Development of Social Studies Aptitude Test for Testing Critical Thinking Skills: Implication for the Achievement of Education for Sustainable Development (ESD)

Morrison Omokiniovo Jessa & John Nwanibeze Odili & Patrick Uzor Osadebe

[1] Teezed Psychometric Laboratory, Nigeria
[2,3] Department of Guidance and Counselling, Delta State University, Abraka, Nigeria
Correspondence: Morrison Omokiniovo Jessa, Nigeria
Email: teezednigeria@gmail.com

Abstract

This study developed a social studies aptitude test (SSAT) that captured and measured critical thinking skills for the achievement of education for sustainable development. Eight research questions guided the study. The instrumentation research design was adopted for the study. The population comprised 72,854 Upper Basic School students, with a total of 1,000 sample size selected using simple random and cluster sampling techniques. A 100-item multiple choice Social Studies Aptitude Test (SSAT) was developed by the researcher, which was used as instrument for the study. The data were collated, entered into a computer system and analysed using chi-square goodness of fit, frequency, percentage, Item Characteristics Curve and factor analysis. The findings revealed that most of the items in the SSAT have a good fit; that the test questions measured the skills of enquiry, intellectual, manipulative and societal values; that the test items are distributed according to Upper Basic Education curriculum; and that the Social Studies Aptitude Test is reliable, having obtained 0.89 index. The study also found that all the 100 items measured a single construct; that most of the items (94 out of 100) were either satisfactory (need no revision), good or moderate (needs little or no revision); that most of the items (89 out of 100) were either very easy or easy; and most of the items (73 out of 100) are not susceptible to guessing. Based on the findings, the study concluded that the Social Studies Aptitude Test items are valid and reliable. The study recommended amongst others, that the developed Social Studies Aptitude Test should be used by Social Studies teachers for the assessment of secondary school students, especially during mock examination, in preparation for external examinations. The study has contributed to knowledge by providing a test that measures the objectives of the revised upper basic education curriculum as well as the achievement of Education for Sustainable Development (ESD).

Keywords: Social Studies Aptitude Test, Critical Thinking Skills, Education for Sustainable Development (ESD), Item Response Theory

Introduction

A social studies aptitude test is an assessment tool often used for selection and placement of students in different classes. It is different from other types of tests that can be used to assess social studies performance. These other tests include achievement,

intelligence, sociometric and personality tests. All these tests vary, based on the domain of learning they assess. For example, personality and sociometric tests assess the affective domain; achievement, intelligence and aptitude tests assess the cognitive domain; while observation assesses the psychomotor domain of learning (Osadebe & Jessa). A social studies aptitude test that is designed to assess students in basic education is expected to be tailored according to the curriculum of basic education in the country.

In a dynamic society such as Nigeria where curriculum changes with societal needs, test items must reflect such changes inherent in the curriculum. It is, however, sad to note that since the revision of the basic school curriculum in Nigeria seven years ago (2014), social studies tests that are used in Delta State remain the same as observation has shown. Most of the social studies test items used in the Delta State Basic Education Certificate Examination (BECE) do not reflect the revised curriculum for basic education, especially in the area of acquisition of critical thinking skills. For instance, in 2016 and 2017, out of 60 multiple-choice items assessed by the researcher, 53 of them, representing 88.3% measured recall, which is a lower cognitive order while the remaining 7 (11.7%) measured thinking, which is a higher cognitive order. In 2018, 90.0% of the 60 test items measured recall while only 10.0% measured critical thinking. Lastly in 2019, out of 60 test items, 91.7% of the items measured recall while only 8.3% measured critical thinking skills. This result implies that the majority of the questions used in the Basic Education Certificate Examination in Delta State from 2016 to 2019 measured mere recall of information learnt. This is a far deviation from the objectives of the revised curriculum for basic schools, which emphasise the acquisition of critical thinking skills. Test developers and teachers seem to have failed to reflect the needed changes in the revised curriculum. In line with this observation, there is an urgent need, therefore, to develop a test that will reflect the changes inherent in the revised curriculum.

The researcher has also observed that most of the social studies test items currently used in Delta State BACE are not designed to measure the extent to which students have acquired intellectual, enquiry, manipulative and societal value skills, which are inherent in the revised curriculum for basic education.

The revised curriculum expects that graduates of Basic Education should be able to organize and synthesize knowledge. They should be capable of independent thinking, to interact in situations demanding clear and rational thoughts, and to combine those thoughts with expressive communicative techniques. Students are expected to think logically and to be able to reason with abstract propositions that they will meet in their daily lives. One way of assessing the acquisition of intellectual skill is by measuring the extent to which students can indulge in critical thinking, problem-solving, formal operation and creativity (Donald, 1985). Measurement of intellectual skills is, therefore, the ability of a test item to measure critical thinking, which is represented in the higher-order cognitive domain. Examples of test items that measure intellectual skills are presented below:

1.   The marriage of Mr and Mrs Akpobome was blessed with three children, namely Oke, Blessing and John. What will be your option if you were told to describe the type of Mr Akpobome's family?
        (a)   compound family
        (b)   extended family
        (c)   friend's family
        (d)   nuclear Family
        (e)   parent's family

        The above test item calls for intellectual skills of evaluation as the student can recognise problems (safety) and finding workable means of reaching solutions without compromising other safety issues. There seems to be a poor representation of the above test items in the social studies test items currently used in Delta State. The implication is that products of this system may not be able to think at the critical level of knowledge, cannot think independently, unable to contribute meaningfully to public discourse and may be prone to negative peer influence. When one cannot think, he or she relies on others to think for him or her.

