Question 2

What is the Overall Model Fit of APM-SV using Nigerian validation sample?

Table 5: Overall Model Fit

Test	Items	Chi-square	df	p	-2LL
Full Test	15	1467.487	504	0.000	15613

TABLE 5 shows the Overall Model Fit with a Chi-Square value of 1467.487, a degree of freedom (df) of 504, a probability of 0.000 and -2 logistic likelihood of 15613. To further appreciate the Overall Model fit, the distribution of the theta estimates for all calibrated items, frequency distribution for the theta estimates, the distribution of the a-, b- and c-parameters and their frequency distributions are presented below: **Figure 1** displays the distribution of the theta estimates for all calibrated items.

Figure 1: Theta Estimates for All Calibrated Items

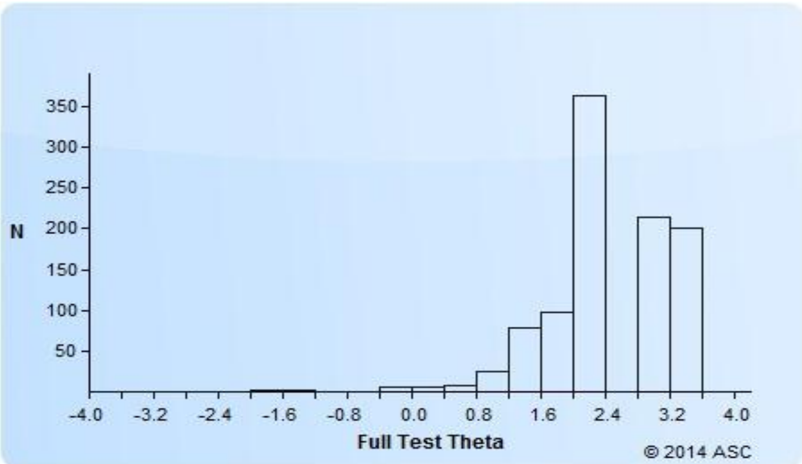


Figure 2 displays the distribution of the b parameters.

Figure 2 : Histogram of the b Parameters

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

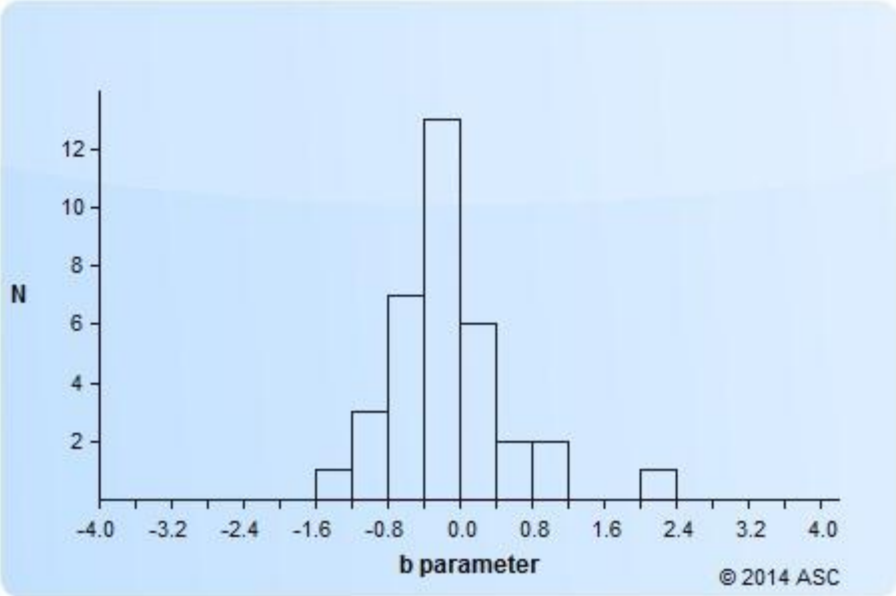


Figure 3 displays the joint distribution of the b parameter by Theta.

Figure 3: b parameter by Theta

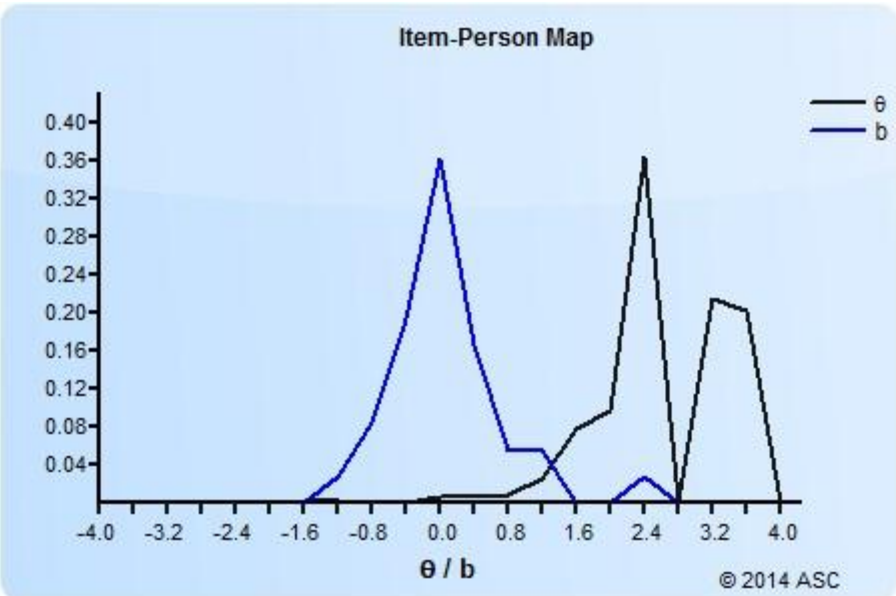


Figure 4 below displays a graph of the Test Response Function (TRF) for all calibrated items. The TRF predicts the proportion or number of items that an examinee would answer correctly as a function of theta. The left Y-axis is in proportion correct units while the right Y-axis is in number-correct units.

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

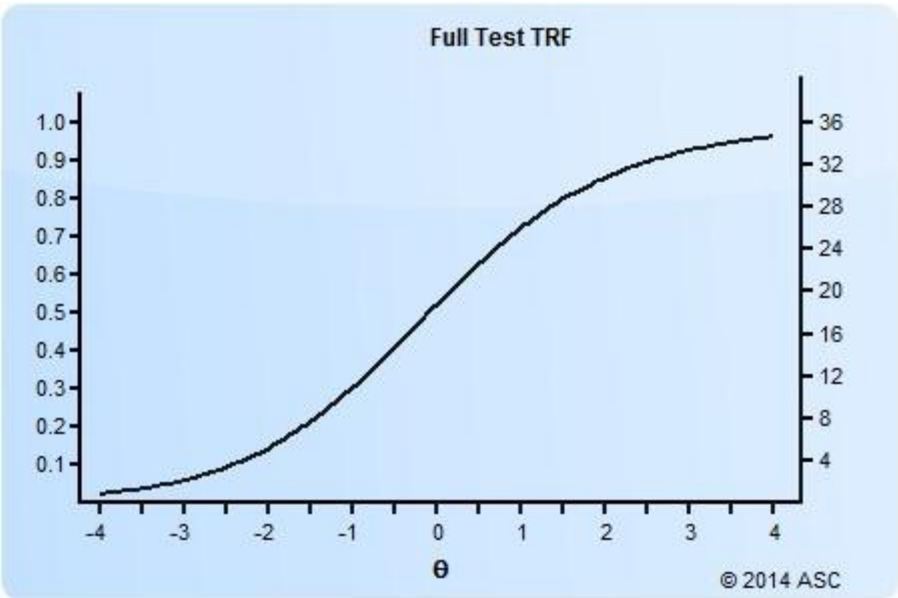
Research Question 3

What is the person separation reliability of APM-SV that can be inferred from the contribution of each of the items to the Test Response Function (TRF)?

Inferred reliability of APM from Test Response Function

The focus of IRT is the contribution of each item to the overall fit of any given instrument. Therefore discussion on reliability of instrument is usually inferred from the Test Response Function (TRF) since the concept of test response function is analogous to the concept of reliability in Classical Test Theory. The TRF is pictorially displayed in **Figure 4**. Reliability in this case is conceived as the person separation reliability or item separation reliability. The person separation reliability is analogous to Cronbach's α. This is the degree to which the APM scale differentiates persons in the test's outcome. The range of course is 0 – 1.

Figure 4 : Test Response Function



Item separation reliability on the other hand is the degree to which item difficulties are differentiated. Again the range of course is 0 – 1. Now due to the sophistication of the X-Calibre software, a power tool designed for analysing IRT, it is now possible to estimate the reliability of an instrument under the IRT models.

Table 6: Summary Statistics for the Total Scores indicating person separation reliability of APM-SV

Test	Items	Alpha	Mean	SD	Skew	Min	Q1	Median	Q3	Max	IQR
Full Test	15	0.920	8.17	2.696	-3.082	1	3.00	8	13.00	15	6.00

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

Table 6 shows the Alpha value of the APM-SV test. The Alpha value is 0.920 which tends towards 1 and it indicates a strong reliability. Therefore the preponderance of APM-SV scale's reliability within the framework of IRT as indicated by the Alpha value is 0.92.

Figure 5 below displays a graph of the Test Information Function for all calibrated items. The TIF is a graphical representation of how much information the test is providing at each level of theta. Maximum information was 8.090 at theta = -0.200.

Figure 5: Test Information Function

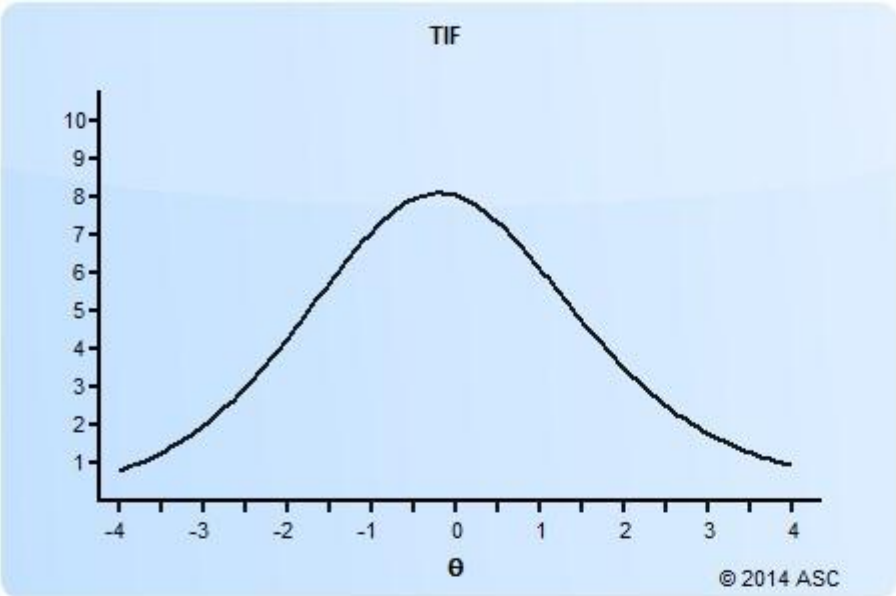
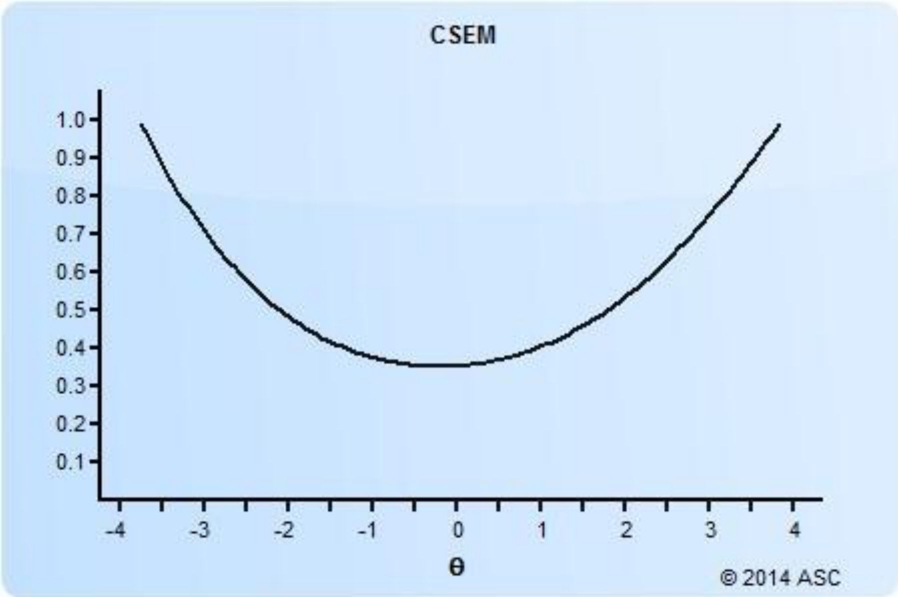


Figure 6 displays a graph of the Conditional Standard Error of Measurement (CSEM) Function. The CSEM is an inverted function of the TIF, and estimates the amount of error in theta estimation for each level of theta. The minimum CSEM was 0.352 at theta = -0.200.

