CALIBRATION OF MATHEMATICS AND GEOGRAPHY ITEMS FOR JOINT COMMAND SCHOOLS PROMOTION EXAMINATIONS OF NIGERIAN ARMY EDUCATION CORPS IN NIGERIA

Valentine O. Enu
&
Eugenia A. Okwilagwe

Abstract

Many schools and organisations examine their students or candidates with items that are not of quality. This could be because they do not have item banks or that they do not know what it takes to develop item bank. Previous studies that calibrated items did so for external examinations for certification or placement. Mathematics and Geography items for Joint Command Schools Promotion Examination, an external examination meant for promotion organised by Nigerian Army Education Corps using Item Response Theory, were calibrated in this study. Survey research design that adopted multistage sampling technique was used in selecting a sample of 600 and 2,400 senior secondary two (SS2) students from Command Day Secondary Schools in Nigeria for the validation and calibration processes respectively. A-200 multiple-choice items for each subject pooled from four year JCSPE were validated. A-100 valid items each were used for the calibration of the items using Bilog-MG and Windstep 3.75 computer software programmes. The average difficulty, discrimination and guessing parameters of mathematics items were 0.63, 0.76 and 0.30 while that of geography were 0.24, -2.64 and 0.00 in that order. Mathematics items with difficulty level ranging from -1.20 to 2.01, discriminating level ranging from 0.24, -2. 1. 45 and guessing parameter ranging from 0.11 to 0.50 were more difficult than Geography items with difficulty level ranging from -4.85 to 3.49, discriminating level ranging from 0.11 to 0.90 and guessing parameter ranging from 0.00 to 0.24. Average mathematics ability of students was 0.25 while that of geography was 0.96, indicating that mathematics items were more difficult than geography items. On the basis of the analysis, it becomes necessary that NAEC should develop item banks in all school subjects that it examines students on to ensure item quality.

Keywords: Item response theory, Command Secondary Schools, Calibration, Mathematics, Nigeria

Introduction

Assessment of learning outcomes faces many challenges in Nigeria and perhaps in many other African countries. One of such challenges is the calibration of items in various school subjects leading to item banking to ease the problem of using poor quality items in students' assessment. Item banking solves and improves assessment practices in both internal and external examinations. The problem arises because many educational institutions and organisations do not have adequate knowledge and expertise to carry out such process. Such educational institutions and organisations like the Nigerian Army Education Corps (NAEC) administer test items on testees that are not of a high quality, standard or do not meet the ability of the testees. This situation leads to poor assessment and judgment of the individual testee. Reasons could be that such institutions and organisations do not have item banks or that they are unaware of what it takes to construct quality test items. Test insecurity has also led to question paper leakages and examination malpractice in internal and external examinations in Nigeria. The presence of item bank has always made such situation easy to overcome at a short notice. The poor performance of students in external examination that is very common in our society today could be attributable to the poor quality test items being used to assess students in internal examinations and to prepare them for external examinations.

An item bank is a collection of test items, organised, classified and catalogued in order to facilitate the construction of a variety of achievement and other types of mental tests (Vale, 2006). According to Rudner (1998), item banks are files of various suitable test items, scaled by subject area, instructional objectives measured and other pertinent item characteristics (item difficulty and item discrimination indices). It is a term used for a repository of test items that belong to a testing programme as well as all information pertaining to the items. Item bank provides such information as: item author, date written, item status (that is whether new, pilot, active or retired), correct answer, item format, classical test theory statistics, item response theory test statistics and user defined field (Vale, 2006). The process of item banking involves the establishment of a common calibration system of test items which are sample free. In item banking, an appropriate selection procedure is used to generate a test from a large pool of questions and catalogued in terms of content and difficulty which is calibrated to a standard scale before it is used. Item bank has a considerable flexibility because the test may be a short or a long test, broad or narrow, hard or easy according to the needs of the test constructor without distorting the test (Rudner, 1998; Akindele, 2004; Thompson, 2009).

With the introduction of standardised tests in education and psychological testing in the year 1905, to solve the problem of retarded children, Alfred Binet opened the door for ensuringthe validity and collection of large scale items in measuring a wide range of human abilities. He was less interested in the measurement of physical properties or their

sensation but was rather decisive entering the domain of the pure mental (cognitive) functions in order to define and measure intelligence (Vale, 2006).

Every year, NAEC conducts a promotion examination called Joint Command Secondary Schools Promotion Examinations (JCSPE) for students. It is centrally conducted for senior secondary class two (SS2) students. Only qualified students who pass the examination with a minimum of five credits including English Language and mathematics are allowed to proceed to senior secondary class three (SS3) in all Command Secondary Schools in Nigeria. Command Secondary Schools are private schools owned by the Nigerian Army and are under the supervision of NAEC. Developing item banks in mathematics and geography for JCSPE is desirable since there is none currently in place. Senior secondary school two teachers construct and send questions and their solutions (keys) in their various teaching subjects to NAEC headquarters every year for the purpose of this promotion examination. The purpose of developing item banks in these two subjects was to ensure that the items of the examination are secured, reliable and of high quality. It will also help to achieve the objective of conducting the examination by the NAEC for its schools which is to improve the existing system of assessment in Command Schools. The study is not aware of any item bank developed in any subject offered in this examination; hence, this study was designed to develop item banks in Mathematics and Geography JCSPE with a view to setting the trend for the improvement of items in other subject areas in the schools.