        Enquiry skills involved the generation and validation of knowledge through scientific investigations. According to the United States' National Research Council (1996), enquiry skills include the abilities to conduct and understand scientific enquiry, including "asking questions, planning and conducting investigations, using appropriate tools and techniques to gather data, thinking critically and logically about relationships between evidence and explanations, constructing and analysing alternative explanations, and communicating scientific arguments" (p. 105). Gormally, Brickmant and Lutz (2012) defined enquiry skills as skills related to two major aspects of scientific literacy: a) recognising and analysing the use of methods of enquiry that lead to scientific knowledge, and b) organising, analysing, and interpreting quantitative data and scientific information. Using enquiry skills, students not only gather new knowledge but also develop their understanding of scientific processes and methods which allow them to explore the surrounding world. Students need these skills as they use scientific reasoning and critical thinking to develop their understanding of science. The development of enquiry skills has a great influence on developing mental processes such as critical thinking and decision making (Shahali, Halim, Rasul, Osman & Zulkifeli, 2017). This is the reason the revised curriculum for basic education emphasizes the acquisition of such skills at the upper basic level. Below, is an example of enquiry skills in a social studies test item.

(1)   Between May and November of every year, Nigeria celebrates the following national days:
        (a)   Easter, Independence Day, Children's Day
        (b)   Independence Day, Eid al-Fitr, Workers' Day

(c)   Workers' Day, Democracy Day, Independence Day

(d)   Democracy Day, Eid al-Adha, Independence Day

The above test item can call for skills of enquiry, as the students will be able to connect bits of information which they learn in social studies with holidays observed during the period in questions. Such test items as observed by the researcher, are not well represented in the social studies test items currently in use for BACE in Delta State. Hence, the need to provide social studies aptitude tests that adequately measure the skills of enquiry.

Manipulative skills can best be described as psychomotor skills that relate individual cognitive function with corresponding physical movement. Students who are competent in manipulative skills will have better opportunity concentrating on the development of enquiry skills which involve skills such as observing, classifying, measuring and using numbers, inferring, predicting, communicating, using space-time relationship, interpreting, defining operationally, controlling variables, making hypotheses and experimenting (Johnstone & Al-Shuaili, 2001). Manipulative skills play an important role in scientific enquiry. Developing these skills is a worthy goal for students to attain. Students who lack exposure to manipulative skills will likely be deficient in scientific enquiry, raising questions and finding the solution to these questions. Such students are less likely to indulge in critical thinking, which is highly needed for technological and economic advancement. Given the above, there is, therefore, the need for test questions to contain items that call for manipulative skills among students. Example of such test items include:

(1)   Categorising the following by placing each in its correct sequential order in a family tree, identify which option best suit your answer

(a)   Son-Daughter-Parents-Grandparents

(b)   Grandparents-Parents-Children

(c)   Uncles-Daddy-Mummy-Children-Grandparents

(d)   Aunties-Parents-Grandparents-Children

The above test item can call for the skills of manipulation because it can allow the students to think chronologically, by placing each item in sequential order, allowing them to think of the whole parts and breaking them down into their separate parts. Such items as observed by the researcher, are not adequately represented in the social studies test questions currently in use in Delta State. This has greatly affected the implementation of the revised upper basic curriculum. Hence, the need to develop test items that will measure skills of manipulation adequately.

Social values are moral beliefs and principles that are accepted by the majority to ensure the continuity of society. It signals not what is there but what should be there in a

society in form of moral imperatives. They are accepted as truthful opinions and criteria set for what is expected by society. Societal value skills are needed for the continuity of society. One of the objectives of social studies education is the enthronement of societal values, where individuals are taught to leave with one another, respect diverse views and the dignity of labour. These skills are well represented in the contents of social studies and are, therefore, included in test items currently in use in Delta State, although the inclusion of such items are rather spontaneous. Example of such test items include:

(1)   _____ is a characteristic of negative group behaviour
- (a)   Humility
- (b)   Peace
- (c)   Unity
- (d)   Violence

Due to the above lapses observed in the social studies aptitude test currently being used in Delta State basic education examination, this study developed and validated a social studies aptitude test (SSAT) for basic education using the 3-parameter IRT model. The purpose was to fill the gap left by the dearth of such questions in the Delta State of Nigeria. The proposed test provided tests or improved on the existing test in the following ways:

First, the test captured the various skills identified in the revised curriculum for basic schools in the areas of critical thinking skills, as well as intellectual, enquiry, manipulative and societal value skills. This was expected to enable the students to be able to think for themselves, indulge in logical thinking and problem-solving exercise in such a way that they would become useful to themselves, their immediate environment and the larger society.

Secondly, the test was constructed in line with the assumptions of Item Response Theory (IRT), in the areas of monotonicity, unidimensionality, local dependence and invariance. These assumptions helped to improve the quality of the test. The assumption of monotonicity indicates that as the trait level is increasing, the probability of a correct response also increases; the assumption of unidimensionality assumes that there is one dominant latent trait being measured and that this trait is the driving force for the responses observed for each item in the measure; the assumption of local independence assumes that responses given to the separate items in a test are mutually independent given a certain level of ability; while the assumption of invariance indicates that item parameters can be estimated from any group of subjects who have answered the item. Hence, the researcher ensured that (1) test items contained stimulus that challenged the trait levels of the students in the areas of intellectual, enquiry, manipulative and societal value skills; (2) test items only measured social studies trait without adding other traits that introduced extraneous factors to the assessment process (3) no test items acted as a link to other items in the test, in other words, all the items in the test were independent of each other, and (4) data that

were used to calibrate the test were obtained from an adequate and representative sample of the population for which the test was made.

