Journal of Education and Learning (EduLearn)
Vol. 20, No. 3, August 2026, pp. 1742~1753
ISSN: 2089-9823 DOI: 10.11591/edulearn.v20i3.23410

Item response theory validation of social studies aptitude test

Morrison O. Jessa, Eric A. Oghounu, Patrick U. Osadebe
Department of Guidance and Counselling, Delta State University, Abraka, Nigeria

Article Info

Article history:

Received Sep 26, 2024
Revised Nov 23, 2025
Accepted Jan 31, 2026

Keywords:

3-parameter item response theory
Measurement
Psychometrics
Social studies aptitude test
Validation

ABSTRACT

This study assessed the psychometric properties of the social studies aptitude test (SSAT) using the 3-parameter item response theory model. The four research questions guided the study. A 100-item multiple-choice SSAT was used as an instrument for the study. The data were collated and analyzed using chi-square goodness of fit and factor analysis. The findings revealed 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); 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. The study recommended, among others, that the developed SSAT should be used by social studies teachers for the assessment of secondary school students, especially during mock examinations.

Corresponding Author:

Morrison O. Jessa
Department of Guidance and Counselling, Delta State University
Abraka, Nigeria
Email: jessa.morrison@delsu.edu.ng

1. INTRODUCTION

Recent reforms in Nigeria's basic education curriculum have emphasized the need for assessment instruments that adequately reflect newly introduced learning objectives and competencies. Following the revision of the basic education curriculum between 2014 and 2022, classroom and large-scale assessments are expected to align with contemporary educational priorities, particularly those promoting critical thinking, civic responsibility, and sustainable development competencies [1]. However, evidence drawn from social studies basic education certificate examination questions administered in Delta State between 2016 and 2019 suggests that several test items failed to adequately capture the range of skills specified in the revised curriculum. Such misalignment poses challenges to the realization of curriculum objectives and undermines broader educational goals associated with education for sustainable development (ESD), which advocates the development of knowledge, values, and competencies required for responsible participation in a sustainable society [2]–[5]. In response to this concern, the present study evaluates the psychometric quality of the social studies aptitude test (SSAT) earlier developed by Jessa *et al*. [6].

The study is anchored on item response theory (IRT), a modern measurement framework widely used in educational and psychological assessment to model the interaction between examinees' latent abilities and their observable responses to test items. Unlike traditional measurement approaches that rely primarily on total test scores, IRT explains performance at the item level by estimating the probability that an individual with a given level of ability will respond correctly to a particular item [7], [8]. Within this framework, both examinee ability and item characteristics are conceptualized along a latent continuum, enabling more precise evaluation of item functioning and measurement accuracy [9]–[11].

The development of IRT emerged from efforts by psychometricians in the mid-twentieth century, particularly Frederic Lord and colleagues, who sought to address limitations associated with classical test

theory, including sample dependency of item statistics and test dependency of examinee scores [12]–[14]. Over time, IRT evolved into a comprehensive statistical paradigm capable of producing invariant item parameters and ability estimates when model assumptions are satisfied [15]–[17]. Central to IRT is the item characteristic curve (ICC), a mathematical function describing how the likelihood of a correct response changes as a function of an examinee's latent trait level. As ability increases, the probability of answering an item correctly also increases systematically and predictably [18].

Application of IRT requires several underlying assumptions to ensure meaningful parameter estimation. These assumptions include unidimensionality, which requires that items measure a dominant latent trait; local independence, which assumes responses to items are conditionally independent given ability; monotonicity, indicating that response probability increases with ability; and the absence of excessive speededness effects [19]. When these assumptions are reasonably satisfied, IRT models provide detailed information about item discrimination, difficulty, and susceptibility to guessing.

Three logistic models are commonly applied in IRT analyses for dichotomously scored items: the one-parameter logistic model (1PLM), which estimates item difficulty; the two-parameter logistic model (2PLM), which incorporates both difficulty and discrimination; and the three-parameter logistic model (3PLM), which additionally estimates guessing behavior. The 3PLM model is particularly suitable for multiple-choice assessments because it accounts for the probability that low-ability examinees may obtain correct answers through guessing.