In the school system, a group of subjects are studied by students not only because of their importance but because of the naturally existing or perceived relation between and among these subjects. Consequently, there are Science-related, Social science-related and Art-related subjects in most schools' planned programmes of instructions for students. The choice of selecting Mathematics and Geography items for calibration in this study was based on this perspective. Furthermore, mathematics is a fundamental mental school subject which acts as a basic index for understanding science, the complexity of modern-day technology, human growth and development (Adeleke, 2007). It is the science of pattern and it is a compulsory school subject in Nigeria educational system, especially at the primary and secondary levels of education. The usefulness of Mathematics is in its everyday life application**.** Mathematics achievement is the proficiency of an individual in any sub-group of Mathematics or all the entire subdivisions of mathematics.

After 20 years of educational research on the issue of mathematics achievement among students, deficiencies in the academic performance of students in Mathematics and Mathematics-related subjects such as the sciences to which Geography belong, persist (Thiessen & Blasius, 2008). As one of the key subjects in the school system in Nigeria, success in Mathematics is a major determinant of some students' future and as such, a

high achievement in it indicates to a great extent the level of these students' thinking ability. As a vital tool for the understanding and application of science and technology, the discipline plays important role in the much-needed technological and national development which has become imperative in the developing nations of the world (Hopkins, 2004). An individual's mathematics knowledge is his/her tendency to respond to perceived mathematical problem situation by reflecting on the problem and its solution in a social context and by constructing or reconstructing mathematical actions, processes, objects and organising these in a way they can be used in dealing with situations. Hence, mathematics constitutes an important component of most if not all examinations, especially in Command Secondary Schools in Nigeria.

Geography is a distinct body of knowledge that deals with the earth as a total system that teaches the effective tools with which a learner helps him/herself to seek for facts which can be used to explain phenomena in the environment (Falaye, 1995). The study of Geography equips the individual with skills to earn a living and contribute to the socio-economic development of a society. Geography balances the societal needs and that of the individual student. Hence, its contribution to the development of the individual, community and the nation in general is enormous. This simply means that Geography equips individual with information that produces informed and united society, especially in a developing nation like Nigeria with many tribes and diverse cultures.

A lot of concerns have been expressed on the state of education in Nigeria, and in spite of the benefit of these subjects, the performance of students in these two subjects (mathematics and geography) and indeed, in other school subjects tends to be modest (Asim, 2007). Among these concerns are: poor teaching methods, inadequate teaching staff, and poor testing items (poor assessment materials) (Okwilagwe, 2002; Grifith, 2005; Njabili, Abedi, Magesse & Kalole, 2005; Asim, 2007). Teachers have also been blamed for the poor performance of students in school subjects in external examinations as a result of teachers' incompetency in assessment (Asim, 2007; Ojo, 2006; Okwilagwe, 2011). The development of item bank for Joint Command School Mathematics and Geography promotion examinations using the pool of items meant to improve the testing system is to ensure that the items in the examination papers are secure and of a high quality. This study, therefore, developed item bank for JCSPE from the item pool of NAEC's JCSPE questions with a view to ensuring that the calibrated items are of high quality.

Research Questions

1. What are the item parameter values of the mathematics and geography test items using item response theory estimation procedures?
2. What are the item parameter estimates of the calibrated items in the mathematics and geography prototype item banks?

Methodology Research Design

Survey research design was adopted in this study. The variables in the study were not manipulated since their manifestations have already occurred.

Target Population

The target population for the study was all the senior secondary school two (SS2) students in Command Secondary Schools in Nigeria in the *2012/2013* session.

Sampling Technique and Sample

Multistage sampling technique was used. Simple random sampling technique was used to select 2nd Mechanised Division of Nigerian Army, Ibadan out of the five Mechanised Divisions in Nigerian Army. Three schools were randomly selected out of six Command Schools in 2nd Mechanised Divisions of Nigerian Army. In each of the selected schools in the Division, simple random sampling technique was used to select one hundred 100 students each for Mathematics and Geography for the validation of the items. Simple random sampling technique was also used to select two Command Schools from each of the remaining four Mechanised Divisions of Nigerian Army for the calibration. In each of the selected schools in the Divisions, simple random sampling technique was used to select 150 students each for Mathematics and Geography for the items calibration. A total of 600 and 2,400 formed the sample for the validation and calibration respectively.

Instrumentation

Two instruments were used for this study. A-200 multiple choice items was pooled from the past four years of Joint Command Secondary Schools Promotion Examination (JCSPE) in Mathematics and Geography for validation. The items from the past JCSPE covered major areas of Mathematics, such as Number and Numerations, Algebra, Geometry, Statistics and Probability. Geography items also covered Physical, Regional and Human/Economic. A-100 multiple-choice items selected from the validated items were used for calibration. With the knowledge of the construction of table of specification, the researchers ensured that the items covered the knowledge, comprehension and application levels of Bloom's taxonomy.

Item Selection

The testees' responses to the items were marked and correct response was awarded 1 while incorrect response and non-response were awarded O. The I-parameter logistic model or Rasch model was used to estimate the difficulty level of each item using Windstep 3.75 IRT package. Winsteps was originally developed by Benjamin Wright and John Michael Linacre at the University of Chicago in the 1980s (Linacre, 2004). This is because previous studies have suggested that a sample as large as 200 to 250 testees would be sufficient enough to estimate parameters using I-parameter (Rasch) model (Linacre, 2004; Baghaei, 2008). According to Rupp (2009), about 1000 testees are needed for 3- parameter model to have stable parameter estimate while about 250 testees is enough for 1 or 2 parameter models for obtaining stable estimate of parameters.