Furthermore, in calibrating the test, the researcher adopted the 3-parameter logistic model. This model sought to predict the probability of a certain response based on the examinee ability/trait level and some parameters which described the performance of the item. These models of IRT included a: the discrimination parameter, an index of how well the item differentiates low from top examinees; typically ranges from 0 to 2, where higher is better, though not many items are above 1.0; b: the difficulty parameter, an index of what level of examinees for which the item is appropriate; typically ranges from -3 to +3, with 0 being an average examinee level; and the c: the pseudo guessing parameter, which is a lower asymptote; typically is focused on 1/k where k is the number of options (Thompson, 2016). These parameters were used to graphically display an item response function (IRF). The unique feature of the 3-parameter model that helped achieve the development and validation of the social studies aptitude test for basic education was its applicability to test items that are susceptible to guessing and the principle of invariance. What differentiates the 3PL from its sisters, the 1PL and 2PL, is that it attempts to model for guessing. This, of course, is highly relevant for multiple-choice items on knowledge or ability assessments. For the fact that the social studies aptitude test that was developed, is a multiple-choice test, the 3-parameter model was, therefore, a great fit for it.

Another advantage of the 3-parameter IRT model is the possibility of comparing the latent traits of individuals of different populations when they are submitted to tests that have certain common items (Bortolotti, Tezza, de Andrade, Bornia & Farias de Sousa, 2013). It also allows for the comparison of individuals of the same population submitted to totally different tests. This is possible because the IRT has the items as its central elements.

Research Questions

The following research questions guided the study:
1. To what extent does the SSAT measure the level of theta (θ)?
2. To what extent does the SSAT measure the skills of enquiry, intellectual, manipulative and societal values?
3. To what extent does the SSAT cover the syllabus for Upper Basic Education in Delta State?
4. To what extent is the SSAT reliable?

Methods
Research Design

This study adopted an instrumentation research design. This design is a good fit for this study because the intention of the researcher was geared towards the development

of a social studies aptitude test for the assessment of Upper Basic School Students.

Population and Sampling

The sample size comprised 1,000 students. The choice of the sample size was based on the recommendation of Lord (1968), that a minimum of 50 items and 1,000 examinees are required to estimate an *a*-parameter with high accuracy. A total of 40 students in each local government area of the state were selected to make a total of 1,000 students. This was done through simple random and cluster sampling techniques. In this case, the schools in each Local Government Area of the state were treated as clusters, such that the researcher randomly selected one school in each Local Government Area to make a total of 25 schools. This was done through a simple random sampling technique of the balloting method. Using this procedure, the researcher wrote the name of all the schools in each local government area on pieces of paper, folded and poured them into a container. He then shuffled them and picked one piece of paper from the container. Schools picked from this process were the selected school in that Local Government Area. This was done for all Local Government Areas until all the 25 schools (one for each Local Government Area) were selected.

The above procedure produced 25 clusters, one for each Local Government Area. For each cluster, the researcher randomly selected one classroom out of the various classrooms in Basic 3. All the students in the selected classroom were used for the study, because they were up to 40 students in the selected school.

Development of Instrument

In line with the objectives of the study, the researcher developed a social studies aptitude test. The test comprised 100 multiple-choice items which were derived from the Basic 3 social studies Syllabus. The syllabus was obtained from the Ministry of Basic and Secondary Education, Asaba, Delta State. The items in the test comprise 5 options; one key and four distracters. After the test development, the test was subjected to scrutiny by the research supervisors, who are experts in Measurement and Evaluation and two in Social Studies. This is to gain their insight as to appropriate matching of the items to the domains and content areas arrangement and appropriateness of the item options. The experts critically reviewed the test and made some suggestions for removal of irrelevant items detected and the addition of more items. They also recommended the rewording of some of the items.

Data Analysis

The chi-square goodness of fit was used to answer research question 1, table of

Specification was used to answer research question 2 and 3, using frequency and percentage. Person separation reliability index and Item Characteristics Curve were used to answer research question 4.

The researcher internally modified the Statistical Package for Social Science (SPSS) version 26, by installing and configuring R extensions and plug-ins as well as integrating the R software into the SPSS. This allowed the SPSS to behave differently from its original purpose of statistical analysis tools. The remodelled SPSS was then used to answer research questions 1. The jmetrik IRT software was used to answer research questions 4. The Microsoft Excel 2016 was used to answer research questions 2 and 3, using a table of specification. The Excel software was used to obtain the percentage distribution of the skills of enquiry, intellectual, manipulative and societal values in the social studies aptitude test for research question 2; and the percentage distribution of the syllabus for Upper Basic Education in Delta State in social studies aptitude test for research question 3.

Results

Research Question 1: To what extent does the SSAT measure the level of theta (θ)?

In answering the above research question, the chi-square goodness of fit was used. The reason for the choice of chi-square goodness of fit is because, it examines the items in the SSAT in order to ascertain if they fit the level of theta (θ) measured by the test. The modified SPSS statistical software was used for the items analysis. The result is presented in Table 3.