Although numerous studies have examined test construction and validation using IRT within and beyond Nigeria, existing research has largely focused on the validation of psychological scales or achievement tests in other subject areas. Empirical evidence specifically addressing the validation of a social studies aptitude test within the Delta State educational context remains limited. Consequently, the present study seeks to fill this gap by applying the three-parameter IRT model to evaluate the psychometric properties of the SSAT. Establishing the validity and reliability of this instrument is essential for ensuring accurate assessment of students' cognitive competencies and for supporting evidence-based educational decision-making. The following research questions guided the study: i) to what extent does the SSAT meet the requirement of unidimensionality? ii) to what extent does the SSAT meet the requirements of a-parameter? iii) to what extent does the SSAT meet the requirement of a b-parameter? and iv) to what extent does the SSAT meet the requirement of a c-parameter?

2. METHOD

2.1. Design

This study adopted an instrumentation research design. This design is suitable for the study because the researcher's intention was geared towards the assessment of the psychometric properties of the social studies aptitude test.

2.2. Participants selection

A total of 1,000 students participated in the study. The choice of the sample size was based on the recommendation of Novick and Lord [13], 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 names of all the schools in each local government area on pieces of paper, folded them, and poured them into a container. The researcher then shuffled them and picked one piece of paper from the container. Schools picked from this process were the selected schools in that local government area. This was done for all local government areas until all 25 schools (one for each local government area) were selected. The 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 there were up to 40 students in the selected school.

2.3. Measure

SSAT, originally developed and described in detail by Jessa *et al*. [6], served as the instrument for this study. The test was designed to assess cognitive competencies embedded within the upper basic social studies curriculum and was constructed in alignment with curriculum objectives specified by the Ministry of Basic and Secondary Education, Delta State. Since the procedures relating to item generation, content validation, and initial development have been comprehensively reported elsewhere, only features relevant to the present validation study are summarized here. The instrument consists of 100 multiple-choice items

derived from approved social studies syllabus content. Each item contains five response options comprising one correct answer and four distractors structured to minimize response bias and encourage analytical thinking. The format was selected to permit estimation of discrimination, difficulty, and guessing parameters under the IRT framework.

Administration of the instrument was conducted directly by the researchers with the assistance of trained research assistants. Before data collection, permission was obtained from participating school authorities, and respondents were briefed on the purpose of the exercise. The test was administered under standardized conditions, and completed scripts were collected immediately after administration to ensure data integrity.

2.4. Data analysis

The data obtained were collated, coded and entered into a computer system through the SPSS, version 26. Factor analysis using principal component analysis of the varimax method was used to answer research question 1. The a, b and c parameters item response theory dichotomous models were used to answer research questions 2, 3 and 4, respectively. The researcher internally modified the SPSS version 26 by installing and configuring R extensions and plug-ins, as well as integrating the R software into SPSS. This allowed SPSS to behave differently from its original purpose of statistical analysis tools. The remodeled SPSS was then used to answer research questions 2-4. These research questions were answered using three different parameter logistic models, that is, the 1PLM for research question 2, 2-PLM for research question 3, and 3PLM for research question 4.

3. RESULTS AND DISCUSSION

3.1. Results

3.1.1. RQ 1: to what extent does the SSAT meet the requirement of unidimensionality?

In answering the research question, the data obtained from the field were subjected to a factor analysis using principal component analysis of the varimax method. Using the Guttman-Kaiser rule, all factors with eigenvalues greater than 1 were retained as the factor that the items measure. Analysis of the scree plot, as shown in Figure 1, shows the underlying construct or unidimensionality of the SSAT. The SPSS statistical software was used in the analysis.

Figure 1 shows the scree plot for the SSAT. From the Figure 1, a careful examination of the scree plot shows that there is only one construct before the breaking point or elbow joint. This, therefore, succinctly shows the unidimensionality of the underlying construct of the SSAT. All the items measure one construct.