According to Green and Franton (2002), a sample size of at least 100 and a minimum of 20 items are enough for obtaining stable indices when using l-parameter (Rasch) model analysis. Item selection in IRT models is based on the intended purpose of the test. The selection of items depends on the amount of information they will contribute to the overall information supplied by the test. Item difficulty parameter was seriously taken into consideration in the selection process to ensure that the items have reasonable spread across the entire test (Umobong, 2004). Small variation in item difficulty is important in measurement of ability level in any target population but the smaller the standard error of measurement, the greater the precision of the measurement. The item logit and person logit separations and reliability were examined before any interpretation of the data. This was found to be 4.48 and 0.95, 2.65 and 0.88 for Mathematics and 4.75 and 0.96, 2.97 and 0.90 for Geography. This separation indicates the number of groups the students can be separated into according to their abilities. According to Eluwa, Idowu and Abang (2011), any item with ZSTD (standardised) infit and/or outfit statistics between -2 and 2 and MNSQ (Mean square) infit and/outfit between 0.6 to 1.4 should be selected. This study adopted these conditions for selecting items.

Data Collection Procedure

Permissions were obtained from the Divisional Education Officer of 2nd Mechanised Division of Nigerian Army for the validation of instruments and the Core Commander Nigerian Army Education Corps (NAEC) of Nigerian Army for the calibration of the items to the Commandants of the selected Command Schools. This is to ensure full cooperation and support of the schools involved.

Analysis Procedure for Calibration

The study employed the approach of the 3-parameter model. This is hinged on the data types used which are dichotomous and are obtained from multiple-choice items. The data was introduced into Bilog-MG software programme for processing. The programme is

designed for wide range of applications of item response theory to testing practical problems. For instance, Bilog has special features, such as: choice of any of the three models of IRT; test of fitness of each item, estimation of standard error of all items, analysis of multiple subsets, among others. Binary scoring method of item response was used that is, right/wrong scoring pattern. This took care of possibility of guessing associated with multiple-choice items.

Analysis and Results

The results of data analysis and discussion as they relate to the research questions are presented as follows:

Research Question 1

What are the item parameter values of the mathematics and geography test items using item response theory estimation procedures?

Table

Item parameter Estimates (Difficulty, Discrimination, Guessing) of Mathematics Items

	Discrimination. 'a'	Difficulty 'b'	Guessing 'c'	Chi -square Probability
1.	0.343	-0.125	0.203	0.9113
2.	0.277	-1.208	0.265	0.3985
3.	0.978	0.802	0.495	0.8257
4.	1.445	1.484	0.446	0.2849
5.	0.836	0.971	0.321	0.2876
6.	0.533	0.117	0.112	0.0824
7.	0.937	1.157	0.423	0.5686
8.	0.466	0.747	0.177	0.2311
9.	0.730	0.702	0.135	0.5425
10.	0.698	0.880	0.253	0.5670
11.	0.784	1.827	0.379	0.0161
12.	0.344	0.583	0.140	0.0001
13.	0.457	0.063	0.181	0.0441
14.	0.656	1.206	0.316	0.3696
15.	0.376	0.166	0.197	0.2889

Items	Discrimination	Difficulty	Guessing	Chi-square Prob.
16.	0.289	-0.197	0.187	0.0114
17.	0.330	-0.319	0.225	0.2943
18.	0.702	0.836	0.308	0.3599
19.	0.794	0.761	0.332	0.8800
20.	0.644	1.106	0.332	0.9721
21.	0.617	0.863	0.311	0.3164
22.	0.548	0.726	0.334	0.2602
23.	0.828	0.746	0.325	0.4741
24,	0.396	0.022	0.167	0.0378
25.	0.960	1.106	0.493	0.3141
26.	1.010	0.682	0,321	0.5754
27,	0.546	1.333	0.500	0.0273
28.	1.298	0.950	0.431	0.8826
29.	0.437	0.272	0.173	0.3424
30.	0.917	0.648	0.346	0.8460
31.	0.520	0.461	0.171	0.1741
32.	1.098	0.955	0.312	0,0096
33	0.406	0.150	0.160	0.0197
34	0.555	1.410	0.467	0.0814
35	0.256	1.192	0.233	0.0252
36	0.445	0.831	0.319	0.7508
37	0.349	0.217	0.252	0.4849
38	0.481	1.230	0.300	0.1775
39	0.355	0.538	0.198	0.3823
40	0.505	0.177	0.206	0.1734
41	0.352	0.039	0.146	0.0023
42	0.522	0.230	0.261	0.3831
43	0.474	0.009	0.174	0.0974
44	0.593	0.624	0.273	0.4262
45	0.596	0.642	0.250	0.6920
46	0.779	0.912	0.393	0.0236
47	0.409	0.302	0.257	0.6361
48	1.365	1.127	0.500	0.8930
49	1.260	0.968	0.500	0.0368
50	0.670	1.005	0.386	0.0674
51	1.226	1.032	0.421	0.1112
52	0.388	0.178	0.184	0.2199
53	1.023	0.825	0.338	0.1537
54	0.767	1.383	0.383	0.2338
55	1.065	0.760	0.417	0.8437
56	0.376	0.387	0.181	0.8294
57	0.631	1.604	0.406	0.0511
58	0.350	2.361	0.264	0.4871
59	0.376	0.065	0.194	0.0018
60	0.404	0.765	0.247	0.1406
61	0.413	0.636	0.264	0.5693
62	0.329	1.049	0.212	0.0854
63	0.480	0.601	0.278	0.4965
64	0.389	-0.270	0.151	0.0200
65	0.758	0.485	0.280	0.4071
66	0.751	0.790	0.264	0.7643
67	0.588	1.006	0.354	0.9017