Table 3: Item Fit Statistics of the SSAT

S/N	Item	$S-X^2$	Sig.	Remark
1.	SSAT56	1.730	.973	√
2.	SSAT74	2.085	.955	√
3.	SSAT49	2.105	.954	√
4.	SSAT39	2.138	.952	√
5.	SSAT4	2.492	.928	√
6.	SSAT62	3.176	.868	√
7.	SSAT94	3.372	.849	√
8.	SSAT35	3.587	.826	√
9.	SSAT30	4.536	.716	√
10.	SSAT93	5.182	.638	√
11.	SSAT87	5.255	.629	√
12.	SSAT77	5.313	.622	√
13.	SSAT47	5.463	.604	√
14.	SSAT26	5.934	.548	√

15.	SSAT81	6.142	.523	√
16.	SSAT13	6.478	.485	√
17.	SSAT37	6.733	.457	√
18.	SSAT75	6.807	.449	√
19.	SSAT65	7.007	.428	√
20.	SSAT91	7.164	.412	√
21.	SSAT98	7.200	.408	√
22.	SSAT32	7.360	.392	√
23.	SSAT43	7.730	.357	√
24.	SSAT12	7.794	.351	√
25.	SSAT97	7.849	.346	√
26.	SSAT27	7.996	.333	√
27.	SSAT8	8.174	.317	√
28.	SSAT84	8.361	.302	√
29.	SSAT36	8.607	.282	√
30.	SSAT3	8.623	.281	√
31.	SSAT96	8.659	.278	√
32.	SSAT22	8.726	.273	√
33.	SSAT76	8.729	.273	√
34.	SSAT34	8.750	.271	√
35.	SSAT18	8.888	.261	√
36.	SSAT86	9.187	.239	√
37.	SSAT48	9.217	.237	√
38.	SSAT9	9.234	.236	√
39.	SSAT99	9.397	.225	√
40.	SSAT79	9.482	.220	√
41.	SSAT19	9.528	.217	√
42.	SSAT55	9.575	.214	√
43.	SSAT61	9.598	.213	√
44.	SSAT59	9.709	.206	√
45.	SSAT51	9.768	.202	√
46.	SSAT45	9.803	.200	√
47.	SSAT33	10.250	.175	√
48.	SSAT92	10.283	.173	√
49.	SSAT71	10.338	.170	√
50.	SSAT28	10.421	.166	√
51.	SSAT66	10.543	.160	√
52.	SSAT80	10.661	.154	√
53.	SSAT60	10.711	.152	√
54.	SSAT82	11.013	.138	√
55.	SSAT38	11.026	.137	√
56.	SSAT78	11.186	.131	√
57.	SSAT67	11.195	.130	√
58.	SSAT31	11.314	.126	√
59.	SSAT17	11.569	.116	√

60.	SSAT89	11.561	.116	√
61.	SSAT52	11.812	.107	√
62.	SSAT2	12.050	.099	√
63.	SSAT73	12.153	.096	√
64.	SSAT24	12.175	.095	√
65.	SSAT1	12.373	.089	√
66.	SSAT68	12.921	.074	√
67.	SSAT25	12.957	.073	√
68.	SSAT85	12.965	.073	√
69.	SSAT83	13.572	.059	√
70.	SSAT21	13.917	.053	√
71.	SSAT88	14.181	.048	X
72.	SSAT29	14.302	.046	X
73.	SSAT23	14.408	.044	X
74.	SSAT5	14.826	.038	X
75.	SSAT72	14.822	.038	X
76.	SSAT11	15.450	.031	X
77.	SSAT70	15.429	.031	X
78.	SSAT58	15.622	.029	X
79.	SSAT20	16.049	.025	X
80.	SSAT40	15.968	.025	X
81.	SSAT7	16.211	.023	X
82.	SSAT90	16.469	.021	X
83.	SSAT10	16.762	.019	X
84.	SSAT41	17.085	.017	X
85.	SSAT95	17.868	.013	X
86.	SSAT42	18.033	.012	X
87.	SSAT53	18.138	.011	X
88.	SSAT16	18.786	.009	X
89.	SSAT54	18.664	.009	X
90.	SSAT15	21.905	.003	X
91.	SSAT44	23.217	.002	X
92.	SSAT63	23.298	.002	X
93.	SSAT46	24.317	.001	X
94.	SSAT6	44.279	.000	X
95.	SSAT14	28.737	.000	X
96.	SSAT50	26.870	.000	X
97.	SSAT57	51.756	.000	X
98.	SSAT64	27.685	.000	X
99.	SSAT69	26.724	.000	X
100.	SSAT100	26.768	.000	X

Key: √ = Good Fit; X = Not Good Fit

Criterion = $p > 0.05$

As shown in Table 3, the p-value ranged from 0.000 to 0.97. Items with p-value greater than 0.05 are regarded as having a good fit in the overall model while items with p-value less than 0.05 are regarded as having no good fit in the overall model. Based on this criterion, out of a total of 100 items, 70 have a good fit in the overall model while 30 did not have a good fit. A total of 70 items final items were therefore selected for final administration. To better visualise the overall model fit of the test, the distribution of the theta estimates for all calibrated items is shown in Fig 4.
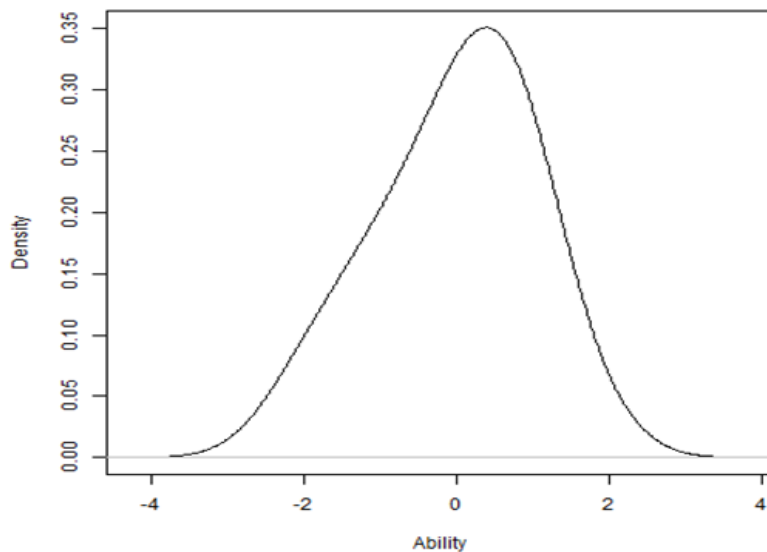
Figure 1: Theta estimates for all calibrated items

Fig 1 shows the graphical representation of how much information the test is providing at each level of theta. In this case, the figure provides satisfactory information over the ability trait range since it takes the shape of a normal distribution curve.