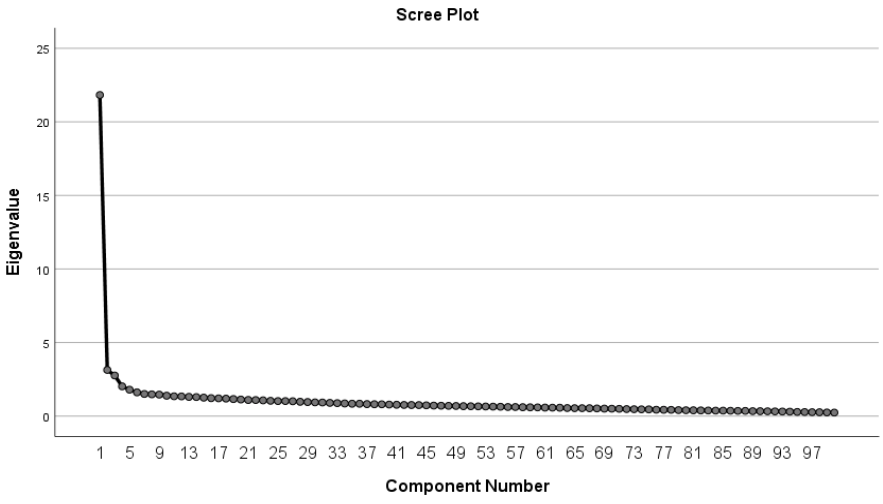

Figure 1. Scree plot for the SSAT

3.1.2. RQ 2: to what extent does the SSAT meet the requirements of a-parameter?

In answering the research question, the modified SPSS statistical software was used to estimate the a-parameter. In analyzing the data, the a-parameter item response theory dichotomous model was utilized.

The aim is to determine the discrimination level of the items in the SSAT. The result is shown in Table 1 (see Appendix).

Table 1 shows the item discrimination index ranged from -0.987 to 9.620, with a higher index indicating a satisfactory item and a lower index indicating a poor item. From the result, items 8, 90, 14, 42, 76 and 65 had a discriminatory index of 0.214, 0.166, 0.139, -0.051, -0.318 and -0.987, indicating poor items that need to be eliminated or revised. Items 58, 32 and 64 had a discriminatory index of 0.624, 0.494 and 0.453, indicating marginal items that need to be revised before they can be used. All other items are either satisfactory (needing no revision), good or moderate (needing little or no revision). The distribution can be visualized in Figure 2.

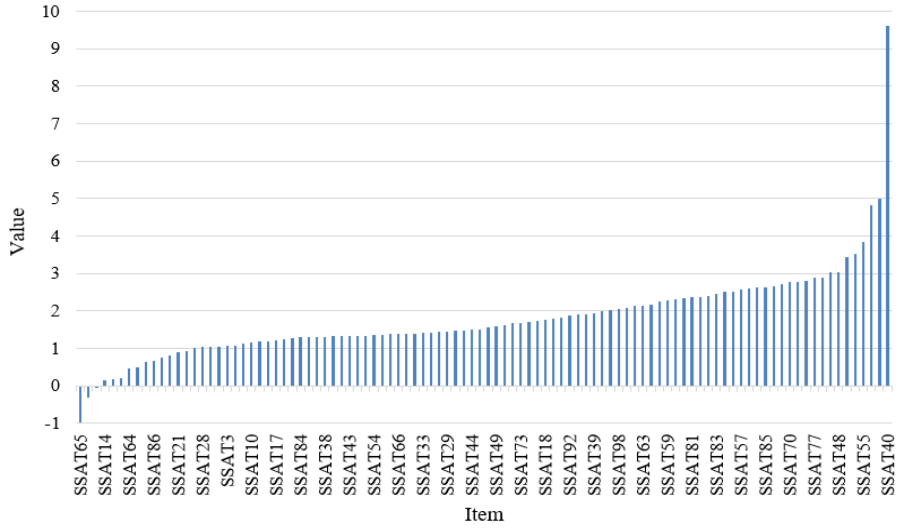

Figure 2. Histogram showing the discriminatory index of the SSAT

3.1.3. RQ 3: to what extent does the SSAT meet the requirement of a b-parameter?

In answering the research question, the modified SPSS statistical software was used to estimate the b-parameter. In analyzing the data, the b-parameter item response theory dichotomous model was utilized. The aim is to determine the difficulty level of the items in the SSAT. The result is shown in Table 2.

Table 2 shows the item difficulty index ranged from -7.702 to 5.236, with a higher index indicating a very difficult item and a lower index indicating a very easy item. From the result, items 40, 45, 38, 53, 8, 90, 84 and 86 had a difficulty index of 2.108, 2.240, 2.464, 2.682, 3.515, 4.192, 4.298 and 5.236, respectively, indicating that the items are very difficult. 47, 95, 97, 99, 94, 87, 18, 14, 78, 46 and 64 had a difficulty index of 1.010, 1.057, 1.142, 1.174, 1.243, 1.248, 1.316, 1.455, 1.594, 1.780 and 1.829, respectively, indicating that the items are difficult. Other items were either very easy or easy. Having obtained a difficulty index that ranged from -7.702 for item 76 to 0.969 for item 69. The distribution can be visualized in Figure 3.