Items	Discrimination	Difficulty	Guessing	Chi-square
68	0.928	0.873	0.361	0.9496
69	0.579	0.923	0.320	0.6110
70	0.406	0.060	0.190	0.3871
71	1.322	1.189	0.387	0.6093
72	0.797	1.009	0.500	0.0632
73	0.988	1.353	0.465	0.2914
74	0.999	1.046	0.466	0.5395
75	0.421	0.634	0.219	0.1480
76	0.682	0.912	0.360	0.6580
77	0.547	1.000	0.269	0.8394
78	0.928	0.958	0.335	0.4786
79	0.427	0.635	0.232	0.2258
80	0.585	1.083	0.383	0.2983
81	0.235	1.617	0.252	0.0605
82	0.308	0.894	0.257	0.2357
83	0.317	0.389	0.240	0.7741
84	0.563	2.006	0.443	0.8755
85	0.416	1.233	0.305	0.9454
86	0.455	0.815	0.216	0.0501
87	0.463	0.662	0.204	0.5092
88	0.508	0.804	0.293	0.0986
89	0.363	-0.129	0.248	0.0291
90	0.547	1.109	0.396	0.8693
91	0.746	1.298	0.431	0.8419
92	0.420	0.992	0.348	0.9116
93	0.391	0.430	0.207	0.9506
94	0.951	1.373	0.497	0.6365
95	0.677	0.996	0.369	0.3835
96	0.804	1.869	0.500	0.2514
97	0.447	0.726	0.344	0.7478
98	0.370	0.250	0.156	0.1450
99	0.920	0.902	0.359	0.9947
100	0.532	0.728	0.238	0.2558

Table 1 reveals that Mathematics difficulty level ranges from -1.208 to 2.006, the discrimination level ranges from 0.235 to 1.445, while the guessing parameter ranges from 0.112 to 0.500.

Table 2

Parameter Estimates for Geography Item

Items	Discrimination	Difficulty	Guessing	Chi-square
12.	0.179	-3.223	0.001	0.393
13.	0.208	-3.603	0.001	0.455
14.	0.210	-2.295	0.001	0.209
15.	0.194	-2.229	0.001	0.013
16.	0.129	-3.735	0.001	0.120
17.	0.141	-3.945	0.001	0.102
18.	0.160	-3.557	0.001	0.420
19.	0.105	-5.743	0.001	0.010
20.	0.147	-4.815	0.001	0.228
21.	0.157	-3.667	0.001	0.258
22.	0.167	-2.971	0.001	0.280
23.	0.179	-3.061	0.001	0.839
24.	0.148	-3.454	0.001	0.344
25.	0.191	-2.711	0.001	0.939
26.	0.174	-3.218	0.001	0.032
27.	0.236	-2.440	0.001	0.055
28.	0.163	-3.798	0.001	0.050
29.	0.214	-2.698	0.001	0.001
30.	0.160	-3.865	0.001	0.057
31.	0.172	-3.059	0.001	0.731
32.	0.274	-1.779	0.001	0.037
33.	0.252	-1.945	0.001	0.037
34.	0.218	-2.500	0.001	0.155
35.	0.154	-3.339	0.001	0.042
36.	0.169	-3.147	0.001	0.013
37.	0.180	-2.919	0.001	0.004
38.	0.151	-3.864	0.001	0.015
39.	0.330	-1.813	0.001	0.009
40.	0.380	-1.727	0.001	0.259
41.	0.333	-1. 732	0.001	0.033
42.	0.281	-1.558	0.001	0.029
43.	0.313	-1.436	0.001	0.179
44.	0.308	-1.621	0.001	0.218
45.	0.330	-1.464	0.001	0.007
46.	0.327	-1.596	0.001	0.287
47.	0.363	-1.635	0.001	0.150
48.	0.400	-1.491	0.001	0.644
49	0.366	-1.525	0.001	0.750
50.	0.357	-1.395	0.001	0.836

51.	0.318	-1.768	0.001	0.114
52.	0.350	-1.432	0.001	0.016
53.	0.359	-1.382	0.001	0.444
54.	0.354	-1.182	0.001	0.853
55	0.291	-1.581	0.001	0.098
56	0.382	-1.296	0.001	0.568
57	0.488	-1.065	0.001	0.800
58	0.431	-1.352	0.001	0.060
59	0.433	-1.322	0.001	0.104
60	0.541	-1.067	0.001	0.104
61	0.428	-0.899	0.001	0.206
62	0.505	-0.894	0.001	0.140
63	0.403	-0.772	0.001	0.022

Items	Discrimination	Difficulty	Guessing	Chi-square
64	0.382	-3.946	0.001	0.005
65.	0.157	-3.231	0.001	0.296
66.	0.212	-2.544	0.001	0.438
67.	0.234	-2.433	0.001	0.183
68.	0.185	-3.455	0.001	0.356
69.	0.211	-2.575	0.001	0.757
70.	0.208	-2.903	0.001	0.019
71.	0.230	-2.407	0.001	0.437
72.	0.175	-3.141	0.001	0.070
73.	0.159	-3.056	0.001	0.057
74.	0.162	-3.091	0.001	0.282
75.	0.178	-2.992	0.001	0.548
76.	0.246	-1.711	0.001	0.309
77.	0.238	-1.931	0.001	0.645
78.	0.194	-2.866	0.001	0.552
79.	0.191	-2.912	0.001	0.184
80.	0.207	-3.043	0.001	0.012
81.	0.192	-3.236	0.001	0.012
82.	0.186	-3.283	0.001	0.016
83.	0.155	-3.991	0.001	0.114
84.	0.176	-3.096	0.001	0.367
85	0.176	-3.392	0.001	0.452
86.	0.203	-2.673	0.001	0.578
87.	0.179	-2.276	0.001	0.278
88.	0.156	-1.023	0.001	0.479
89.	0.186	-2.996	0.001	0.050
90.	0.134	-4.185	0.001	0.205
91.	0.177	-3.110	0.001	0.824
92.	0.138	-3.759	0.001	0.173
93.	0.150	-3.395	0.001	0.062
94.	0.136	-3.521	0.001	0.299
95.	0.162	-3.179	0.001	0.059
96.	0.240	-2.103	0.001	0.025
97.	0.286	-1.833	0.001	0.763
98.	0.196	-3.303	0.001	0.864
99.	0.177	-3.540	0.001	0.635
100.	0.148	-4.430	0.001	0.097

Table 2 shows that the difficulty levels of Geography items ranges from -4.85 to 3.487, the discrimination index ranges from 0.105 to 0.896 while the guessing parameter ranges from 0.001 to 0.235. The table reveals that there is no Geography item with negative discrimination index or guessing index and that only item 1 has guessing parameter greater than 0.001.