Research Question 2: To what extent does the SSAT measure the skills of enquiry, intellectual, manipulative and societal values?

In answering the above research question, a Table of Specification was used, using frequency and percentage. The table of specification is a two-dimensional table in which the vertical axis contains the content of the course and the horizontal axis contains the behaviour process to be measured. In this table, the vertical axis contains the content area of the Social Studies Aptitude Test while the horizontal axis contains the objectives of enquiry, intellectual, manipulative and societal values. The Microsoft Excel 2016 professional edition was used in the analysis. The result is presented in Table 4.

Table 4: Percentage distribution of questions according to the objectives of enquiry, intellectual, manipulative and societal values

	Enquiry Skill (26%)	Intellectual Skill (26%)	Manipulative Skill (21%)	Societal Value Skill (27%)	Total 100%
Safety (18%)	6	5	3	4	18
Family (18%)	4	6	4	4	18
Marriage (17%)	6	5	5	1	17
Group Behaviour (40%)	8	9	8	15	40
Drug Abuse & Trafficking (2%)	0	0	0	2	2
Human Trafficking (5%)	2	1	1	1	5
Total	26	26	21	27	100

As shown in Table 4 enquiry skills had a total of 26 items, representing 26%, intellectual skills had a total of 26 items, representing 26%, manipulative skill had a total of 21 items, representing 21% while societal values had a total of 27 items, representing 27% of the 100 items in the Social Studies Aptitude Test (SSAT).

Research Question 3: To what extent does the SSAT cover the syllabus for Upper Basic Education in Delta State?

In answering the above research question, a Table of Specification was used, using frequency and percentage. The table of specification is a two-dimensional table in which the vertical axis contains the content of the course and the horizontal axis contains the behaviour process to be measured. In this table, the vertical axis contains the content area of the Social Studies Aptitude Test while the horizontal axis contains the behavioural objectives measured by the test in the Bloom's taxonomy of education objectives. The Microsoft Excel 2016 professional edition was used in the analysis. The result is presented in Table 5.

Table 5: Percentage distribution of questions across the topics in the curriculum

Subject Content Area	Behavioural Objectives						Total 100%
	Knowledge 9%	Comprehension 8%	Application 13%	Analysis 27%	Synthesis 12%	Evaluation 31%	
Safety (18%)	2	2	2	5	1	6	18
Family (18%)	1	2	2	4	4	5	18
Marriage (17%)	3	2	3	3	2	4	17
Group Behaviour (40%)	3	2	4	12	5	14	40
Drug Abuse & Trafficking (2%)	0	0	1	1	0	0	2
Human Trafficking (5%)	0	0	1	2	0	2	5
Total = 100%	9	8	13	27	12	31	100

From the result of Table 5, the items in the test are adequately distributed according to the curriculum of Upper Basic Education especially in the higher order thinking skills. This means that the test measure thinking in the higher order thinking.

**Research Question 4:** To what extent is the SSAT reliable?

In answering the above research question, the data obtained were subjected to item analysis using Person separation reliability index and Item Characteristics Curve. The Jmetrik psychometric software was used for the analysis. The result is presented in Table 6.

Table 6: Summary Statistics for the Total Scores indicating person separation reliability of the SSAT

Test	Items	Alpha	Mean	SD	Skew	Min	Median	Max	IQR
Full Test	100	0.89	19.88	3.34	-0.05	9.00	20.00	31.00	4.00

As shown in Table 6, the Alpha value is 0.89 which tends towards 1.00, indicating a strong reliability. Therefore, the reliability of the SSAT instrument is high when compared within the framework of IRT as indicated by the Alpha value is 0.89. In order to visualise the data, fig 5 has been provided, which capture the Item Characteristics Curve