3.1.4. RQ 4: to what extent does the SSAT meet the requirement of a c-parameter?

In answering the above research question, the modified SPSS statistical software was used to estimate the c-parameter. In analyzing the data, the c-parameter item response theory dichotomous model was utilized. The aim is to determine the guessing level of the items in the SSAT. The result is shown in Table 3 (see Appendix).

Table 3 shows the item guessing index ranged from 0.000 to 0.444, with a higher index indicating a guessable item and a lower index indicating not guessable item. The recommended range should be between 0.00 and 0.20. This is because the SSAT has 5 alternatives; a low examinee should have a 1/5=0.20 chance of guessing the correct answer. Since c=0.20 for this 5-alternative item, once the right key is isolated, the examinees will be guessing among the remaining four options. From the result, 27 items had a guessing index above the recommended 0.20, meaning that they are guessable. Other items (73 out of 100) had a guessing index that ranged from 0.000 to 0.192, indicating that they were not guessable. The distribution can be visualized in Figure 4.

Table 2. b-parameter index for the SSAT

Item	Value	Standard error	Z	Remark	Item	Value	Standard error	Z	Remark
SSAT76	-7.702	2.720	-2.832	Very easy	SSAT92	-.087	.129	-.678	Very easy
SSAT65	-4.971	2.668	-1.863	Very easy	SSAT50	-.086	.084	-1.020	Very easy
SSAT2	-1.999	.149	-13.417	Very easy	SSAT74	-.077	.088	-.876	Very easy
SSAT15	-1.939	.161	-12.061	Very easy	SSAT93	-.077	.101	-.768	Very easy
SSAT20	-1.797	.176	-10.202	Very easy	SSAT41	-.053	.084	-.631	Very easy
SSAT3	-1.678	.150	-11.187	Very easy	SSAT89	.011	.144	.078	Easy
SSAT16	-1.369	.105	-12.998	Very easy	SSAT77	.024	.067	.365	Easy
SSAT43	-1.328	.484	-2.743	Very easy	SSAT96	.252	.057	4.396	Easy
SSAT11	-1.326	.102	-12.970	Very easy	SSAT81	.293	.085	3.459	Easy
SSAT1	-1.108	.914	-1.213	Very easy	SSAT48	.373	.077	4.828	Easy
SSAT9	-1.033	.103	-10.056	Very easy	SSAT88	.385	.040	9.676	Easy
SSAT17	-1.010	.100	-10.136	Very easy	SSAT19	.431	.174	2.474	Easy
SSAT29	-.974	.089	-10.918	Very easy	SSAT30	.452	.169	2.672	Easy
SSAT26	-.872	.259	-3.371	Very easy	SSAT85	.469	.053	8.828	Easy
SSAT21	-.828	.610	-1.356	Very easy	SSAT58	.477	.131	3.644	Easy
SSAT52	-.824	.148	-5.579	Very easy	SSAT71	.490	.089	5.529	Easy