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

Figure 6: CSEM Function



Research Question 4

What is the Item Response Function (item-by item) evidence of reliability of APM-SV in Nigeria?

Item Response Function (item-by- item) reliability of APM-SV

The item-by-item results of the analysis shows that each scored item has four tables and a plot of the item response function (IRF). The item-by-item analysis report is a sequel to the Test Information Function (TIF). The red line (fit line) represents the observed proportion correct conditional on theta. In almost all the items of the APM-SV scale, there were no large deviations of the red line from the IRF which are suggestive of good item fit. Thus, the fit line further identifies why and how the particular item fits the chosen 3-Parameter Logistic IRT model. There are four tables presented for each item by the X-Calibre analysis, but due to secrecy and confidentiality of the items, only those for the 1st, 18th and 36th items are presented here for illustration.

1. Item information table: records the information supplied by the control file (or Classic Data Header) for this item.

2. Classical statistics table: classical statistics for the item.

3. IRT parameters table: item parameter estimates for the item.

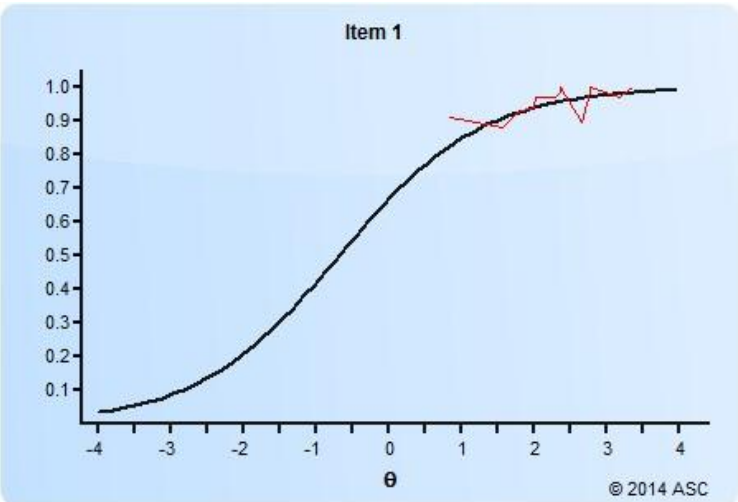
4. Option/Category statistics: detailed statistics for each item, which helps diagnose issues in items with poor statistics. In the case of the APM-SV, there are no items with poor statistics under the chosen 3-PLM.

The classical statistics presents classical summary statistics for the item. For multiple choice items instrument like the instrument of this study: APM-SV, the P value and the point-biserial

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

correlations are presented in the first three columns of the table. The P value is the proportion of examinees that answered an item in the keyed direction and 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 Alpha w/o is Cronbach's alpha computed with the current item excluded. The item-total correlation is a measure of the discriminating power of the item and is related to the IRT discrimination parameter. The IRT parameters table presents the IRT item parameters and the fit statistics. The latent trait theta is expressed on a standardized scale, so a one unit change equals a one standard deviation change. The "a" parameter indexes the discrimination of the item, as larger values for "a" will result in a greater steepness of the slope of the IRF or Item Characteristic Curve (ICC) and indicate the item differentiates examinees well. The "b" parameter is the item difficulty parameter and equals the location on the theta continuum where the probability of a correct response equals .50 + (c/2). It follows that multiple choice items with more positive "b" parameters are more difficult for examinees, as a higher trait level is required to endorse the keyed response 50% of the time. The "c" parameter equals the probability of an examinee of infinitely low theta obtaining a correct response due to guessing. Thus, "c" is also the lower asymptote of the IRF or ICC. The standard errors (SE) for each item parameter estimate are also presented in the item parameter table. A large SE for an item parameter (compared to the other items) indicates that the item parameter was poorly estimated. The IRT standardized (z) residual is the last entry in the item parameter table. It indexes the fit of the data to the Item Response Function. 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. The chi-square fit statistic and its degrees of freedom are reported for each item.

Presented below are three examples of the item-by-item report.



Item information

Seq.	*ID*	*Model*	*Scored*	*Num Options*	*Domain*	*Flags*
1	1	1PL	Yes	8	1	

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

Classical statistics

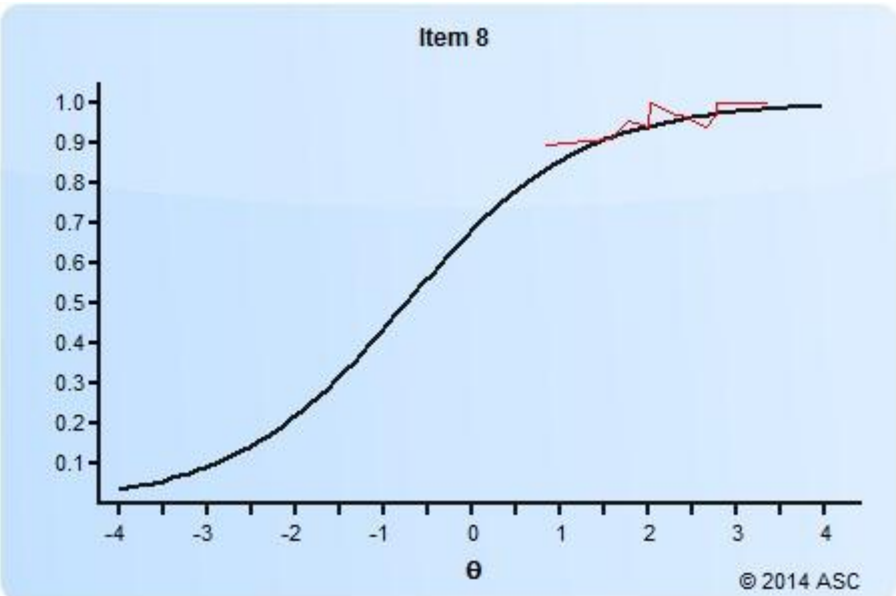
N	*P*	*S-Rpbis*	*T-Rpbis*	*Alpha w/o*
1000	0.962	0.030	0.119	0.672

IRT parameters

a	*b*	*a SE*	*b SE*	*Chi-sq*	*df*	*p*	*z Resid*	*p*	*InMSQ*	*InZstd*	*OutMSQ*	*OutZstd*
1.000	-0.622	0.053	0.144	26.892	14	0.020	2.639	0.008	0.789	-1.862	0.736	-1.536

Option statistics

Option	*N*	*Prop.*	*S-Rpbis*	*T-Rpbis*	*Mean*	*SD*	
A	10	0.010	0.018	-0.031	2.274	0.628	
B	5	0.005	-0.017	-0.053	1.906	0.521	
C	1	0.001	-0.022	-0.036	1.563	0.000	
D	14	0.014	-0.030	-0.081	1.964	0.706	
E	962	0.962	0.030	0.119	2.560	0.851	
F	2	0.002	-0.064	-0.068	1.246	0.775	
G	4	0.004	-0.014	-0.048	1.902	0.437	
H	2	0.002	0.036	0.019	2.901	0.720	
Omit	0						
Not Admin	0						



ISSN 2055-0154(Print), ISSN 2055-0162(Online)

Item information

Seq.	*ID*	*Model*	*Scored*	*Num Options*	*Domain*	*Flags*
8	8	1PL	Yes	8	1	

Classical statistics

N	*P*	*S-Rpbis*	*T-Rpbis*	*Alpha w/o*
1000	0.968	0.079	0.142	0.670

IRT parameters

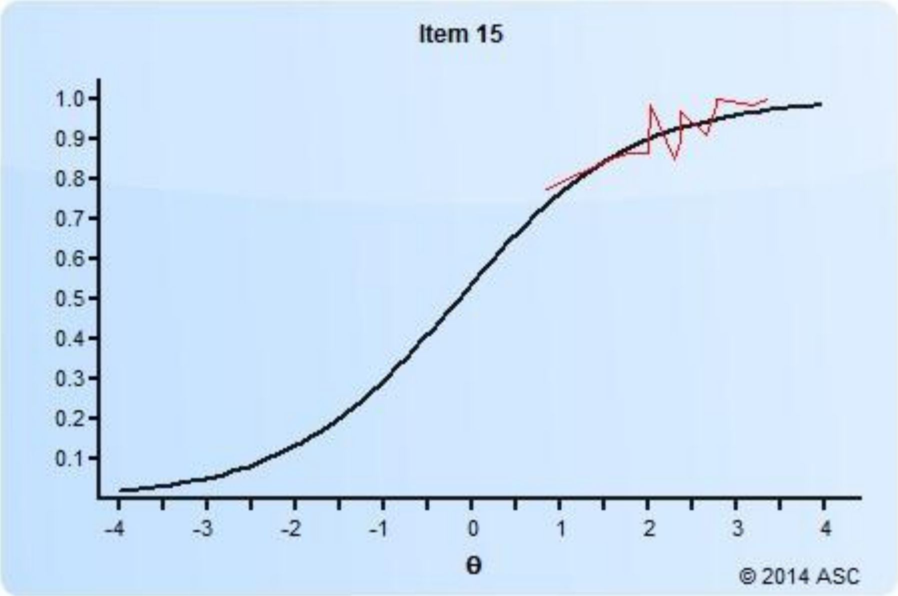
a	*b*	*a SE*	*b SE*	*Chi-sq*	*df*	*p*	*z Resid*	*p*	*InMSQ*	*InZstd*	*OutMSQ*	*OutZstd*
1.000	-0.691	0.053	0.148	16.189	14	0.302	2.798	0.005	0.694	-2.724	0.581	-2.557

Option statistics

Option	*N*	*Prop.*	*S-Rpbis*	*T-Rpbis*	*Mean*	*SD*	
A	968	0.968	0.079	0.142	2.562	0.845	
B	15	0.015	-0.013	-0.074	2.033	0.577	
C	3	0.003	-0.044	-0.066	1.510	0.314	
D	9	0.009	-0.100	-0.102	1.630	1.110	
E	0	0.000	--	--	--	--	
F	3	0.003	-0.017	-0.041	1.912	0.788	
G	1	0.001	-0.010	-0.028	1.794	0.000	
H	1	0.001	0.026	0.010	2.814	0.000	
Omit	0						
Not Admin	0						