(ICC). The ICC 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 X-axis is in number-correct units. In this case ICC will predict 89% or its equivalent of the score of each examinee on the SSAT.
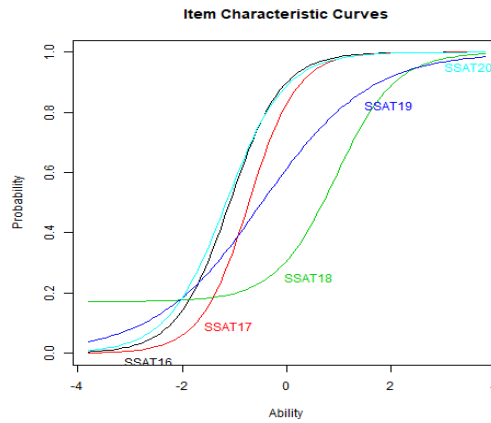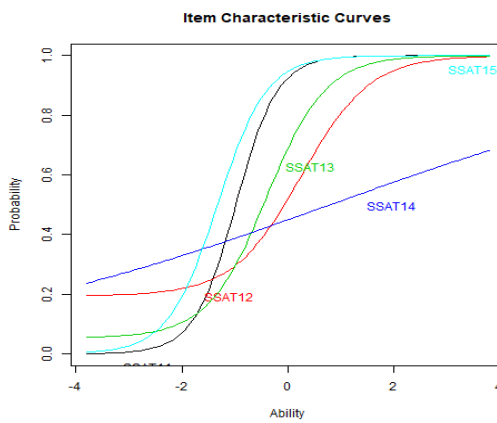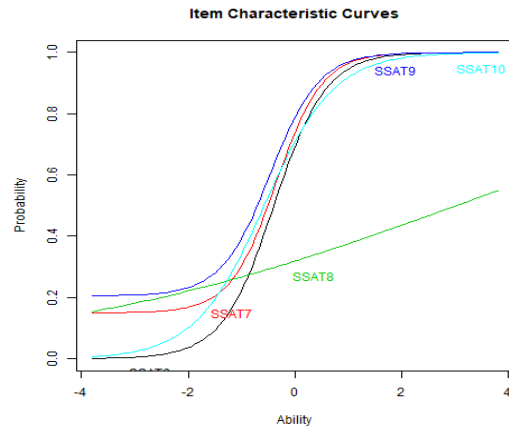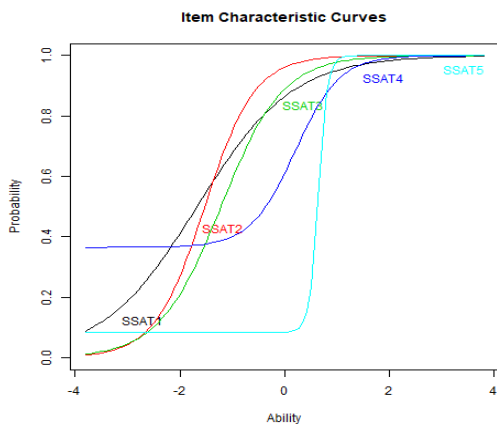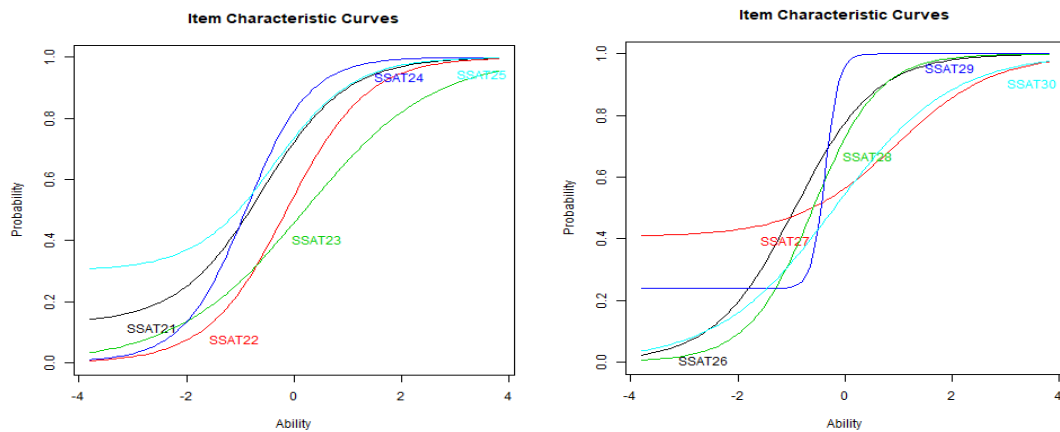
Fig 2: Item Characteristics Curve for the first 30 items of the SSAT
Discussion

Level of Theta (θ) Measured by the Social Studies Aptitude Test (SSAT)

The first finding revealed that most of the items in the SSAT have a good fit. The finding further showed that at of the 100 test items, 70 had a good fit while 30 did not. One of the methods used in estimating the theta level of students that is measured by tests, is Maximum Likelihood Estimation (MLE). It is used in estimating the parameters of an assumed probability distribution, given some observed data. This is achieved by maximizing a likelihood function so that, under the assumed statistical model, the observed data is most probable.

When a parametric model is fitted to data, item-level fit is usually assessed to guide item revision/deletion. To evaluate item fit, numerous statistical procedures have been introduced in IRT literature (Chon, Lee, & Dunbar, 2010; Haberman, Sinharay, & Chon, 2013; LaHuis, Clark, & O'Bruen, 2011; Li & Rupp, 2011; Ranger & Kuhn, 2012; Wang, Shu, Shang, & Xu, 2015). Among them, chi-square-based item fit indices (i.e., Q_1, G^2, S-X^2, and S-G^2) are the most popular family of statistical indices and have been used to examine model misspecification under dichotomous and/or polytomous items (Chon et al., 2010; Wang et al., 2015).

The aforementioned finding shows that the test is generally within the testees' abilities, which explains why the students' overall test score is so high. The testees have a good chance of answering the bulk of the questions correctly. This conclusion supports Thompson's (2009) main premise of IRT, which states that the likelihood of a correct or keyed answer is a function of an underlying attribute or ability, indicated by the Greek letter theta (θ) with a scale typically depicted as ranging from -3 to 3, with 0.0 representing average ability.

Several researchers have recommended and used Maximum Likelihood Estimation for the development of assessment tools. According to Han (2016), because of its capacity to give unbiased estimates, the Maximum Likelihood Estimation (MLE) method is one of the most commonly utilized methods. In their study of the development and standardization of an agricultural science achievement test for senior secondary school students in Taraba State, Nigeria, Egunsola, Denga, and Pev (2014) used the Maximum Likelihood estimation technique to analyze their test and found that the test items had high validity based on a one-parameter model of item response theory.