SSAT28	-.818	.103	-7.938	Very easy	SSAT4	.502	.099	5.070	Easy
SSAT33	-.814	.084	-9.708	Very easy	SSAT37	.504	.172	2.936	Easy
SSAT60	-.804	.219	-3.670	Very easy	SSAT32	.535	.166	3.219	Easy
SSAT10	-.703	.090	-7.774	Very easy	SSAT59	.663	.074	8.988	Easy
SSAT61	-.675	.107	-6.327	Very easy	SSAT75	.738	.131	5.650	Easy
SSAT36	-.672	.122	-5.499	Very easy	SSAT68	.740	.055	13.526	Easy
SSAT31	-.670	.276	-2.430	Very easy	SSAT27	.783	.161	4.875	Easy
SSAT13	-.633	.090	-7.033	Very easy	SSAT83	.822	.063	13.005	Easy
SSAT35	-.618	.127	-4.864	Very easy	SSAT66	.843	.106	7.977	Easy
SSAT67	-.606	.105	-5.758	Very easy	SSAT100	.908	.324	2.799	Easy
SSAT39	-.587	.159	-3.685	Very easy	SSAT42	.912	10.803	.084	Easy
SSAT55	-.570	.077	-7.388	Very easy	SSAT23	.951	.071	13.392	Easy
SSAT56	-.549	.070	-7.810	Very easy	SSAT51	.956	.126	7.566	Easy
SSAT44	-.545	.071	-7.632	Very easy	SSAT5	.960	.066	14.612	Easy
SSAT79	-.541	.248	-2.183	Very easy	SSAT69	.969	.164	5.926	Easy
SSAT57	-.536	.081	-6.594	Very easy	SSAT47	1.010	.126	7.991	Difficult
SSAT62	-.473	.091	-5.194	Very easy	SSAT95	1.057	.104	10.206	Difficult
SSAT70	-.469	.108	-4.350	Very easy	SSAT97	1.142	.107	10.653	Difficult
SSAT63	-.460	.083	-5.558	Very easy	SSAT99	1.174	.086	13.620	Difficult
SSAT82	-.457	.167	-2.736	Very easy	SSAT94	1.243	.077	16.149	Difficult
SSAT34	-.444	.328	-1.354	Very easy	SSAT87	1.248	.119	10.461	Difficult
SSAT6	-.429	.193	-2.224	Very easy	SSAT18	1.316	.122	10.758	Difficult
SSAT7	-.408	.311	-1.309	Very easy	SSAT14	1.455	.931	1.563	Difficult
SSAT25	-.401	.394	-1.018	Very easy	SSAT78	1.594	.091	17.437	Difficult
SSAT54	-.388	.195	-1.985	Very easy	SSAT46	1.780	.130	13.651	Difficult
SSAT72	-.364	.073	-4.979	Very easy	SSAT64	1.829	.340	5.382	Difficult
SSAT73	-.363	.117	-3.108	Very easy	SSAT40	2.108	.109	19.254	Very difficult
SSAT24	-.235	.142	-1.653	Very easy	SSAT45	2.240	.254	8.809	Very difficult
SSAT91	-.222	.159	-1.395	Very easy	SSAT38	2.464	.425	5.801	Very difficult
SSAT49	-.184	.169	-1.084	Very easy	SSAT53	2.682	.600	4.473	Very difficult
SSAT98	-.153	.104	-1.476	Very easy	SSAT8	3.515	1.243	2.828	Very difficult
SSAT12	-.137	.383	-.358	Very easy	SSAT90	4.192	1.924	2.179	Very difficult
SSAT80	-.125	.089	-1.408	Very easy	SSAT84	4.298	3.102	1.385	Very difficult
SSAT22	-.096	.271	-.354	Very easy	SSAT86	5.236	4.589	1.141	Very difficult