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

European Journal of Statistics and Probability
Vol.4, No.3, pp.20-60, July 2016
Published by European Centre for Research Training and Development UK (www.eajournals.org)



Item information

Seq.	*ID*	*Model*	*Scored*	*Num Options*	*Domain*	*Flags*
15	15	1PL	Yes	8	1	

Classical statistics

N	*P*	*S-Rpbis*	*T-Rpbis*	*Alpha w/o*
1000	0.927	0.163	0.238	0.665

IRT parameters

a	*b*	*a SE*	*b SE*	*Chi-sq*	*df*	*p*	*z Resid*	*p*	*InMSQ*	*InZstd*	*OutMSQ*	*OutZstd*
1.000	-0.086	0.051	0.117	26.462	14	0.023	2.962	0.003	0.853	-1.687	0.718	-2.232

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

Option statistics

Option	*N*	*Prop.*	*S-Rpbis*	*T-Rpbis*	*Mean*	*SD*	
A	48	0.048	-0.073	-0.165	1.917	0.641	
B	927	0.927	0.163	0.238	2.597	0.823	
C	15	0.015	-0.196	-0.158	1.450	1.449	
D	6	0.006	-0.006	-0.048	2.011	0.392	
E	0	0.000	--	--	--	--	
F	0	0.000	--	--	--	--	
G	2	0.002	0.019	-0.008	2.392	0.000	
H	0	0.000	--	--	--	--	
Omit	2	0.002	-0.074	-0.076	1.089	0.120	
Not Admin	0						

Research Question 5

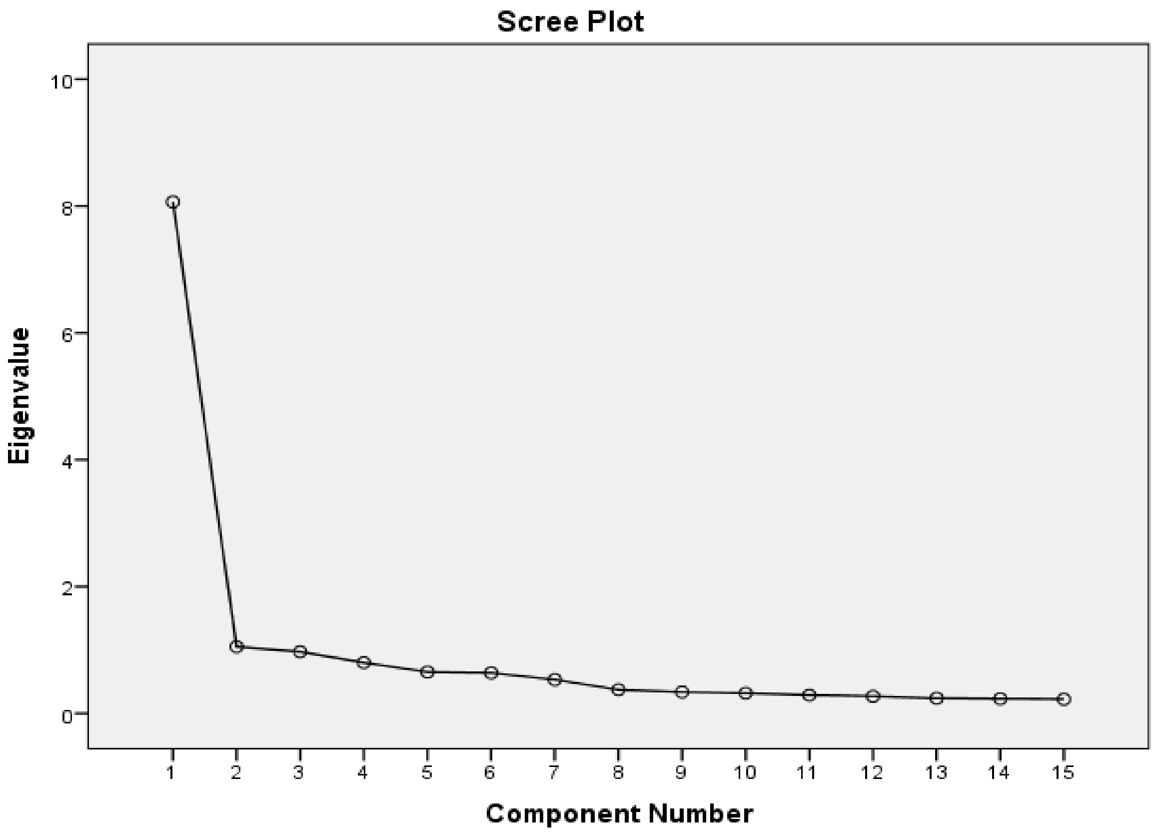
What is the evidence of unidimensionality, if any, of APM-SV in Nigeria?

Unidimensionality evidence of APM-SV Sclae in Nigeria

Unidimensional models require a single trait (ability or domain) dimension, $\boldsymbol{\theta}$. (Table 4.8 contained in the full report of this work), part of the X-Calibre report indicated that the scale essentially examined one trait, construct or domain. However in order to extrapolate the unidimensionality of the scale, factor analysis was performed for the scores generated from the APM-SV test administered to the 1000 university undergraduates and senior secondary school students. Field (2005) wrote that "factors or traits or underlining constructs can be extrapolated or established through the use of eigenvalues and variance, scree plot and communalities." Georgiev (2008), Morsanyi, Primi, Handley (2009), Raven (2000), Van der Ven and Ellis (2000), Raven, Raven and Court (1997) as well as the WPS (2015) stressed the need for extrapolation of unidimensionality of an instrument that is indeed measuring only one domain or construct; or even more than one factor with the use of eigenvalues. Using Guttman-Kaiser rule, "all factors with eigenvalues greater than 1 should be retained as the factors that the scale measures." Guttman-Kaiser also suggested that "factors which account for 70% and above of the variance should be accepted as the underlining construct." Analysis of the scree plot is another way to determine the underlining construct or unidimensionality of a scale. The rule of thumb in analysing the scree plot is very simple. Traits or constructs or factors before the breaking point or elbow joint in the scree plot graph is assumed to be the main construct under examination. Furthermore it is also important to check the communalities after construct extraction. If the communalities are low, the extracted constructs account for only a little part of the variance, and therefore more constructs might be deemed to be in view which might provide better account for the total variance. Dimension reduction analysis was utilized to determine significant unidimensionality extraction at greater than 0.50. The choice of 0.50 was made by the researcher because according to Thomson (2004) "determining the number of factors or construct to be extracted or extrapolated requires judgment." In this analysis, promax rotation was utilized to maximize the establishment of the construct under examination. The choice of promax rotation was made because orthogonality is not assumed in this case and

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

therefore the items of the construct to be examined are expected to correlate. A careful examination of the scree plot shown below shows that there is only one construct before the breaking point or elbow joint. This therefore succinctly shows the unidimensionality of the underlining construct of the APM-SV scale, namely fluid ability. All the 15 items measure one construct, the intelligence of the test taker. From the table of communalities provided by the SPSS analysis result, the main construct measured by APM-SV scale which is intelligence explains 8.17 or 81.754% of the total variance. This is incontrovertibly a sizable chunk of the model. Therefore the underlining construct is effectively examined by the scale and it ensures its unidimensionality. Since the assumption of unidimensionality is met by this model, it invariably means that local independence holds



Research Question 6

What is the b-parameter index (item difficulty parameter) for each APM-SV item in Nigeria?

Item Difficulty Parameter (b-parameter) of APM-SV

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

Table 7: Item Parameters for All Calibrated Items

Seq.	Item ID	P	R	a	b	c	Flag(s)
1	1	0.546	0.159	0.640	-0.322	0.399	
2	2	0.560	0.158	0.756	-0.524	0.089	
3	3	0.358	0.045	0.663	-0.549	0.133	
4	4	0.254	0.060	0.772	-1.002	0.248	
5	5	0.335	0.089	0.607	-0.442	0.152	
6	6	0.365	0.042	0.746	-1.264	0.152	
7	7	0.221	0.158	0.834	-0.305	0.149	
8	8	0.250	0.044	0.674	0.291	0.053	
9	9	0.134	0.035	0.550	0.371	0.053	
10	10	0.153	0.085	0.752	0.589	0.051	
11	11	0.130	0.049	0.608	1.316	0.053	
12	12	0.134	0.108	0.677	1.316	0.052	
13	13	0.030	0.129	0.760	4.406	0.151	
14	14	0.097	0.104	0.704	2.624	0.055	
15	15	0.014	0.136	0.748	2.086	0.052	

The answer to the sixth research question is glaringly obvious in **Table 3** (already presented above). The *b*-parameter is the item difficulty parameter and indicates the location on the theta (***θ***) continuum where the probability of a correct response equals *c*/2 + .50. Thus, the *b*-parameter is the centre of the IRF and is where the slope steeps most to show the discriminating power of the item maximally. Since the APM scale is centred on the examinees drawn from the university undergraduates and senior secondary school students, the *b* parameter shows the examinee's ***θ*** value for which the item is appropriate. Higher *b*-parameters (> 1.0) indicate that the item is more difficult; a value below -1.0 indicates that the item is very easy. The purpose of the APM test is the measure of fluid ability, the test is conceived and designed by its constructor in such a way that the difficulty index graduates from very easy item to very difficult item. According to X-Calibre manual, the difficulty index "ranges in theory from negative to positive infinity, but in practice from -3.0 (very easy) to +3.0 (very difficult)." A careful examination of the *b* parameter column in Table 7 displayed above shows that the values of *b* for item 1 is -0.322, item 2 is -0.524, items 14 and 15 have *b* values of -2.624 and 2.086 respectively. Thus the *b* parameter kept graduating in difficulty just like its parent test-the APM. The *b* parameter is related to the classical *P* statistic, as items with low *P* values will tend to have higher (more positive) *b* parameters and items with high *P* values will tend to have lower (more negative) *b* parameters.

Research Question 7

What is the range of Differential Item Function (discriminatory index) popularly referred to in IRT as a-parameter of the APM-SV in the country?