Research Question 2

What are the item parameter estimates of the calibrated items in the Mathematics and geography prototype item banks?

Tables 3 and 4 present the extracts of the estimates of the ability of the testees in mathematics and geography. The tables show the candidates' numbers, the number of items tried, the total number of items the candidate got right, the ability estimates and the standard error. The extracts were given because it will be very cumbersome to present the abilities of one thousand two hundred (1, 200) testees in an article of this nature. The extracts in Table 3 show that the estimate of mathematics abilities ranges from -1.32 to 2.45. It also shows that the average mathematics ability is 0.25 with an average standard error of 0.21. The average Mathematics score is also found to be 52.2 with a standard deviation of 12.4. The extracts in table 4 shows that the abilities of the students in Geography ranges from -0.53 to 2.45 and the average ability is 0.96 with mean standard error of 0.23. The table also shows an average total score of 7l.6 and a standard error of 7.4.

Table 3
Mathematics Ability Estimates

Candidate's No	**No. Tried**	**No. Right**	**Ability**	**Standard Error**
124	100	92	2.45	0.37
193	100	92	2.45	0.37
474	100	91	2.42	0.36
566	100	91	2.42	0.36
846	100	91	2.42	0.36
113	100	91	2.42	0.36
233	100	91	2.42	0.36
148	100	90	2.40	0.33
186	100	90	2.40	0.33
242	100	90	2.40	0.33
290	100	90	2.40	0.33
520	100	89	2.36	0.31
183	100	89	2.36	0.31
"	"	"	"	"
		"		"
"	"			"
"		"	"	
			"	
285	100	31	-0.84	0.24
803	100	30	-0.86	0.22
830	100	30	-0.86	0.22
833	100	30	-0.86	0.22
1160	100	30	-.086	0.22
288	100	28	-0.93	0.23
372	100	27	-0.99	0.23
349	100	26	-1.04	0.23
518	100	26	-1.04	0.23
1171	100	25	-1.09	0.23
310	100	21	-1.32	0.25
Mean		**55.2**	**0.25**	**0.21**
S.D		**12.4**	**0.56**	**0.02**

Table 4
Geography Ability Estimate

Candidate's No	No. Tried	No. Right	Ability	Standard Error
49	100	92	2.45	0.37
291	100	91	2.32	0.35
335	100	91	2.32	0.35
1196	100	91	2.32	0.35
40	100	89	2.09	0.32
321	100	89	2.09	0.32
606	100	89	2.09	0.32
900	100	89	2.09	0.32
1191	100	89	2.09	0.32
48	100	88	2.00	0.31
51	100	88	2.00	0.31
300	100	88	2.00	0.31
"	"	"	"	"
"	"	"	"	"
"	"	"	"	"
"	"	"	"	"
"	"	"	"	"
"	"	"	"	"
"	"	"	"	"
"	"	"	"	"'
"	"	"	"	""
692	100	50	0.00	0.20
782	100	49	-0.04	0.20
898	100	49	-0.04	0.20
1110	100	49	-0.04	0.20
1152	100	49	-0.04	0.20
1163	100	49	-0.04	0.20
1185	100	49	-0.04	0.20
952	100	47	-0.12	0.20
295	100	46	-0.16	0.20
508	100	46	-0.16	0.20
269	100	41	-0.36	0.20
283	100	37	-0.53	0.21
Mean		**71.6**	**0.96**	**0.23**
S. D.		**7.4**	**0.38**	**0.02**

Discussion

Findings from the mathematics item parameter estimates indicate that the items are difficult, with an average difficulty level of the items greater than 0.5 logits. The higher the logits, the more difficult an item is. Only four items (Items 1, 2, 16 and 17) were found to be very easy with difficulty values of -0.13, -1.21, -0.20 and -0.32 respectively and this represents only 4% of the total items. Although mathematics items were difficult, their discrimination values which ranged from 0.277 to 1.45 were in order when compared to that of Geography. This is because items with high discriminating powers contribute more to measurement precision than items with low discrimination value (Nworgu & Agah, 2012; Ojerinde, Popoola & Onyeneho, 2012). In a simple logistic model, discrimination index explains the contribution of each item to the assessment of ability. Mathematics Items 4, 26, 28, 32, 48, 49, 51, 55, and 71 contributed more to measurement precision than other items.

Mathematics item parameter estimates also indicate that only three items have discrimination index less than 0.3 and only one item has difficulty index greater than 2.95, the bench mark for the rejection of an item (Baker, 2001; Ojerinde, 2013). However, some mathematics items have "c" (guessing parameter) that is greater than 0.4 which can be corrected by changing their position in the test. These high guessing parameters could be as a result of the fact that mathematics items were difficult and students resorted to guessing. However, only 15 mathematics items did not fit 3-parameter model because their chi-square probability was less than .05 while 85 items fitted the model.

In the same vein, geography parameter estimates show that Geography items were with a negative average difficulty level. The difficulty level ranges from -5.74 to 3.05, while the discrimination level ranges from 0.001 to 0.90. It was only Item 1 that has guessing parameter of 0.24, while others have guessing parameters of 0.001. This could be because Geography items were easy and there was no need to guess. There are also variations in the "a" and "c" of mathematics and geography items, because 3-parameter model of IRT was used in the estimation of the parameters. Also, 24 geography items did not fit the 3-parameter model because their chi-square was less than .05.