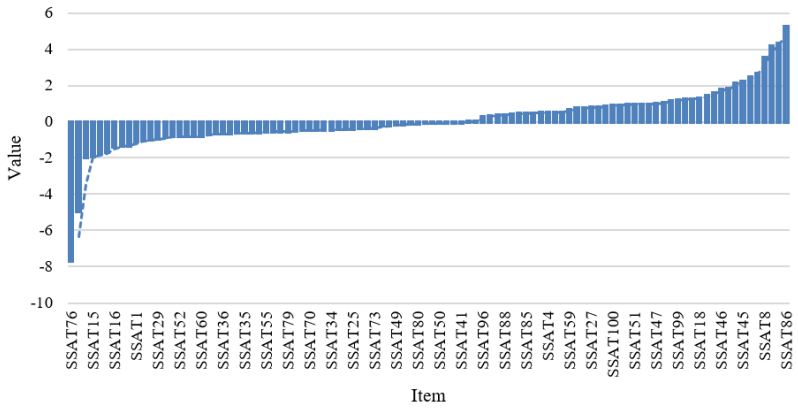

Figure 3. Histogram showing the difficulty index of the SSAT

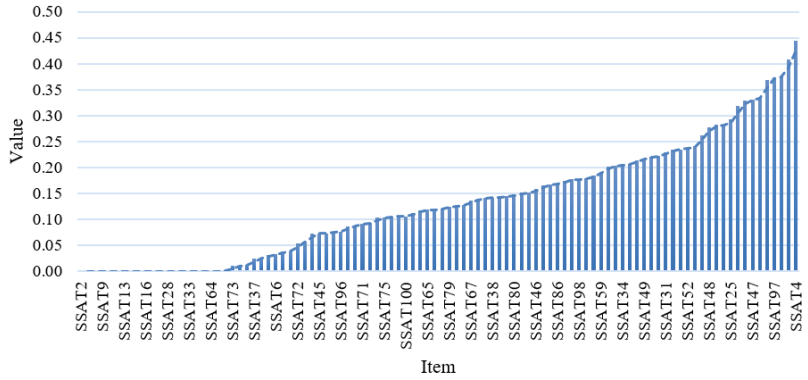

Figure 4. Histogram showing the guessing index of the SSAT

3.2. Discussion

3.2.1. Index of unidimensionality of the items in the SSAT

The analysis of the factor structure revealed that all items of the SSAT measured a single dominant construct, as indicated by the scree plot results. This outcome supports the assumption of unidimensionality required for the application of IRT models. Unidimensional measurement implies that examinees' responses are primarily governed by one underlying latent ability rather than multiple unrelated traits. The presence of a clearly defined dominant factor suggests that the SSAT functions as a coherent instrument capable of assessing a unified domain of competence within social studies. This finding aligns with the theoretical expectation that instruments developed to assess a specific cognitive domain should demonstrate structural consistency across items. Similar validation studies have reported comparable outcomes when factor analytic procedures were employed to examine latent structure before IRT modelling. For example, earlier psychometric investigations confirmed that properly constructed aptitude measures typically exhibit a dominant factor representing the intended construct [20]–[22]. The confirmation of unidimensionality, therefore, provides empirical justification for proceeding with parameter estimation using the three-parameter logistic model.

a. Parameter index for each item in the SSAT

Results relating to item discrimination indicated that the majority of the test items demonstrated acceptable to strong discriminating power. Specifically, most items were classified as satisfactory, good, or moderate, suggesting that they effectively differentiate between examinees with higher and lower levels of ability. In measurement terms, discrimination reflects the sensitivity of an item to variations in the latent trait being assessed. Items with higher discrimination parameters contribute more meaningfully to accurate measurement because they provide clearer distinctions among learners of differing competence levels. The observed pattern suggests that the SSAT possesses adequate internal functioning, with only a small number of items requiring revision or removal. Comparable findings have been reported in previous IRT-based validation studies where well-constructed multiple-choice instruments yielded favorable discrimination indices under the three-parameter model [22]. The predominance of adequately discriminating items, therefore, indicates that the test is capable of producing reliable differentiation among students during assessment.

b. Parameter index (item difficulty parameter) for each item of the SSAT

Analysis of the difficulty parameter showed that a large proportion of the items were categorized as easy or very easy, while a smaller number were identified as difficult or very difficult. Within the IRT framework, item difficulty represents the ability level required for an examinee to have a 50 percent probability of answering an item correctly. The distribution obtained in this study suggests that most items are appropriately aligned with the ability level of the target population. An assessment instrument typically benefits from a range of item difficulties so that learners across varying ability levels can be measured accurately. The presence of both easier and more challenging items enhances score interpretation by covering a broader segment of the ability continuum. Previous psychometric literature similarly notes that difficulty parameters theoretically span a wide continuum but are most informative when distributed across realistic ability levels encountered among examinees [23], [24]. Consequently, the SSAT appears suitably calibrated for evaluating upper basic students without imposing excessive cognitive demand beyond expected competence levels.

c. Parameter index for each item (item discriminatory parameter) of the SSAT

Findings concerning the guessing parameter indicated that the majority of items were not highly susceptible to guessing effects. The c-parameter estimates the probability that an examinee with very low ability may answer an item correctly purely by chance. Because the SSAT consists of five response alternatives, an acceptable guessing probability is expected to remain relatively low. The results demonstrated that most items fell within recommended limits, indicating minimal distortion of scores due to random guessing. Low guessing indices enhance measurement validity because they ensure that correct responses more accurately reflect genuine knowledge rather than chance performance. Earlier studies applying IRT to educational assessments similarly observed that lower c-values correspond to improved item quality and stronger measurement precision [25]–[30]. The limited number of guessable items identified in this study, therefore, suggests that the distractors function adequately and that the instrument largely satisfies IRT assumptions relating to response behavior.

4. CONCLUSION

Based on the findings, it can be concluded that the SSAT items have a good psychometric property and can therefore be used for the assessment of upper basic school students in the cognitive domain. The test is unidimensional and, hence, measures a single trait. Based on the findings, the authors recommended that the developed SSAT should be used by social studies teachers for the assessment of secondary school students, especially during mock examinations, in preparation for external examinations; that 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 be sound; and that most of the items should be modified in terms of their difficulty, discrimination and guessing power so that they will be useful in the assessment of Upper Basic School students. The researchers suggest that a similar study should be conducted in other states, so that more good tests will be available for use. Also, a further study should be carried out on other subject areas.