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

Differential Item Function (Discriminatory Index) or *a*-parameter of APM-SV in Nigeria

Differential item functioning (DIF) occurs when the performance of an item differs across groups of examinees with equal latent trait as an evidence of item bias which leads to test bias. In this study, the university undergraduates and senior secondary school students' responses to the APM-SV test were examined for DIF across gender (i.e. males and females), age (i.e. university undergraduates Vs senior secondary school students), and ethnic groups in Nigeria (i.e. Hausa, Igbo, Yoruba and Minority). The goal of this analysis was to flag items that are potentially biased against one group in favour of another. The X-Calibre's Mantel-Haenszel statistical analysis tool, where each group is split into several ability levels, and the probability of a correct response compared between the groups for each level was applied to the results of the respondents to the APM scale test. The Mantel-Haenszel (M-H) coefficient is reported for each item as an odds ratio. The coefficient is a weighted average of the odds ratios for each θ level. If the odds ratio is less than 1.0, then the item is more likely to be correctly endorsed by one group than the other group(s). Likewise, odds ratios greater than 1.0 indicate that one group was more likely to correctly endorse the item than other group(s). According to Brouwers, Van de Vijver, & Van Hhemert (2009), 'the M-H coefficient is standardized through a log transformation, which is referred to as M-H DIF. The transformed value less than 0 indicates a reference group advantage whereas a value greater than 0 indicates the item is more likely to be correctly endorsed by a particular group than the other group or groups.' These ratios were used to determine if the DIF present in the responses to the APM scale was constant for all abilities (uniform DIF) or varied conditional on θ (crossing DIF). The M-H coefficient is not sensitive to crossing DIF, so null results were checked to confirm that crossing DIF was present or not present (Dorans & Holland 1993). Subsequently the X-Calibre ***z-test Statistic*** was also applied so that the negative of the natural logarithm of the M-H odds ratio was divided by its standard error to obtain the *z*-test statistic used to test the significance of the M-H against a null of zero DIF (odds ratio of 1.0). The two-tailed *p* value associated with the *z* test for DIF was then prorated. Items with *p* values less than .05 were flagged as having significant DIF. Thus the group that the item or items of the scale is/are ***Bias Against*** are flagged. This then is the group the item or items is/are disfavouring, or "biased against" when the *p* value is less than .05. In the context of the M-H test for DIF, the group that the item is disfavouring has a lower probability of a correct response than the other group, controlling for ability level. Below are the results of the analysis for each of the identified groups:

Gender: Males and Females

Table 8: Subgroup statistics for the Full Test

Subgroup	Examinees	Mean Theta	SD Theta
MALES	540	-1.276	0.369
FEMALES	460	-1.291	0.342

Table 8 shows the gender subgroup statistics for the APM-SV test. The Meanθ and SDθ values for male with a total number of 540 are -1.276 and 0.369 respectively, while the Meanθ and SDθ values for female with a total number of 460 are -1.291and 0.472 respectively. Again as in the full APM-SV scale test, both values are not far apart indicating no element of bias towards any gender. The APM-SV scale was equally weighted and rated for both genders.

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

Table 9: Mantel-Haenszel's Item Parameters for All Calibrated Items for 2 Groups: Male Vs Female

Seq.	Item ID	P	R	a	b	c	Flag(s)
1	1	0.546	0.159	0.340	4.000	0.039	
2	2	0.560	0.158	0.456	4.000	0.049	
3	3	0.358	0.045	0.363	4.000	0.013	
4	4	0.254	0.060	0.472	4.000	0.048	
5	5	0.335	0.089	0.307	4.000	0.052	
6	6	0.365	0.042	0.446	4.000	0.052	
7	7	0.221	0.158	0.534	4.000	0.049	
8	8	0.250	0.044	0.374	4.000	0.015	
9	9	0.134	0.035	0.250	4.000	0.015	
10	10	0.153	0.085	0.452	4.000	0.051	
11	11	0.130	0.049	0.308	4.000	0.053	
12	12	0.134	0.108	0.377	4.000	0.032	
13	13	0.030	0.129	0.460	4.000	0.131	
14	14	0.097	0.104	0.404	4.000	0.025	
15	15	0.014	0.136	0.448	4.000	0.012	

Again, as shown in Table 9 the Mantel-Haenszel's item parameter for all calibrated items assigned equal *b* parameter values to all items of the APM-SV scale from the analysis of the scores generated from both genders. Similarly, the discriminatory, *a*, parameter did not show much discrepancies. All the *a parameters* ranged from 0.250 to 0.472. This is an indication that the APM-SV did not discriminate the test outcome in terms of group. These conclusively show that the items of the APM-SV were bias free towards the gender groups of males and females. This conclusively shows that the items of the APM-SV were bias free towards gender. In other words the items are gender blind. The items are not biased towards any gender.

Age: (University Undergraduates Vs Senior Secondary School Students)

Table 10: Subgroup statistics for the Full Test

Subgroup	Examinees	Mean Theta	SD Theta
UG	500	-0.541	0.105
SS	500	-0.539	0.103

Table 10 above shows the school category subgroup statistics for the APM-SV test. The Meanθ and SDθ values for undergraduates (UG) with a total number of 500 are -0.541 and 0.105 respectively, while the Meanθ and SDθ values for senior secondary (SS) with a total number of 500 are -0.539 and 0.103 respectively. Both values are not far apart indicating no element

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

of bias towards any age. The APM-SV scale was equally weighted and rated for the two age groups and categories.

Table 11: Mantel-Haenszel's Item Parameters for All Calibrated Items for 2 Groups: Male Vs Female

Seq.	Item ID	P	R	a	b	c	Flag(s)
1	1	0.546	0.159	0.240	4.000	0.039	
2	2	0.560	0.158	0.256	4.000	0.049	
3	3	0.358	0.045	0.263	4.000	0.013	
4	4	0.254	0.060	0.272	4.000	0.048	
5	5	0.335	0.089	0.207	4.000	0.052	
6	6	0.365	0.042	0.346	4.000	0.052	
7	7	0.221	0.158	0.434	4.000	0.049	
8	8	0.250	0.044	0.274	4.000	0.015	
9	9	0.134	0.035	0.150	4.000	0.015	
10	10	0.153	0.085	0.352	4.000	0.051	
11	11	0.130	0.049	0.208	4.000	0.053	
12	12	0.134	0.108	0.277	4.000	0.032	
13	13	0.030	0.129	0.360	4.000	0.131	
14	14	0.097	0.104	0.304	4.000	0.025	
15	15	0.014	0.136	0.348	4.000	0.012	

Again, as shown in **Table 11**, the Mantel-Haenszel's item parameter for all calibrated items assigned equal *b* parameter values to all items of the APM-SV scale from the analysis of the scores generated from both age groups. Similarly, the discriminatory, *a*, parameter did not show much discrepancies. All the *a* parameter had values that ranged from 0.150 to 0.434. This is an indication that the APM-SV did not discriminate the test outcome in terms of group. These conclusively show that the items of the APM-SV scale were bias free towards the age groups of undergraduates of ages 16 to 40 years and senior secondary school students of ages 11 to 20 years. The items are not biased towards any group on the account of age.

Ethnic Groups: Hausa, Igbo, Yoruba and Minority

Table 12: Subgroup statistics for the Full Test

Subgroup	Examinees	Mean Theta	SD Theta
HAUSA	210	-1.176	0.109
IGBO	270	-1.391	0.102
YORUBA	180	-1.131	0.097
MINORITY	340	-1.487	0.267

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

Table 12 shows the school category subgroup statistics for the APM-SV test. The Meanθ and SDθ values for Hausa with a total number of 210 participants are -1.176 and 0.109 respectively. The Meanθ and SDθ values for Igbo with a total number of 270 are -1.291 and 0.102 respectively. The Meanθ and SDθ values for Yoruba with a total number of 180 participants are -1.131 and 0.097 respectively. The Meanθ and SDθ values for Minority with a total number of 340 participants are -1.487 and 0.267 respectively. What can be reasonably deduced from the above iterations is that there is a close parity between the values of Meanθ and SDθ and therefore the claim of bias towards any of the ethnic groups is not sustainable. The APM-SV scale was equally weighted and rated for all the ethnic groups.

Table 13: Mantel-Haenszel's Item Parameters for All Calibrated Items for 4 Groups: Hau, Igb, Yor & Min

Seq.	Item ID	P	R	a	b	c	Flag(s)
1	1	0.002	0.008	0.294	4.000	0.041	
2	2	0.001	0.021	0.394	4.000	0.041	
3	3	0.003	0.012	0.254	4.000	0.042	
4	4	0.021	-0.050	0.291	4.000	0.052	
5	5	0.002	-0.003	0.181	4.000	0.046	
6	6	0.010	0.007	0.174	4.000	0.056	
7	7	0.004	0.031	0.174	4.000	0.054	
8	8	0.010	0.010	0.170	4.000	0.046	
9	9	0.002	-0.003	0.173	4.000	0.031	
10	10	0.009	0.004	0.169	4.000	0.036	
11	11	0.008	0.036	0.170	4.000	0.025	
12	12	0.003	-0.013	0.171	4.000	0.022	
13	13	0.037	0.109	0.156	4.000	0.065	
14	14	0.050	0.075	0.152	4.000	0.054	
15	15	0.001	0.040	0.172	4.000	0.031	

Again, as shown above in **Table 13**, the Mantel-Haenszel's item parameter for all calibrated items assigned equal *b* parameter values to all items of the APM-SV scale from the analysis of the scores generated from all ethnic groups. Similarly, the discriminatory, *a*, parameter did not show much discrepancies. All the *a parameters* ranged from 0.152 to 0.394. This is an indication that the APM-SV did not discriminate the test outcome in terms of ethnic groups. These conclusively show that the items of the APM-SV were bias free towards the ethnic groups in Nigeria: Hausa, Igbo, Yoruba and Minority. The items are not biased towards any group on the account of ethnicity or culture.

Research Question 8

What is the c-parameter (probability of guessing) for each APM-SV item as evidence of bias culturally (ethnic group, school type/age, and sex)?

Probability of Guessing (c-parameter) of APM-SV items in Nigeria

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

Table 14 : Item Parameters for All Calibrated Items

Seq.	Item ID	P	R	a	b	c	Flag(s)
1	1	0.546	0.159	0.640	-0.322	0.399	
2	2	0.560	0.158	0.756	-0.524	0.089	
3	3	0.358	0.045	0.663	-0.549	0.133	
4	4	0.254	0.060	0.772	-1.002	0.248	
5	5	0.335	0.089	0.607	-0.442	0.152	
6	6	0.365	0.042	0.746	-1.264	0.152	
7	7	0.221	0.158	0.834	-0.305	0.149	
8	8	0.250	0.044	0.674	0.291	0.053	
9	9	0.134	0.035	0.550	0.371	0.053	
10	10	0.153	0.085	0.752	0.589	0.051	
11	11	0.130	0.049	0.608	1.316	0.053	
12	12	0.134	0.108	0.677	1.316	0.052	
13	13	0.030	0.129	0.760	4.406	0.151	
14	14	0.097	0.104	0.704	0.6 24	0.055	
15	15	0.014	0.136	0.748	1.086	0.052	

The *c* parameter equals the probability of an examinee of infinitely low θ obtaining a correct response due to guessing. Thus, *c* is also the lower asymptote of the IRF. The inclusion of a non-zero *c* parameter affects the location of *a* and *b* on the θ scale. The *c* parameter is expected to equal approximately 1 divided by the number of alternatives for multiple-choice tests. Therefore, for the APM that has 8 alternatives, a low examinee should have 1/8 = 0.125 chance of guessing the correct answer. Since *c* = 0.125 for this 8-alternative item, once the right key is isolated, the examinees will be guessing among the remaining seven options. Therefore where guessing is pronounced, the value of *c* will be much lower than 0.125. Higher value will mean that guessing is not strongly evidenced. From Table 14 already presented above, *c parameter* values range from 0.051 to 0.399 indicating limited or no guessing. Therefore the degree of guessing can be said to be low amongst the undergraduates as well as the senior secondary school students in Nigeria.

The Advanced Progressive Matrices-Smart Version (Apm-Sv) and the Correlates of Fluid Measures

In order to address the remaining research question of this study, the various instruments correlated with the Raven's Advanced Progressive Matrices Smart Version (APM-SV) scale, a modified version of the original Advanced Progressive Matrices-the full test, were administered alongside with the APM-SV scale. Samples of the instruments are provided in the appendixes.

The instruments are

1. Digital Span scale of the WAIS-R (Wechsler, 2006). The Digit Span scale of the WISC–R was used to measure working memory. This task was chosen following Ackerman, Beier & Boyle, (2005) and Colom, Florez-Mendoza & Rebollo, (2003), who

employed a simple digit span to assess the relationship between fluid intelligence and working memory. Participants were given sequences of numbers and were asked to recall them forward and backward. The series began with two digits and kept increasing in length, with two trials for each level.