The mathematics ability extracts in Table 3 show that although mathematics items were found to be difficult, students average score was 52.2 which indicates that students scored above 50 marks on the average with a standard deviation of 12.4. This high average score could be as a result of some exceptional students who might have scored very high. The high standard deviation is an indication that the performance is not well spread. The geography ability extract reveals that many students scored very high in the geography items; this could be because Geography items were found to be simple at this

level of the students. The 71.6 average mark in geography clearly shows that many students scored high and a standard deviation of 7.4 shows that the marks were better spread than that of mathematics. Mathematics and geography extracts indicate that the average abilities of testees on mathematics and geography achievement tests are 0.25 and 0.96 respectively. This low average ability in mathematics is still an indication that mathematics items were difficult at this level of the students while the high average ability in geography indicates that the items were easy at this level of education. The parameter estimates of the calibrated items in the mathematics and geography prototype item banks are shown in Appendices 1 and 2 respectively.

Conclusion and Recommendation

Schools and organisations should ensure that their examination test items are of a quality and secure. These they could achieve by calibrating the items for the examination with modern techniques such as item response theory. The results of JCSPE are used for the improvement of the system and as an indicator of the quality of education offered in Command schools. It is therefore recommended that the management of Command schools should encourage test experts in their schools to use Item Response Theory to estimate item characteristics of their examination test items.

References

Adeleke, J. O. (2007). Identification and effect of cognitive entry characteristics of students learning outcomes in bearing in mathematics. Ph.D thesis. Institute of Education, University of Ibadan.

Akindele, B. P. (2004). The development of an item bank for selection test into Nigerian Universities: An exploratory study. Ph. D thesis. Department of Guidance and Counselling, University of Ibadan, Ibadan.

Asim, A. E. (2007). Examination ethics and school-based assessments in science, technology and Mathematics: A 60 Proceedings of episteme 3 critical concern for Universal Basic Education. *Proceedings of the 9th National Conference of National Association of Evaluators and Researchers. Ago-Iwoye. Nigeria.*

Baghaei, P. (2008). The Rasch model as a construct validation tool. Rasch Measurement Transactions. 22(1), 1144-1146. Available online at http://www.rasch.org/rmt/rmt22a.htm

Baker, F. (*2001). The Basics of item response theory,* ERIC Clearing house on Assessment and Evaluation, University of Maryland, College Park, MD.

Chong, HO Yu, (2010). A simple guide to the item response theory (IRT) and Rasch modeling: http://www.wisdom.com

Eluwa, A. N.; Idowu, E. 0., Abang, B. K. (2011). Evaluation of Mathematics

achievement test: A comparism between classical test theory and item response theory (IRT). *Journal of Education and Social Research*, 1(40), 99- 106.

Falaye F.V. (1995). Comparative effectiveness of delayed and immediate feedback on senior secondary school students' performance in Geography. PhD Thesis. International Centre for Educational Evaluation, University of Ibadan.

Green, K. E., Franton,C. G. (2002). Survey development and validation with the Rasch model. *A paper presented at the International Conference on Questionnaire Development.* 14 November, Evaluation and Testing, Charleston S.

Griffith, S. A. (2005). Assuring fairness in school-based assessment: Mapping the boundaries of teachers' involvement. *Paper Presented at the 31st Annual Conference of International Association for Educational Assessments, 4-9 September, 2005.* Abuja.

Hopkins, T. M. (2004). Gender Issues III Mathematics' achievement in Tennessee: Does rural school location matter? Ph.D. Dissertation, Graduate School. University of Tennessee, Knoxville.

Igbokwe, E. (2004). Development of Mathematics item bank for placement of students into federal unity schools in Nigeria. PhD Thesis International Centre for Educational Evaluation, University of Ibadan.

Linacre, J. M. (2004). Test validity and Rasch measurement: construct, content, etc. *Rasch Measurement Transactions,* 18(1), 970-971. Retrieved from:htpp://www.rasch.org

Njabili, A. F., Abedi, S., Magesse, M. W., Kalole, S. A. (2005). Equity and school-based assessment: The case of Tanzania. *A paper presented at the 31st Annual Conference of International Association for Educational Assessment,* 4-9 *Sept, 2005.* Abuja, Nigeria. www.unesco/oc.unesco.orgieducationleta-report/2003-pdf/chapter3.pdf.

Nworgu, B. G.,&Agah, J. J. (2012). Application of three- parameter logistic model in the calibration of a mathematics achievement test.*Journal of Educational Assessment in Africa,*7.

Ojerinde, D, Popoola, K. Ojo, F., Onyeneho, P. (2012): *Introduction to item response theory, parameter* modeles*, estimation and application. Abuja:* Marielouse Mike Press Ltd.

Ojerinde, D. (2013). Classical test theory (CTT) vs item response theory (IRT): An evaluation of comparability of item analysis results. *A paper presented at the Institute of Education University of Ibadan.23rd* May *2013 lecture series.*

Ojo, G. O. (2006). Relationships between teachers characteristics and students academic performance in secondary schools in Ogun State, Nigeria. *Journal of Educational Focus,*7, 74-86.

Okwilagwe, E. A. (2002). Modem trends in secondary school Geography teaching. In S.O Ayodele (Eds). *Teaching strategies for Nigerian secondary schools, secondary*

pedagogical support. 1: 121-139. A Power House Publication.

Okwilagwe, E. A. (2011). *Teaching and learning secondary school geography in Nigeria* . Ibadan: Stirling-Horden Publishers Ltd.

Rudner, L. 1998. Item banking: Practical assessment. *Research and Evaluation.* 6.4: 72-87.