The findings are also consistent with Oku and Iweka (2018), who used the Maximum Likelihood Estimation Method to perform item analysis on each item and discovered that 99 of the test items fit the One-Parameter Model (1-PLM). The findings back up the findings of Ani (2014), who used the maximum likelihood estimate approach of BILOG-MG computer programming to analyze the data and discovered that all 50 Economics test items passed item calibration.

Distribution of Questions According to the Objectives of Enquiry, Intellectual, Manipulative and Societal Values

The second finding revealed that the test questions measured the skills of enquiry, intellectual, manipulative and societal values. The result further showed that enquiry skills had a total of 26 items, representing 26%, intellectual skills had a total of 26 items, representing 26%, manipulative skill had a total of 21 items, representing 21% while societal values had a total of 27 items, representing 27% of the 100 items in the Social Studies Aptitude Test (SSAT).

The above finding suggests that the Social Studies Aptitude Test adequately measured the objectives of the Revised Upper Basic Education Curriculum. The revised 9-Year Basic Education Curriculum is based on the philosophy that every learner who has completed 9 years of basic education should have acquired appropriate levels of literacy, numeracy, manipulative, communicative, and life skills, as well as the ethical, moral, and civic values required for laying a solid foundation for life-long learning and as a foundation for scientific and reflective thinking. Students' critical thinking abilities must be fostered as part of the curriculum, which will aid in the achievement of educational goals. Students must be developed in the areas of intellect, enquiry, manipulation, and societal value skills in order to achieve this.

The finding also suggests that the Social Studies Aptitude Test can serve as a useful tool for the achievement of Education for Sustainable Development (ESD). This is because one of the roles of ESD is to help people develop the attitudes, values, skills and knowledge to make informed decisions. When the test is applied on the students, there is a likelihood that overtime, they will be able to possess intellectual, enquiry, manipulative and societal

value skills.

The above finding is in consistent with the Nigerian Educational Research and Development Council's (NERDC, 2013) assertion that the Nigerian Basic Education School Curriculum (BESC) was developed to ensure that learners at the Basic Education level receive well-rounded education capable of competing favourably anywhere in the world in terms of knowledge, skills, techniques, values and aptitude. The finding also agrees with Sarita (2017), who stated that skills of enquiry is fundamental for the development of higher-order thinking skills as outlined in Bloom's taxonomy.

Distribution of Questions Across the Topics in the Curriculum

The third finding shows that the test items are distributed according to Upper Basic Education curriculum. This was due to high levels of agreement among experts on the percentage weights allocated to the cognitive domain objectives and subject areas of Upper Basic School Social Studies. The result suggests that the table of specifications adequately covers all of the objectives and content categories. This demonstrates that the test has good content validity and may be used to assess Delta State's upper basic education students. The test can be used to predict students' ability in Social Studies during a mock exam or in the classroom.

The above finding is consistent with the finding of Ani (2014). The author developed and validated multiple-choice test in Economics. The result revealed that the test is valid according to the tenets of Item Response Theory of test development. The finding also agrees with Orangi and Dorani (2010), who carried out research to develop a social studies aptitude test for high school students based on item-response theory (IRT). They found that the test was valid and reliable.

Reliability Coefficient of the Social Studies Aptitude Test (SSAT)

The fourth finding revealed that the Social Studies Aptitude Test is reliable, having obtained 0.89 index. Reliability in Item Response Theory (IRT) is quite different from that of Classical Test Theory (CTT), though they share the same implication. For instance, while in CTT, the reliability measure is simply the proportion of true to observed variance, in IRT, reliability is conceived as the person separation reliability or item separation reliability, which is similar to Cronbach alpha in CTT. The range is usually 0–1, with a higher value indicating a higher reliability index. Item separation reliability is the degree to which item difficulties are differentiated.

The above finding is in line with Orangi and Dorani (2010), who carried out research to develop a social studies aptitude test for high school students based on item-response theory (IRT), and found that the constructed forms were of high reliability, they were at the same time acknowledgeable through the analysis based on the classical Method and they were also in accordance with the three–factors of the Item Response Theory.

The finding also agrees with Ezechukwu, Chinecherem, Oguguo, Ene and Ugorji (2020), who determined the psychometric properties of the Economics Achievement Test (EAT) using Item Response Theory (IRT). Two popular IRT models namely, one-parameter logistics (1PL) and two-parameter logistics (2PL) models were utilized just like in the current study. Reliability and validity for each item and for the whole test were established according to the one-parameter and two-parameter logistic models. The finding revealed that the instrument was highly reliable and fit for use.

Conclusion and Recommendations

Based on the findings, it can be concluded that the Social Studies Aptitude Test items are valid and reliable. It has a good psychometric property and can therefore be used for the assessment of Upper Basic School students in the cognitive domain. The test is in line with the ability level of the students in the appropriate level of schooling.

Based on the findings from this study the following recommendations were made:

The developed Social Studies Aptitude Test should be used by Social Studies teachers for the assessment of secondary school students, especially during mock examination, in preparation for external examinations;

The test should be added to the already existing item bank domiciled in the Ministry of Basic and Secondary Education, since the psychometric properties of the test has been shown to be sound;

Implication for the Achievement of Education for Sustainable Development (ESD)

To achieve sustainable national development, the development of the social studies aptitude test was guided by the tenets of education for sustainable development. The role of ESD is to help people develop the attitudes, values, skills and knowledge to make informed decisions. There is, therefore, the need for students to be able to develop critical thinking ability, which will aid in the achievement of the objectives of education. To do this, students need to possess intellectual, enquiry, manipulative and societal value skills.