2. Heuristics and biases literature tasks (Gilovich, Graffin, & Kahneman, 2002).
3. Mathematics Ability Test. (Redman, A. 2012).
4. Bennett Mechanical Comprehension Test (BMCT; Bennett, 1969).
5. Culture Fair Intelligence Test (CFIT) developed by R. B. Cattel.
6. Advanced Progressive Matrices (APM) full test.

Each participant was asked to respond to two instruments. In other words, respondents were asked to take two tests. The APM-SV was compulsory for all participants. Then each participant was required to do a second test from any of the six other tests (see the above for the list of the tests) administered to them. A total of 1000 respondents made up of 500 university undergraduate and 500 senior secondary school students took part in these examinations. They were 540 males and 460 females. Out of the 1000 participants, 301 took the Digit Span scale of the WISC–R, 250 responded to the heuristics and biases literature tasks, 62 responded to Alan Redman's Mathematics Ability Test, 67 took the Bennett Mechanical Comprehension Test (BMCT), 196 took the Culture Fair Intelligence Test (CFIT) and 224 took the Advanced Progressive Matrices (APM) full test. The correlation analyses were performed using the SPSS.

The results of the analyses were applied to treat the remaining research questions.

Research Question 9

What are the Correlations between APM-SV scale and other measures of intelligence such as working memory measures, Mechanical Reasoning ability, Mathematical Reasoning ability, Probabilistic Reasoning ability, APM and CFIT?

Research Question 9(i)

Table 15: APMSV and APM Correlation

		N	Correlation	Sig.
Pair 1	APMSV & APM	224	.960	.000

As can be seen from the **Table 15** above the correlation between the modified version of the APM scale dubbed APM-SV and the full APM scale recorded a positive correlation of 0.96. This is a testimony to the fact that both instrument are strongly related.

Research Question 9(ii)

Table 16: APMSV and Digit Span Correlation

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

		N	Correlation	Sig.
Pair 1	APMSV & DIGITSPAN	150	.742	.000

There exist a positive moderately strong correlation between working memory measure-the Digit Span scale of the WISC–R. The correlation coefficient of 0.74 was the outcome of the analysis of the scores generated for this purpose. The result is displayed in Table 35 above.

Research Question 9(iii)

Table 17: APMSV and Mechanical Comprehension Correlation

		N	Correlation	Sig.
Pair 1	APMSV4 & MECH	62	.590	.001

There is also a noticeable relationship that is statistically significant between the APM-SV scale and Bennett Mechanical Comprehension Test (BMCT). A correlation value of 0.60 was the resultant computation of the data thereto. The value is displayed in **Table 17** above.

Research Question 9(iv)

Table 18: Paired Samples Correlation

		N	Correlation	Sig.
Pair 1	APMSV & MATHS	67	.842	.000

The APM-SV scale showed a very strong positive relationship with Mathematical Reasoning ability. The correlation value of 0.84 is displayed in **Table 18** above.

Research Question 9(v)

Table 19: APMSV and Heuristic Probabilistic Reasoning test Correlation

		N	Correlation	Sig.
Pair 1	APMSV & HEURISTIC	196	.037	.598

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

Off all the correlations performed in this exercise with the APM-SV, the relationship between the APM-SV and Probabilistic Reasoning was found to be the least in strength. The correlation recorded a small but positive value of 0.037. The figure is shown in **Table 19**

Research Question 9(vi)

Table 20: APMSV and Culture Fair Intelligence Test Correlation

	N	Correlation	Sig.
Pair 1 APMSV & CFIT	224	.860	.000

The culture fair intelligence test, which in itself is a measure of fluid ability and has established a strong correlation with academic achievement correlated very strongly with the modified APM, the APM-SV. It recorded a correlation value of 0.86. The result is displayed above in **Table 20**. This is very significant for this study. It is interesting to recall that CFIT scale has been normed and standardized in Nigeria. The APM itself is also a measure of fluid ability and invariably a correlate of academic achievement.

CONTRIBUTIONS

The Raven's Advanced Progressive Matrices (APM) scale, even though constructed in 1936 by John C. Raven, is not yet a very popular scale in Africa. Therefore it is not commonly employed for research works in the area of intelligence investigation in the African continent, particularly Nigeria, whereas the APM instrument is reputed to be in high demand in America, Europe and Asia. The APM scale is internationally well acclaimed, well known and highly patronized in research work across the globe. Evers (2011) had attested to the fact that "the Raven's Matrices are in the fourth position among the ten most used tests in Europe." According to him "they are widely employed to assess fluid ability in adolescents and adults and have also been recommended as a useful measure for identifying academic potentials." From the review of literature, it was only in 2011 and more recently, in 2012, that Chiesi, Ciancaleoni, Galli, Morsanyi & Primi (2011 and 2012) "applied for the first time the IRT models in investigating the Arthur and Day's Advanced Progressive Matrices-Short Form (APM-SF) Scale." Both investigations (in 2011 and 2012) were carried out by them in and limited to Europe. There were no evidences in their reports to show that samples were drawn from any part of Africa. Therefore, this current investigation may be the first of its kind anywhere in Africa. This current study utilized the IRT to analyse the APM and IRT for the analysis of the APM-SV in Nigeria. Both endeavours are relatively novel. There's been a growing emphasis on IRT in the recent times. In fact the International Test Commission has recommended that IRT be used for the proper description and evaluation of existing and widely used psychological instruments for confirmation or otherwise of their psychometric properties (Muñiz, 2011). Each confirmation

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

of suitably high psychometric properties as in Kpolovie and Emekene (2016) and in the current investigation implies that IRT could be employed in test development.

The results of this current study has conclusively affirmed the fact that the 3-Parameter Logistic IRT Model is the most suitable for examining the Advance Progressive Matrices-Smart Version (APM-SV) scale. Some authors (Gallini, 1983 and Raven, Prieler & Benesch, 2005) suggested that the three-parameter logistic model (3-PLM)-discrimination, difficulty and guessing – is preferable because there is a guessing component due to the multiple-choice format of the matrices. Some others (Çikrikçi-Demįrtaşli, 2000 and Georgiev, 2008) argued that guessing is irrelevant as each matrix has eight response options, and they opted for the two-parameter model (2PL). Although from the results obtained in this study there were only limited or no guessing involved in the test outcome, nevertheless the 3-PLM was found to be the most suitable for the short form-the Advanced Progressive Matrices-Smart Version (APM-SV) created by this researcher. In fact all the 15 items of the test for the abridge form-the APM-SV made a perfect fit under the 3-PLM in this current study. The APM-SV had an overall model fit with a Chi-Square value of 1467.487, a degrees of freedom (df) of 504, a probability of 0.000 and -2 logistic likelihood of 15613. The APM-SV scale within the framework of IRT yielded a strong reliability as indicated by the Alpha value of 0.83. This appears to be the highest reliability ever in the history of the investigations of the short forms of the APM. This is probably due to the in-depth analytical ability of X-Caliber 4.2. In the past, Arthur & Day (1994) reported reliability values that ranged from 0.58 - 0.66 for the 12 short-form items extracted from the full 36-item version and 0.72 for the full form. Thus their 12-item scale resulted in a lower reliability compared to the full APM test. Sefcek (2007) reported a moderate reliability value of 0.79 for the APM-18 scale which he created. Chiesi, Ciancaleoni, Galli, Morsanyi & Primi (2011) reported a reliability value of 0.62 for the investigation they carried with Arthur and Day's APM-SF with 12 items.

The Test information function (TIF) at each level of theta for the APM-SV provided enormous information at each level of theta to a maximum information degree of 5.825 at theta = -2.050. The one construct examined by the APM-SV scale in this study, which is intelligence explained 23.063 or 81.754% of the total variance. This is incontrovertibly a sizable chunk of the model. Therefore the underlining construct ostensibly being examined was indeed effectively examined by the scale and it ensured its unidimensionality. And since the assumption of unidimensionality was met by the 3-PLM, it invariably meant that the local independence held. Chiesi, Ciancaleoni, Galli, Morsanyi & Primi, (2011) reported in their own study that the unidimensionality assumption was met. This current study has again confirm their findings. Specifically, Chiesi, et al (2011) wrote that "unidimensionality was sustained across the five groups CFI and TLI with values that ranged from .98 to .99 and these indicated very good fit." The unidimensionality of the underlining construct of the APM-SV scale was also checked and assured. The instrument indeed measures consistently the fluid ability of those who take the test. All the 15 items measure one construct, the intelligence of the test taker. The results of both the APM and APM-SV in this current study have confirmed what Forby & Ben-Porath (2007) wrote with regards to the issue and importance of unidimensionality of instruments designed to measure one construct. They wrote that "research designed to assess the impact of violations of the unidimensionality assumption has suggested that the unidimensional IRT models are relatively robust with respect to moderate violations of strict unidimensionality, and that the most important issue concerns the relative degree to which the item pool is dominated by a single latent trait."

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

According to X-Calibre manual, the difficulty index "ranges in theory from negative to positive infinity, but in practice from -3.0 (very easy) to +3.0 (very difficult)." A careful examination of the *b* parameter column of the 3-PLM used for examining the APM scale in this current study yielded the values of *b* that ranged from -2.595 to +2.133. Thus *b* parameter graduated from very easy to very difficult. The APM did not discriminate the test outcomes in terms of groups, whether gender groups, age groups or ethnic or cultural affiliations in Nigeria. Similarly the *b* parameter graduated in value from -0.322 to 2.086 showing an increasing order of difficulty. The items of the APM-SV were also bias free towards all the groups that were investigated: gender, age and ethnicity. The degree of guessing was found to be low amongst the undergraduates as well as the senior secondary school students. The *c parameter* values ranged from 0.189 to 617. This indicated a limited or no guessing in the test outcome. Therefore the degree of guessing was low amongst the undergraduates as well as the senior secondary school students. Chiesi, et al (2011) also reported that "the DIF analysis across age revealed that one item displayed a significant difference. This item was designated as a study item for the DIF analysis. The eleven remaining items were identified as anchor items. Item 23 showed DIF for the difficulty parameter (*b*) as indicated by the significant difference between the -2loglikelihood. The NCDIF index was .02. Its magnitude confirmed that Item 23 showed non-ignorable DIF. Specifically, parameters indicated that it was easier for older respondents.