Rupp, A. A. (2009). Item response modeling with Bilog-MG and Multilog for windows. *International Journal of Testing*. 3(4), 365-384

Thiessen, V., Blasius, J. (2008). Mathematics achievement and Mathematics learning strategies: Cognitive competences and construct differentiation. *International Journal of Educational Research.* 47(6), 362-371.

Thompson, N. A. (2009). Rasch transaction. *Educational Research.* 47(6), 362-371.

Umobong, M. E.(2004). Item response theory introducing objectivity into educational measurement and evaluation in Nigeria. In Omaze A. Afemikhe and J. Gbenga Adewale (Eds.)*Educational research and study group*, In honour of Wole Falayajo.Institute of Education, University of Ibadan, Ibadan, Nigeria.

Vale, C.D. (2006). Computerised item banking. In S.D. Downing&T. M. Haladyna,(Eds.).

The Handbook of Test Development. Routledge.

Appendix 1

BANK 3						BANK 4					
Ability	**Item Para**	**Discr Para**	**Diff**	**Guessing**	**Subset**	**Ability**	**Item Para**	**Discr Para**	**Diff**	**Guessing**	**Subset**
0.13	88	0.508	0.804	0.293	AIg	-0.17	93	0.391	0.430	0.207	Geo
	3	0.978	0.802	0495	Geo		83	0.317	0.389	0.240	S/P
	66	0.751	0.790	0.264	Alg		56	0.376	0.387	0.181	S/P
	60	0.404	0.765	0.247	Geo		47	0.409	0.302	0.257	S/P
	19	0.914	0.761	0.332	N/N		29	0.437	0.272	0.173	N/N
	55	1.065	0.760	0.417	Geo		98	0.370	0.250	0.156	Geo
	8	0.466	0.747	0.177	Alg		42	0.522	0.230	0.261	Geo
	23	0.828	0.746	0.325	N/N		32	0.349	0.217	0.252	Geo
	100	0.532	0.728	0.238	S/P		52	0.388	0.178	0.184	Geo
	97	0.447	0.726	0.344	Geo		40	0.505	0.177	0.206	Alg
	22	0.548	0.726	0.344	Geo		15	0.376	0.166	0.197	N/|N
	9	0.730	0.702	0.351	N/n		33	0.406	0.150	0.160	Geo
To	26	1.010	0.682	0.321	N/N		6	0.533	0.117	0.122	Geo
	87	0.463	0.662	0.204	Geo		59	0.376	0.065	0.194	Alg
	30	0.917	0.648	0.346	S/P		13	0.457	0.063	0.181	N/N
	45	0.56	0.642	0.250	S/P		70	0.406	0.060	0.190	Alg
	61	0.413	0.636	0.264	N/N		41	0.362	0.039	0.146	Geo
	79	0.420	0.635	0.232	Geo		24	0.396	0.022	0.167	N/N
	.75	0.421	0.634	0.219	N/N		43	0.474	0.009	0.174	N/N
	44	0.593	0.24	0.273	Alg		1	0.343	-0.125	0.203	N/N
	63	0.480	0.601	0.278	N/n		89	0.363	-0.129	0.248	S/P
	12	0.344	0.583	0.140	N/N		16	0.289	-0.197	0.187	Geo
	39	0.355	0.538	0.198	Alg		64	0.389	-0.270	0.151	Geo
	65	0.758	0.485	0.280	N/N		17	0.330	-0.139	0.225	Alg
	0-18	31	0.520	0.461	0.171	S/P	1.32	2	0.277	-0.208	0.265	N/N

Joint Command Schools Promotion Examination:Mathematics Item Ban

BANK 1

Ability	Items	Discr Para	Diff Para	Guessing	Subset
2.45	58	1.065	2.361	0.417	N/N
	84	0.563	2.006	0.443	Geo
	96	0.804	1.869	0.500	N/N
	11	0.784	1.827	0.379	S/P
	81	0.235	1.617	0.252	Alg
	57	0.631	1.604	0.406	N/N
	4	1.445	1.484	0.446	Alg
	54	0.767	1.383	0.383	Geo
	94	0.951	1.373	0.497	Alg
	73	0.988	1.353	0.465	Alg
	27	0.546	1.333	0.500	N/N
	91	0.746	1.298	0.431	N/N
	85	0.416	1.233	0.305	Alg
To	38	0.481	1.230	0.300	Alg
	14	0.656	1.206	0.316	N/N
	35	0.256	1.192	0.233	Ale
	71	1.322	1.189	0.387	Geo
	7	0.937	1.157	0.423	Alg
	34	0.555	1.141	0.467	Alg
	48	1.365	1.127	0.500	S/p
	90	0.547	1.109	0.396	Geo
	25	0.960	1.106	0.493	Alg
	20	0.644	1.106	0.332	S/P
	80	0.585	1.083	0.383	S/p
0.57	62	0.329	1.049	0.212	Geo

BANK 2

Ability	Items	Discr Para	Diff Para	Guessing	Subset
0.56	74	0.999	1.046	0.466	Alg
	51	1.226	1.032	0.421	Alg
	72	0.797	1.099	0.500	Geo
	67	0.588	1.006	0.354	N/N
	50	0.676	1.005	0.386	Alg
	77	0.547	1.000	0.269	Alg
	95	0.677	0.996	0.396	Alg
	92	0.420	0.992	0.348	N/N
	5	0.836	0.971	0.321	Alg
	49	1.260	0.968	0.500	S/P
	78	0.928	0.958	0.335	N/N
	32	1.098	0.955	0,312	ALg
	28	1.298	0.950	0.431	N/N
To	69	0.579	0.923	0.320	N/N
	46	0.779	0.912	0.393	N/N
	76	0.682	0.912	0.360	N/N
	99	0.920	0.992	0.359	Geo
	82	0.308	0.894	0.257	Alg
	10	0.698	0.880	0.253	N/N
	68	0.928	0.873	0.361	S/P
	21	0.617	0.863	0.311	Geo
	18	0.702	0.836	0.308	Alg
	36	0.542	0.831	0.319	N/N
	53	1.023	0.825	0.338	Alg
0.14	86	0.459	0.815	0.216	Geo