Intellectual Skills

A commonly recognized intellectual skill is critical thinking, which is the ability to analyse and interpret information to assess its context and validity. In other words, for students to be able to develop and utilise intellectual skills, they require higher-order skills. Hence, the majority of the test items of the social studies aptitude test will call for much higher-order skills. One way of writing tests for higher-order thinking test, needed for the development of intellectual skills is through real-world scenario-based questions. These

are questions that simulate real-world experiences.

Enquiry Skills

These are skills needed for the pursuit of scientific investigation. Questions used in basic education ought to call for these skills. It will enable the students to develop their understanding of scientific processes and methods which allow them to explore the surrounding world. One way of writing test items that call for enquiry skills is to raise questions that require them to use their existing bank of knowledge to solve immediate problems.

Manipulative Skills

These are skills needed by students for the manipulation of objectives around their environment. Such skills as throwing, catching, kicking, rolling, cutting and writing are increasingly needed for scientific enquiry, which is highly needed for technological and economic advancement. In writing the social studies aptitude test items, the test items will call for such skills.

Societal Value Skills

For basic education to help achieve sustainable national development, there is a need for students to develop societal value skills. Such skills are needed to be able to appreciate the environment and work towards developing it. Hence, the test items of the social studies aptitude test will call for these skills.

References

Bloom, B. S. (1956). *Taxonomy of educational objectives.* New York: David Mckay.

Bortolotti, S. L. V., Tezza, R., de Andrade, D. F., Bornia, A. C., & Farias de Sousa Junior, A. (2013). Relevance and advantages of using the item response theory. *Qual. Quant.,* 47, 2341–2360.

Chon, K. H., Lee, W., & Dunbar, S. B. (2010). A Comparison of Item Fit Statistics for Mixed IRT Models. *Journal of Educational Measurement,* 47(3), 318–338. https://doi.org/10.1111/j.1745-3984.2010.00116.x

Donald, J. G. (1985). Intellectual skills in higher education. *The Canadian Journal of Higher Education,* 15(1), 53-68.

Egunsola, A., Denga, L., & Pev, I. (2014). Development and standardization of agricultural science achievement test for senior secondary school students in

Taraba State Nigeria. *Journal of Education and Leadership Development,* 6(2), 72-85.

Ezechukwu, R. I., Chinecherem, B., Oguguo, E., Ene, C. U., & Ugorji, C. O. (2020). Psychometric Analysis of Economics Achievement Test Using Item Response Theory. *World Journal of Education,* 10(2), 59-68.

FGN (2014). *National Policy on Education* (Revised). Federal Ministry of Education. Lagos: NERDC.

Gormally, C., Brickmant, P., & Lutz, M. (2012). Developing a test of scientific literacy skills (TOSLS): Measuring undergraduates' evaluation of scientific information and arguments. *CBE—Life Sciences Education,* 11, 364-377.

Haberman, S. J., Sinharay, S., & Chon, K. H. (2013). Assessing item fit for unidimensional item response theory models using residuals from estimated item response functions, *Psychometrika,* 78(3), 417–440. https://doi.org/10.1007/s11336-012-9305-1

Han, K. T. (2016). Maximum Likelihood Score Estimation Method with Fences for Short-Length Tests and Computerized Adaptive Tests. *Applied Psychological Measurement,* 40(4), 289–301. https://doi.org/10.1177/0146621616631317

Johnstone, A. H., & Al-Shuaili, A. (2001). Learning in the laboratory: Some thoughts from the literature. *Uni. Chem. Ed,* 5(2), 42-91.

LaHuis, D. M., Clark, P., & O'Brien, E. (2011). An examination of item response theory item fit indices for the graded response model. *Organizational Research Methods,* 14, 10-23.

Li, Y., & Rupp, A. A. (2011). Performance of the S-X2 statistic for full-information bifactor models. *Educational and Psychological Measurement,* 71, 986-1005.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading Mass: Addison-Wesley.

National Research Council (1996). *National science education standards.* Washington, DC: The National Academy Press.

Oku, K., & Iweka, F. (2018). Development, Standardization and Application of Chemistry Achievement Test Using the One-Parameter Logistic Model (1-Plm) of Item Response Theory (Irt). *American Journal of Educational Research,* 6(3), 238-257. doi: 10.12691/education-6-3-11

Orangi A. M., & Dorani, K. (2010). Developing a social studies aptitude test for high school students based on item-response theory (IRT). *Journal of Psychological Models and Methods,* 1(1), 1-13.

Osadebe, P. U., & Jessa, M. O. (2018). Development of social studies achievement test for assessment of secondary school students. *European Journal of Open Education and E-Learning Studies,* 3(1), 104-124.

Ranger, J., & Kuhn, J. T. (2012). Assessing fit of item response models using the

information matrix test. *Journal of Educational Measurement,* 49, 247-268.

Ross, J. (1966). An empirical study of a logistic mental test model. *Psychometrika,* 31, 325-340.

Sarita, J. (2017). Inquiry-based learning: Necessary for teachers and students. *International Journal of Applied Research,* 3(6), 163-164.

Shahali, E. H. M., Halim, L., Rasul, M. S., Osman, K., & Zulkifeli, M. A. (2017). STEM learning through engineering design: Impact on middle secondary students' interest towards STEM. *EURASIA Journal of Mathematics, Science and Technology Education,* 13(5), 1189-1211.

Thompson, N. (2016). *Introduction to Item Response Theory.* A paper presented at the CAT/IRT workshop at the University of Brasilia.

Thompson, N. A. (2009). *Ability Estimation with Item Response Theory.* Assessment Systems Corporation.

Wang, C., Shu, Z., Shang, Z., & Xu, G. (2015). Assessing item-level fit for the DINA model. *Applied Psychological Measurement,* 39, 525-538.