Nonetheless, since only one item exhibits DIF (less than 10% of the total number of items that composed the full scale while the APM-SF was considered equivalent across age." In the case of this present study no item yielded a non-equivalent measure across all the groups, namely gender, age and cultural affiliations. In a three parts study carried out by Ablard & Mills (1996) titled "Evaluating Abridged Versions Of The Raven's Advanced Progressive Matrices For Identifying Students With Academic Talent" found out that older students performed significantly better than younger students, for Set I, t(219) = 4.35, P< .001, and Set II, t(219) = 5.34, P< .001 and significance tests for independent alpha coefficients showed that the alpha coefficients were not significantly different between younger and older students *p*=.05 for either Set 1 or Set 2 of the APM." They concluded therefore that "the APM set 1 and set 2 have equivalent internal reliability for younger and older students who ranged from fifth to ninth grades." This present study showed little or no noticeable differences if any between the performance of the university undergraduates of 16-40 years old and senior secondary school students of 12-20 years old that represented the two age groups investigated. Without discountenancing the reports of Ablard & Mills' (1996) work, the present study has simply confirmed the fact that the APM are suitable for both adolescents of age 12 above as well as adults of any age. In a validation study conducted by Rushton, Skuy & Bons (2004) titled “Construct Validity of Raven’s Advanced Progressive Matrices for African and Non-African Engineering Students in South Africa” with a primary concern to test the hypothesis that the Raven’s Advanced Progressive Matrices has the same construct validity among African university students as it does in non-African students, they examined data from 306 highly brilliant of 17- to 23-year olds in the Faculties of Engineering and the Built Environment at the University of the Witwatersrand (177 Africans, 57 East Indians, 72 Whites; 54 women, 252 men). Analyses using the CTT models were made of the Matrices scores, an English Comprehension test, and the Similarities subscale from the South African Wechsler Adult Intelligence Scale, end-of-year university grades, and high-school grade point average. Out of the 36 Matrices problems, the African students solved an average of 23; East Indian students, 26; and White students, 29 (po.001), placing them at the 60th, 71st, and 86th percentiles, respectively, and yielding IQ equivalents of 103, 108, and 118 on the 1993 US norms. For the Raven’s Advanced Progressive Matrices, all calculations were based on raw scores, with each

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

of the 36 items scored as 0 (incorrect) or 1 (correct). Internal consistencies based on Cronbach's alpha were 0.86 for the sample as a whole (n5306), 0.86 for the Africans (n5177), 0.79, for the East Indians (n557), and 0.75 for the Whites (n572). The SPSS output for the percentile computations for this present study contains the normalized standard scores on the APM test that have been converted into percentile ranks for it to be more readily understood by all. Details of which is contained in the appendix section of the main report of this work.

Following a four prong rule defined by this researcher and the precedents set by Arthur and Day, Sefcek and other researchers, a modified version of the APM has now been created. The current researcher selected items based on their order of increasing difficulty and fair contributions to the item total correlation. Now the Advanced Progressive Matrices-Smart Version (APM-SV) has 15 items. The items are: 1, 2, 7, 10, 12, 15, 17, 21, 23, 24, 28, 29, 31, 33, and 35. These items were selected from the 36-items of set 2 of the APM on the basis of their high item-total correlation coefficients. The APM-SV scale correlated positively well with other measures of fluid ability. APM-SV and the full APM scale recorded a positive correlation of 0.96. APM-SV and the Digit Span scale of the WISC–R, a working memory measure had a correlation coefficient of 0.74. APM-SV scale and Bennett Mechanical Comprehension Test (BMCT) had a correlation value of 0.60. The APM-SV scale showed a very strong positive relationship with Mathematical Reasoning ability with a correlation value of 0.84. APM-SV and Probabilistic Reasoning was found to be the least in strength even though positive in orientation. The correlation recorded a small but positive value of 0.037. The Culture Fair Intelligence Test, which in itself is a measure of fluid general mental ability correlated very strongly with the modified APM, the APM-SV. It recorded a correlation value of 0.86. This is very significant for this study. It is interesting to recall that CFIT scale has been standardized with norms, reliability, validity, etc locally derived in Nigeria using a Nigerian sample (Kpolovie, 2016b; 2015; 2016a; Kpolovie & Emekene, 2016).

A study conducted by Mogle, Lovett, Stawski, and Sliwinski (2008) examined the relations among working memory capacity (WMC), secondary memory (SM), and fluid intelligence (*g*F) measured with the Raven's APM via a latent variable analysis. They found out that a latent SM variable was uniquely related to a measure of *g*F as captured by the APM, and that SM accounted for the variance shared between *g*F and WMC. Specifically, two of the SM tasks (story recognition and paired associates) correlated with the *g*F measure at 0.26 and 0.25 (respectively), whereas the word recognition measure correlated appreciably higher, with the *g*F measure at 0.39. All of the WMC tasks, however, correlated with *g*F at 0.29. In their own research work Chiesi, Ciancaleoni, Galli, Morsanyi & Primi, (2011) reported that "the correlation between the APM-SF and digit span scores was $r(N = 653) = 0.42$, $p< 0.001$, the correlations between the APM-SF score and the reasoning measures were $r(N = 921) = 0.35$, $p< 0.001$ for probabilistic reasoning, $r(N = 202) = 0.27$, $p< 0.001$ for mechanical reasoning, and $r(N = 151) = .47$, $p< .001$ for mathematical reasoning while the correlations with the achievement measures indicated that the APM-SF was related to high school final grades ($r(N = 126) = .25$, $p<.01$), final mathematics grades $r(N = 1157) = .24$, $p< .001$), and final introductory statistics grades $r(N = 115) = .27$, $p< .01$)." This present study has attested to the fact that the APM is indeed a measure of intelligence and that a strong correlation exist between it and other measures of intelligence.

In summary this work applied the item response theory (IRT), a well-known theory of psychometric analysis on the Raven's Advanced Progressive Matrices. From the review of literature and as it has now been confirmed by this current research work, the Raven's

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

Advanced Progressive Matrices (APM) is a non-verbal multiple choice measure of the reasoning (or, better, 'meaning-making') component of Spearman's ***g***. The Spearman's ***g*** is often referred to as "general intelligence" (Raven, Raven & Court, 2003, updated 2012). Raven's Progressive Matrices are largely employed by researchers and practitioners in the field of psychometrics, education, medicine and the social sciences. This present study has also confirmed that the instrument is indeed very suitable for cross-cultural studies of intelligence, appropriate for measuring cognitive ability free of verbal interference, helpful in assessing ethnically diverse populations and serves well as an intelligence test that detects "sub-optimal performance" especially the discovery of gifted or talented individual. It is an instrument that has no biases towards gender or age groups. The instrument was designed to serve adolescents of age 12 and above as well as the adults of any age. True to the intensions of the test maker, this study revealed an even performance between the adolescent and adult samples of this study. The adolescent sample was made up of 1600 senior secondary school students while the adult sample comprised of 1500 university undergraduates. The results of the analysis of scores collected from the respondents to the APM scale's test confirmed the fact that the instrument serves to minimize the impact of language skills and cultural biases and therefore they are particularly well suited for measuring the intelligence of individuals whose native language is not English, as well as those who may have reading problems or hearing impairment. It was also found out from the review of literature that the APM test helps in determining managerial skills amongst executives of corporate organizations, intellectual efficiency amongst learners, the speed and accuracy of high level cognition work, cognitive processes or organic dysfunction amongst children and the elderly. Raven, Raven and Court (2012) said the "Raven's Advanced Progressive Matrices (APM) measure two complementary components of general intelligence: the capacity to think clearly and make sense of complex data (eductive ability); and the capacity to store and reproduce information (reproductive ability)".

The Advanced Progressive Matrices (APM) is in high demand as an instrument of choice among researchers in America, Europe and Asia. Meanwhile, the instrument is hardly known let alone employed in research work in Africa, particularly Nigeria (Kpolovie & Emekene, 2016). One of the likely reasons might be due to the fact that it has never been standardized with norms in Nigeria. Therefore this research work on the application of psychometric analyses on the Advanced Progressive Matrices (APM) was focused on modifying the APM scale by locally generating fit statistics, reliability, parameter indexes, and so on in Nigeria. This it is hoped will make it available to the research community in Africa, particularly Nigeria as an instrument of choice for intelligence related research work. Thus the huge task of applying IRT on the Raven's Advanced Progressive Matrices (APM) for use in Nigeria was embarked upon by the current researcher. The result of this exercise has effectively brought about the establishment of the relevant statistical parameters for describing the APM in Nigeria henceforth. The APM which is an important and powerful tool for measuring intelligence can now be incorporated into the Nigerian educational system. The APM has been found to be reliable, valid and bias-free; and is in use in several foreign countries. In each country where the instrument is effectively in use, it was first standardized and validated. Its use in Nigeria or any country without first validating and standardizing (i.e. establishing its reliability, validity and norms, using that country's sample), will amount to abuse of the test. This is because it is psychometrically wrong to use a test standardized on one population for another population. This study was chiefly aimed at making the APM more user-friendly for the purpose of achieving a productive search for the academically talented among our children and the many youths of today in Nigeria by developing a short form of the APM now dubbed as the Advanced Progressive Matrices-Smart Version (APM-SV). The development of the short form was made

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

possible through the computing power, mathematical precision and the robust analytical nature of the item response theory.

Two research questions were posed with the ultimate goal of determining the reliability, validity and norms of the APM in Nigeria for it to be appropriately put to use in this country. Multiple triangulation research design, which permits flexible and robust approaches in establishment of psychometric properties and norms of the test was employed. Stratified random sampling was adopted to obtain 1500 university undergraduates within the age category of 16 to 40 years and 1600 senior secondary school students with the age category of 12 to 20 years from four ethnic groups (Hausa, Igbo Yoruba and Minority). The X-Calibre 4.2, a powerful software for iterating IRT related data and SPSS together with Microsoft excel were deployed for the analyses of the data generated for this work. The analyses included the iteration of the parameter logistic IRT models, the computation of the test information function (TIF), the determination of differential item function (DIF) among the test items and their consequential biases where present, the calculations of the item total correlations so as to appreciate the contribution of each item to the total goodness of fit statistics of the test. The results showed that the 3-Parameter Logistic IRT Model was the most suitable for examining the Advanced Progressive Matrices-Smart Version (APM-SV) scales. The APM scale had an overall model goodness of fit statistics with a Chi-Square value of 1120.993, a degree of freedom (df) of 432, a probability of 0.000 and -2 logistic likelihood of 40593. The reliability of the APM scale within the framework of IRT as indicated by the Alpha value for the full test with 36 items is 0.95. The TIF provided enormous information at each level of theta to a Maximum information degree of 5.825 at theta = -2.050. The one construct examined by the APM scale, which is intelligence explains 33.005 or 85.686% of the total variance. This is incontrovertibly a sizable chunk of the model. Therefore the underlining construct ostensibly being examined was indeed effectively examined by the scale and it ensured its unidimensionality. And since the assumption of unidimensionality was met by the 3-PLM, it invariably meant that the local independence held. According to X-Calibre manual, the difficulty index "ranges in theory from negative to positive infinity, but in practice from -3.0 (very easy) to +3.0 (very difficult)." A careful examination of the *b* parameter column of the 3-PLM used for examining the APM scale yielded the values of *b* that ranged from -2.595 to +2.133. Thus *b* parameter graduated from very easy to very difficult. The APM did not discriminate the test outcomes in terms of groups, whether gender groups, age groups or ethnic or cultural affiliations in Nigeria. The *c parameter* values ranged from 0.189 to 617. This indicated a limited or no guessing in the test outcome. Therefore the degree of guessing was low amongst the undergraduates as well as the senior secondary school students. This indicated that the APM-SV scale was not biased towards the ethnic or age groups. The stated purposes of this research endeavour abinitio included a thorough and detail application of psychometric analyses on the Advanced Progressive matrices (APM) with a view towards establishing its concomitant psychometric properties in Nigeria by developing an abridged form of it, compute the most appropriate IRT model that has a more viable goodness of fit statistics and provide relevant and usable parameter modules that can be used by experts in the field of psychometrics as well as non-experts including those who are not familiar at all with the language of testing. The results of this investigation will certainly mark an eon in the chronicles of educational development and Psychological testing in Africa, particularly Nigeria. This is an explicit investigation that is total in its mission and delivery. It will serve as a milestone in the psychometrics and testing industry in Nigeria and beyond. In the past in this country, perhaps the only internationally acclaimed instrument for measuring intelligence available to the psychometric community might be the Culture Fair Intelligence Test (CFIT) standardized with

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

norms and other statistical parameters in Nigeria by Kpolovie in 2001 (Kpolovie, 2010). This will be an additional valid and scientifically proven measure of intelligence in this country. The study of the abridged form of the Advanced Progressive Matrices has been furthered through this research work. And this is coming right from Nigeria. The need for an abridgement of the APM testing scale is obvious. Today, people are on the go. Everyone lives online. Everyone has a psychometric location, placement and at least an address online. The language of today has changed from the old metaphor of slow and steady win the race. Speed and accuracy are the in things now. They are the defining characteristics of today's world of multi-tasking. Anything that must accomplish objective must be sharp, smart and deliver on point. Hence the creation of the Advanced Progressive Matrices-Smart Version.