Appendix 2

Joint Command Schools Promotion Examination: Geography Item Bank

BANK 1

Ability	Item Para	Discr Para	Diff	Guessing	Subset
2.45	1	0.896	3.0487	0.235	Phy
	63	0.403	-0.772	0.001	Reg
	62	0.505	-0.984	0.001	Phy
	61	0.428	-0.899	0.001	Phy
	88	0.156	-0.023	0.001	Phy
	57	0.488	-0.065	0.001	Phy
	60	0.541	-.0067	0.001	Reg
	54	0.354	-0.182	0.001	Phy
	56	0.382	-1.296	0.001	Phy
	59	0.433	-0.322	0.001	Reg
	58	0.431	-0.352	0.001	Phy
To	53	0.359	-0.382	0.001	Phy
	50	0.357	-0.395	0.001	Reg
	52	0.350	-0.432	0.001	Phy
	43	0.313	-1.436	0.001	Reg
	45	0.330	-1.464	0.001	Reg
	48	0.400	-1.491	0.001	Phy
	49	0.366	-1.525	0.001	Reg
	42	0.281	-1.558	0.001	Reg
	55	0.291	-1.581	0.001	Phy
	46	0.327	-1.596	0.001	Reg
	44	0.308	-1.621	0.001	Phy
	47	0.363	-1.635	0.001	Phy
	76	0.246	-1.711	0.001	Phy
1.22	40	0.380	-1.727	0.001	H/E

BANK 2

Ability	Item Para	Discr Para	Diff Para	Guessing	Subset
1.21	41	0.333	-1.732	0.001	H/E
	51	0.318	768	0.001	Reg
	32	0.274	-1.779	0.001	H/E
	39	0.330	-1.813	0.001	H/E
	97	0.286	-1.833	0.001	Phy
	77	0.238	-1.931	0.001	Phy
	33	0.252	-1.945	0.001	Phy
	96	0.240	-2.103	0.001	Reg
	15	0.194	-2.229	0.001	H/E
	87	0.179	-2.276	0.001	Phy
	14	0.210	-2.295	0.001	H/E
To	71	0.230	-2.407	0.001	Phy
	67	0.234	-2.433	0.001	Reg
	27	0.236	-2.440	0.001	Phy
	34	0.218	-2.500	0.001	Reg
	8	0.192	-2.508	0.001	Reg
	66	0.212	-2.544	0.001	Phy
	69	0.211	-2.575	0.001	H/E
	86	0.203	-2.673	0.001	Phy
	29	0.214	-2.698	0.001	Phy
	25	0.191	-2.711	0.001	Phy
	78	0.194	-2.866	0.001	Phy
	70	0.208	-2.903	0.001	Phy
	79	0.191	-2.912	0.001	Phy
0.96	37	0.180	-2.919	0.001	H/E

BANK 3

Ability	Item	Discr Para	Diff Para	Guessing	Subset
0.95	22	0.167	-2.971	0.001	H/E
	75	0.178	-2.992	0.001	Phy
	89	0.186	-2.996	0.001	Phy
	80	0.209	-3.043	0.001	Phy
	73	0.159	-3.056	0.001	Reg
	31	0.172	-3.059	0.001	Phy
	23	0.179	-3.061	0.001	H/E
	74	0.162	-3.091	0.001	Phy
	4	0.174	-3.096	0.001	Phy
	84	0.196	-3.096	0.001	Reg
	91	0.177	-3.110	0.001	Reg
	10	0.169	-3.122	0.001	H/E
	5	0.177	-3.138	0.001	Reg
	72	0.175	-3.141	0.001	Phy
	36	0.169	-3.147	0.001	Phy
	95	0.162	-3.179	0.001	Phy
	26	0.174	-3.218	0.001	Phy
	12	V179	-3.223	0.001	Phy
	65	0.157	-3.231	0.001	Reg
	81	0.192	-3.236	0.001	Reg
	82	0.186	-3.283	0.001	Reg
	98	0.196	-3.303	0.001	H/e
	35	0.154	-3.339	0.001	Reg
	85	0.176	-3.392	0.001	Reg
0.67	93	0.150	-3.395	0.001	Phy

BANK 4

Ability	Item	Discr Para	Diff Para	Guessing	Subset
0.66	24	0.148	-3.454	0.001	H/E
	68	0.185	-3.455	0.001	Reg
	6	0.145	-3.488	0.001	Phy
	9	0.167	-3.515	0.001	Phy
	94	0.136	-3.521	0.001	Reg
	99	0.177	-3.540	0.001	Phy
	18	0.160	-3.557	0.001	H/E
	3	0.156	-3.568	0.001	H/E
	13	0.208	-3.603	0.001	H/E
	21	0.159	-3.667	0.001	H/E
To	7	0.154	-3.716	0.001	Reg
	16	0.129	-3.735	0.001	Phy
	92	V138	-3.759	0.001	Reg
	11	0.166	-3.780	0.001	Reg
	28	0.163	798-3.	0.001	Phy
	38	0.151	-3.864	0.001	Reg
	30	0.160	-3.865	0.001	Phy
	17	0.141	-3.945	0.001	H/E
	64	0.382	-3.946	0.001	Reg
	83	0.155	-3.991	0.001	H/E
	2	0.150	-4.037	0.001	Phy
	90	0.134	-4.185	0.001	Phy
	100	0.148	-4.430	0.001	Reg
	20	0.147	-4.815	0.001	Reg
-0.53	19	0.105	-5.743	0.001	H/E