The power of the IRT was succinctly tested and proven in this study. The International Test Commission had recommended that IRT be used for the proper description and evaluation of existing and widely used psychological instruments (Muñiz, 2011). Hence IRT was fully deployed and applied in this study. CTT obviously has its own place of pride in the psychometrics industry but the superior power of the IRT cannot be ignored anymore. If examination bodies in Nigeria, all our institutions of learning and indeed all testing agencies particularly those that high labour and or recommend existing staff for promotions for corporate organizations will embrace and utilize the robust, in-depth and diverse inclusiveness of IRT, examination malpractices and inequalities in hiring and promotions might soon be a thing of the past in Nigeria. Biases of all forms in test items can be easily eliminated and public examinations can then become more credible (Kpolovie & Iderima, 2016; Kpolovie, Joe & Okoto, 2014; Kpolovie, 2012; 2016a; 2016b; Joint Admission and Matriculation Board JAMB, 2016). All instruments of assessment, any kind of assessment, whether in the institutions of learning or within the corporate world, when subjected to the computing potency of IRT can be easily adjusted and upgraded so as to deliver on their accurate objectives. One of the unstated missions of this research work was to make a very strong statement that IRT is the way, not just a way, to go within the testing industry in Nigeria if things must be done right. It will also be right to submit at this juncture that the capacity and expertise needed to handle and execute IRT related tasks are now available, at least within the department of Educational Psychology, Faculty of Education, University of Port Harcourt, Nigeria if not in other climes as well. Based on the discussions above, the following recommendation were made:

1. That the Raven's Advanced Progressive Matrices (APM) scale, a very popular and reputed measure of intelligence, be incorporated into the pool of research instruments in Nigeria.

2. A deliberate efforts should be made by psychometricians in Nigeria to promote the use of this measure of fluid ability.

3. That public examination bodies such as Joint admission and matriculation board, West Africa Examinations Council, National Examination Council and other similar bodies to more vigorously review their existing instruments and the new ones being constructed using the item response theory models since this has been recommended by the International Testing Commission.

4. All instruments of assessment, any kind of assessment, whether in the school system or within the corporate practice be subjected to the rigorous item response theory analysis especially the Differential Item Functioning, for scrutiny so as to eliminate all kinds of biases from testing instruments because the purpose of IRT among other things is to

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

provide a framework for evaluating how well assessments work and how well individual items on assessment work.

5. All information generated from Test Information Function be used to review testing instruments so as to upgrade and make them deliver their objectives on point. It has been empirically proven that IRT has a superior ability in helping to develop high stake tests.

6. The abridged version of the APM scale known as the Advanced Progressive Matrices-Smart Version (APM-SV) is now available to the research community in Nigeria and beyond.

7. The APM-SV scale, a measure of fluid ability just like its parent scale, the full form APM, yielded a strong and positive correlation with the Culture Fair intelligence test. Culture Fair Intelligence Test on the other hand, which in itself is an undisputable measure of intelligence is known to be positively correlated with Achievement tests. It is therefore recommended that teachers, lecturers, trainers and other persons that are involved with the impartation of intellectual skills adopt and use the APM-SV scale for a routine assessment of their trainees. It will serve as quick check on how well their students are catching the lessons. Anyone who does well in the APM test should do well in their regular lessons or instructions. A good performance in the APM test could be an indication of the innate ability to do well in any kind of instruction, training, lesson or lecture.

CONCLUSION

In conclusion, the stated purposes of this research endeavoured included a thorough and detail application of psychometric analyses on the Advanced Progressive matrices (APM) with a view towards establishing its concomitant psychometric properties in Nigeria, develop an abridged form of it, compute the most appropriate IRT model that has a more viable goodness of fit statistics and provide relevant and usable norms and standardization modules that can be used by experts in the field of psychometrics as well as non-experts including those who are not familiar at all with the language of testing. The results of this investigation will certainly mark an eon in the chronicles of educational development and Psychological testing in Africa, particularly Nigeria. This is an explicit investigation that is total in its mission and delivery. This study will serve as a milestone in the psychometrics and testing industry in Nigeria and beyond. In the past in this country, perhaps the only internationally acclaimed instrument for measuring intelligence available to the psychometric community might be the Culture Fair Intelligence Test (CFIT) normed and standardized in Nigeria by Kpolovie in 2001 (Kpolovie, 2010). This will be an additional valid and scientifically proven measure of mental ability that could be used in this Nigeria. The study of the abridged form of the Advanced Progressive Matrices has been furthered through this research work. And this is coming right from Nigeria. The need for an abridgement of the APM testing scale is obvious. Today, people are on the go. Everyone lives online. Everyone has a psychometric location, placement and at least an address online. The language of today has changed from the old metaphor of slow and steady win the race. Speed and accuracy are the in things now. They are the defining characteristics of today's world of multi-tasking. Anything that must accomplish objective must be sharp, smart and deliver on point. Hence the creation of the Advanced Progressive Matrices-Smart

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

Version. This, it is hoped will help decision makers within and outside the education sector in Nigeria.

From this study on psychometric advent of Advanced Progressive Matrices – Smart Version (APM-SV) in Nigeria, the following contribution to knowledge have been made:

1. The study used the IRT models (1, 2, and 3PLM) to derive item parameters, reliability and validity, Test Information Function, Differential Item Functioning (DIF), Item Response Function (IRF), Unidimensionality using factor analysis through data reduction method and Goodness of Fit Statistics among other statistical measures for the APM-SV scale in Nigeria.

2. The study has created a more user friendly version of the APM dubbed APM-SV.

3. The APM-SV will enhance quick and easy the identification, placement, acceleration and enrichment of the gifted/talented students in Nigerian as well as ability selection and placement within and outside the education management climes is now available to the research community in Nigeria.

4. This work has significantly furthered the study on the short form of the APM.

5. Results from this study have again confirmed the potency of the IRT models. The findings about the suitability of the 3-PLM was an eye opener to the fact that a perfect fit statistic can be achieved and it enhances the credibility of any measuring instrument because it can easily show the directions and dimensions of the items and how well the items are performing the intended objectives they were designed to serve in the testing instrument.

6. The study correlated APM-SV scale with other measures of mental ability such as working memory measures, Mechanical Reasoning ability, Mathematical Reasoning ability, Probabilistic Reasoning ability, APM and CFIT and thereby confirmed that the short form of the APM could indeed be a good measure of fluid ability.

7. The APM-SV did not discriminate the test outcomes in terms of groups, whether gender, age or ethnic or cultural affiliations in Nigeria. APM-SV were also bias free towards all the groups that were investigated: gender, age and ethnicity. These succinctly confirmed the fact that APM is indeed a perfect measure of fluid ability among those whose native language is not English.

8. The study also confirmed the fact that with eight response option the possibility of guessing will be greatly reduced in multiple choice examinations. Examination bodies in Nigeria might want to validate this inference.

9. A four prong rule defined by this researcher with which the APM was modified to create a more user friendly version of the APM can be used by other researchers to modify similar instruments. The four prong rules are:

   i. The APM was divided into 6 sections with each section containing 6-items, based on difficulty. The sections are: Section 1: items 1-6, Section 2: items 7-12, Section 3: items 13-18, Section 4: items 19-24, Section 5: items 25-30 and Section 6: items 31-36.

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

ii. The two items with the highest Pearson point-biserial correlation (*r-pbis*) were selected from each section: based on this rule the following items were selected: items 1, 2, 7, 12, 15, 17, 21, 23, 28, 29, 31, and 35.

iii. One additional item with the highest Pearson point-biserial correlation (*r-pbis*) next to the two already selected in that section was chosen from sections 2, 4 and 6 leaving out sections 1, 3 and 5. Based on this rule, the following items were selected: Section 2 - item 10, section 4 - item 24, section 6 - item 33.

iv. And in the case of a tie, including the item that resulted in the largest drop in internal consistency provided it was not excluded from the full test.

Based on findings of this investigation, it is strongly recommended that:

1. That the Raven's Advanced Progressive Matrices (APM) scale, a very popular and reputed measure of intelligence, be fully incorporated into the pool of research instruments in Nigeria.
2. A deliberate efforts should be made by psychometricians in Nigeria to promote the use of this measure of fluid ability.
3. That public examination bodies such as Joint admission and matriculation board, West Africa Examinations Council, National Examination Council and other similar bodies to more vigorously review their existing instruments and the new ones being constructed using the item response theory models since this has been recommended by the International Testing Commission.
4. Further research is required in the area of using the APM to identify academically talented students.
5. Further study is also now required to establish norms for the abridged version created by this researcher or the earlier ones created by other researchers such as the 12 items' Advanced Progressive Matrices-Short Form (APM-SF) or the 18 items' Advanced Progressive Matrices-Medium Form (APM-MF).
6. Similar studies as the current research work are also required for the Standard Progressive Matrices and the Coloured Progressive Matrices.

REFERENCES

Abad, FJ, Colom R, Rebollo I, & Escorial, S (2004). Sex differential item functioning in the Raven's Advanced Progressive Matrices: Evidence for bias. *Personality and Individual Differences*, 36 (2004), pp. 1459–1470

Ackerman, PL, Beier, ME, & Boyle, MO (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131*, 30–60

Arthur, W & Day D (1994). Development of a short form for the Raven Advanced Progressive Matrices test Educational and Psychological Measurement, 54 (1994), pp. 395–403

Bennett, CK (1969). Bennett mechanical comprehension test, San Antonio: The Psychological Corporation.

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

Bors, DA & Stokes TL (1998). Raven's Advanced Progressive Matrices: Norms for the first-year university students and the development of a short form Educational and Psychological Measurement, 58 (3) (1998), pp. 382–398

Cattell, RB (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54, 1-22.*

Chiesi, F, Ciancaleoni, M, Galli, S & Primi, C (2012, March 26). Using the Advanced Progressive Matrices (Set I) to Assess Fluid Ability in a Short Time Frame: An Item Response Theory–Based Analysis. *Psychological Assessment*. Retrieved July 15, 2012, from http://my.apa.org/portal.

Chiesi, F, Ciancaleoni, M, Galli, S, Morsanyi, K, & Primi, C (2011). Item response theory analysis and differential item functioning across age, gender and country of a short form of the Advanced Progressive Matrices. *Learning and Individual Differences.* Retrieved June 3, 2012, from http://www.sciencedirect.com/science/article/

Chiesi, F, & Primi, C (2009). Recency effects in primary-age children and college students using a gaming situation. *International Electronic Journal of Mathematics Education, 4, 3,* 259 –274. Retrieved June 3, 2012 from www.iejme.com

Chiesi, F, Primi, C & Morsanyi, K (2011). Developmental changes in probabilistic reasoning: The role of cognitive capacity, instructions, thinking styles and relevant knowledge. *Thinking & Reasoning, 17,* 315–350 Retrieved June 3, 2012, from http://www.sciencedirect.com/science/article/

Çikrikçi-Demįrtaşli, N (2000). A study of Raven Standard Progressive Matrices Test's item measures under Classical and Item Response Model Paper presented at 31st European Mathematical Psychology Congress, Austria: Graz. Retrieved June 3, 2012, from http://www.sciencedirect.com/science/article/

Colom, R, Florez-Mendoza, C & Rebollo, I (2003). Working memory and intelligence. *Personality and Individual Differences, 34,* 33–39

Evers, A (2011). Testing practices and attitude towards tests and testing: The results of a global survey Paper presented at the 12th European Congress of Psychology, Istanbul (2011, July). Retrieved July 15, 2012, from http://my.apa.org/portal

Federal Ministry of Education (2014). *Staff and students data in our educational institutions*. Abuja: FME press.

Field, A (2005). Discovering statistics using SPSS for windows. London: Sage Publication.

FRN National Population Commission (2014). *Nigeria Demographic and Health Survey 2013*. USA: ICF International.

Gallini, JK, (1983) A Rasch analysis of Raven item data The Journal of Experimental Education, 52 (1) (1983), pp. 27–32.Retrieved July 15, 2012, from http://my.apa.org/portal

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

Gilovich, T, Griffin, D, & Kahneman, D (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment.* New York, NY: CambridgeUniversity Press.

Joe, I A (2000). Fundamental Statistics for Education and the behavioural sciences. Ibadan: Kraft Books Ltd.

Joint Admissions and Matriculation Board JAMB (2016). *Vital issues in the introduction of Computer-Based Testing in Large-Scale Assessment*. Abuja, Nigeria: Marvelous Mike Press Ltd.

Kpolovie, PJ (2016). *Excellent research methods*. Indiana: Partridge Africa Publishing.

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

Kpolovie, PJ (2016a). Single-subject research method: the needed simplification. *British Journal of Education.* 4(6), 68-95. http://www.eajournals.org/wp-content/uploads/Kpolovie-Peter-James.pdf

Kpolovie, PJ; Joe, AI; Okoto, T (2014). Academic achievement prediction: Role of interest in learning and attitude towards school. *International Journal of Humanities, Social Sciences and Education*. 1(11), 73-100. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.686.1998&rep=rep1&type=pdf

Kpolovie, PJ (2016b). Intelligence and academic achievement: A longitudinal survey. *International Journal of Recent Scientific Research*. ISSN 0976-3031

Kpolovie, PJ; Emekene, CO (2016). Item Response Theory validation of Advanced Progressive Matrices. *British Journal of Psychology Research*. 4(1), 1-32. http://www.eajournals.org/wp-content/uploads/Item-Response-Theory-Validation-of-Advanced-Progressive-Matrices-in-Nigeria1.pdf

Kpolovie, PJ; Iderima EC (2016). Learners' readiness for xMOOCs: Inequity in Nigeria. *European Journal of Computer Science and Information Technology*. 4(3), 16-46. http://www.eajournals.org/wp-content/uploads/Learners----readiness-for-xMOOCS-Inequity-in-Nigeria.pdf

Kpolovie, PJ (2015). Indispensability of intelligence testing in the repositioning and revitalization of Nigerian education. *Multidisciplinary Journal of Research Development*. 6(4), 1-11. http://www.globalacademicgroup.com/journals/nard/Kpolovie.pdf

Kpolovie, PJ (2014). *Tests, Measurement and Evaluation in Education. Second Edition*. Owerri: Springfield Publishers Ltd.

Kpolovie, PJ & Obilor, IE (2013a). Higher education for all in need through the National Open University of Nigeria: A paradox in policy practice. *Merit Research Journal of Education and Review (MRJER).* http://meritresearchjournals.org/er/content/2013/September/Kpolovie%20and%20Obilor.pdf ISSN: 2350-2282. 1(8), 172-180.

Kpolovie, PJ & Obilor, IE (2013b). Nigerian universities bag ludicrous ranks in world rankings of universities. *Universal Journal of Education and General Studies*. http://www.universalresearchjournals.org/ujegs 2(9), 303-323.

Kpolovie, PJ & Obilor, IE (2013c). Adequacy-inadequacy: Education funding in Nigeria. *Universal Journal of Education and General Studies*. http://www.universalresearchjournals.org/ujegs ISSN: 2277-0984. 2(8), 239-254.

Kpolovie, PJ (2012). E*ducation Reforms without Evaluation Designs: Nigeria at Risk*. New Owerri: Springfield Publishers Ltd.

Kpolovie, PJ (2011). *Statistical Techniques for Advanced Research*. New Owerri: Springfield Publishers Ltd.

Kpolovie, PJ (2010). *Advanced Research Methods*. New Owerri: Springfield Publishers Ltd.

Kpolovie, PJ (1999). Reliability of the Culture Fair Intelligence Test in Nigeria. *Journal of Education in Developing Areas (JEDA).* A Journal of the Faculty of Education, University of Port Harcourt. Nigeria. XIII, 18-24.

Michell, J (1999). *Measurement in Psychology*. Cambridge: Cambridge University Press.

Mogle, JA, Lovett, BJ, Stawski, RS, & Sliwinski, MJ (2008). What's so special about working memory? An examination of the relationship among working memory, secondary memory, and fluid intelligence. *Psychological Science*, 19, 1071-1077.

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

Morsanyi, K, Primi, C, Chiesi, F & Handley, SJ, (2009).The effects and side-effects of statistics education. Psychology students' (mis-)conceptions of probability Contemporary Educational Psychology, 34 (2009), pp. 210–220

Muñiz, J (2011). International strategies to improve tests and testing. Paper presented at the 12th European Congress of Psychology, Istanbul (2011, July)

Muñiz, J (2009). The role of EFPA in setting standards for tests and test use Paper presented at the 11th European Congress of Psychology, Oslo (2009, July). Retrieved June 3, 2012, from http://www.sciencedirect.com/science/article/

NEEDS Assessment of Nigerian Universities Committee (2013). *NEEDS Assessment of Nigerian Universities Reports*. Abuja: FME Press.

OECD (2015). *Education at a glance 2015 OECD indicators*. UK: OECD Publishing. http://dx.doi.org/10.1787/eag-2015-en

Ojerinde, D; Popoola, K; Ojo, F; and Ariyo, A (2014). *Practical applications of item response theory in large-scale assessment*. Nigeria: Marvelous Mike Press Limited.

Ojerinde, D; Popoola, K; Ojo, F; and Onyeneho, P (2012). *Introduction to item response theory: Parameter models, estimation and application*. Nigeria: Marvelous Mike Press Limited.

Ololube, NP & Kpolovie, PJ (2013). Literature and focus group analysis of the approaches and obstacle to effective educational planning in higher education in an emerging economy. *International Journal of Scientific Research in Education (IJSRE)*. http://www.ijsre.com/Vol.,%206 3-Ololube%20&%20 Kpolovie.pdf 6(3): 233–254.

Ololube, NP & Kpolovie, PJ (2012). Approaches to conducting scientific research in education, arts and the social sciences. *Online Journal of Education Research* 1(3): 44–56. http://onlineresearchjournals.org/IJER/pdf/2012/june/ Ololube%20and%20Kpolovie.pdf

Ololube, NP, Emejuru, PI, Kpolovie, PJ, Amaele, A, & Uzorka, MC (2012). Excellence in higher education: a powerful instrument for social and economic progress. In, N. P. Ololube and P. J. Kpolovie. *Educational Management in Developing Economies: Cases 'n' School Effectiveness and Quality Improvement*. Saarbucken, Germany: LAP LAMBERT Academic Publishing. ISBN 978-3-8465-8931-1. http://www.amazon.com/EducationalManagement-Developing-Economies-Effectiveness/dp/3846589314 93–112.

Ololube, NP, Kpolovie, PJ & Makewa, LN (2015). Handbook of Research on Enhancing Teacher Education with Advanced Instructional Technology. PA, USA: *Information Science Reference (an imprint of IGI Global)*. ISBN 13: 978146668162; EISBN 13: 9781466681637; DOI: 10.4018/978-1-4666-8162-0 http://www.igi-global.com/book/handbook-research-enhancing-teacher-education/120264

Orluwene, GW (2012). *Fundamentals of Testing and Non-Testing Tools in Educational Psychology*. Port Harcourt: Lucky Press Ltd.

Rathus, SA (1990). *Psychology: Annotated instructor's edition*. Chicago: Hoilt, Rinehart and Winston.

Raven, J (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology*, *41*(1), 1-48. Retrieved June 15, 2012, from http://www.wpspublish.com

Raven, JC (1962). *Advanced progressive matrices*. London: Lewis & Co. Ltd.

Raven, J. C., Court, J. H., & Raven, J. *(1983). Manual for Raven's Progressive Mamas and vocabulary scales, section* 4: *Advanced Progressive Matrices, Sets I and II*. London: H. K. Lewis.

ISSN 2055-0154(Print), ISSN 2055-0162(Online)

Raven, JC Raven J (Eds.), (2008). *Uses and abuses of intelligence: Studies advancing Spearman and Raven's quest for non-arbitrary metrics*. Unionville, New York: Royal Fireworks Press,.

Raven, J, Raven, JC, & Court, JH (2003). *Raven manual section 1: General overview*. Oxford: Oxford Psychologists Press.

Raven, J, Raven, JC & Court, JH (2012). *Raven manual section 1: General overview*. Oxford: Oxford Psychologists Press.

Raven, J, Raven, JC & Court, JH (1997). *Mill Hill Vocabulary Scale: 1998 Edition.* Oxford: Oxford Psychologists Press.

Raven, J, Raven, JC & Court, JH (1993). *Raven manual section 1: Generaloverview*. Oxford: Oxford Psychologists Press.

Raven, J, Raven, JC & Court, JH (1998). *Advanced Progressive Matrices*. Oxford: Oxford Psychologists Press.

Rushton, JP Skuy, M & Bons, TA (2004) Construct Validity of Raven's Advanced Progressive Matrices for African and Non-African Engineering Students in South Africa. *International Journal of Selection and Assessment* volume 12 number 3 September 2004, Retrieved from *www.charlesdarwinresearch.org/Ravens%20inIJSA04.pdf*

Sefcek, JA, (2007). Development of an 18-item short form of the Ravens Advanced Progressive Matrices (RAPM- 18). Retrieved from http://www.highbeam.com/doc/1G1-108790099.html

Spearman, C (1927).*The abilities of man*. London: Macmillan.

Thompson, B (2004). Five methodology errors in educational research: The pantheon of statistical significance and other faux pas. In B. Thompson (Ed.), *Advances in social science methodology* (pp. 23-86). Stamford, CT: JAI Press.

Van der Ark, AA (2010). Using item response theory to score the Advanced Progressive Matrices: Rationale and research findings. *Journal of Psychological Tests*, 37, 16-60.

Van der Ven AHGS & Ellis, JL (2000) A Rasch analysis of Raven's Standard Progressive Matrices. *Personality and Individual Differences*, 29 (2000), pp. 45–64.

Vigneau, F Bors DA (2005) Items in context: Assessing the dimensionality of Raven's Advanced Progressive Matrices. *Educational and Psychological Measurement*, 65 (1) (2005), pp. 109–123.

Wechsler, D (2013). *WISC–III.Wechsler Intelligence Scale for Children-Terza Edizione-Manuale* [*WISC–III. Wechsler Intelligence Scale for Children-Third edition-Handbook*]. Florence, Italy: Organizzazioni Speciali.

Western Psychological Services California (2015). *Western Psychological Services*. (www.wpspublish.com)

Wikipedia, (2012). Item Response theory. *Retrieved June 3, 2012 from* www.**wikipedia**.org/