ACKNOWLEDGMENTS

The authors thank the research assistants, who helped in data collection and those who helped to edit the language of the paper.

FUNDING INFORMATION

The authors personally funded this research from their personal pockets. No funding was obtained from external bodies.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Morison O. Jessa	✓	✓	✓	✓	✓	✓		✓	✓	✓				
Eric A. Oghounu		✓				✓				✓		✓		
Patrick U. Osadebe		✓								✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest as far as the research is concerned.

INFORMED CONSENT

The authors asked for the informed consent of the students who participated in the study. The students willingly provided informed consent. They were assured that the data they provided would only be used for the purpose of research. They were also assured that the process was completely voluntary. No student was coerced into taking the test.

ETHICAL APPROVAL

Ethical approval was obtained first from the Faculty of Education, Delta State University, Abraka, Nigeria. Then, approval was taken from individual principals of the various schools used in the study.

DATA AVAILABILITY

The data that support the findings of this study will be made available on request from the corresponding author, [MOJ]. The data, which contain information that could compromise the privacy of research participants, are not publicly available due to certain restrictions.

REFERENCES

[1] The Nigerian Educational Research and Development Council, *Revised 9-year basic education curriculum*. Abuja, Nigeria: NERDC Press, 2014.

[2] L. M. Ribeiro *et al.*, "Educating for the sustainable development goals through service-learning: university students' perspectives about the competences developed," *Frontiers in Education*, vol. 8, 2023, doi: 10.3389/feduc.2023.1144134.

[3] S. J. Niu, S. Väyrynen, O. P. Malinen, and I. Ruokonen, "Editorial: education for the future: learning and teaching for sustainable development in education," *Frontiers in Education*, vol. 9, 2024, doi: 10.3389/feduc.2024.1476869.

[4] United Nations Educational Scientific and Cultural Organization, *Reimagining our futures together: a new social contract for education*. United Nations Educational, Scientific and Cultural Organization, 2021, doi: 10.54675/ASRB4722.

[5] Department of Economic and Social Affairs, "Transforming our world: the 2030 agenda for sustainable development," United Nations, 2015.

[6] M. O. Jessa, J. N. Odili, and P. U. Osadebe, "Development of social studies aptitude test for testing critical thinking skills: implication for the achievement of education for sustainable development (ESD)," *Canadian Journal of Educational and Social Studies*, vol. 3, no. 4, 2023, doi: 10.53103/cjess.v3i4.163.

[7] N. Myszkowski, *Item response theory for creativity measurement*. Cambridge University Press, 2024.

[9] R. J. De Ayala, *The theory and practice of item response theory*. New York: The Guilford Press, 2022.

[10] K. Hori, H. Fukuhara, and T. Yamada, "Item response theory and its applications in educational measurement Part I: item response theory and its implementation in R," *WIREs Computational Statistics*, vol. 14, no. 2, Mar. 2022, doi: 10.1002/wics.1531.

[11] P. U. Osadebe and M. O. Jessa, "Development of social studies achievement test for assessment of secondary school students," *European Journal of Open Education and E-learning Studies*, vol. 5, no. 1, 2018.

[12] F. Lord, *A theory of test scores*. Richmond: Psychometric Corporation, 1952.

[13] M. R. Novick and F. M. Lord, *Statistical theories of mental test scores*. Boston: Addison-Wesley, 1968.

[14] R. Hambleton and M. Jodoin, "Item response theory: models and features," in *Encyclopedia of Psychological Assessment*, *R. Fernández-Ballesteros*, Ed. Oliver's Yard: SAGE Publications Ltd, 2003, doi: 10.4135/9780857025753.n110.

[15] R. K. Hambleton, "Item response theory: A broad psychometric framework for measurement advances," *Psicothema*, vol. 6, no. 3, pp. 535–556, 1994.

[16] R. K. Hambleton, H. Swaminathan, and J. H. Rogers, *Fundamentals of item response theory*. Thousand Oaks: Sage Publications, Inc., 1991.

[17] C. S. Wells, "Item response theory," in *Assessing Measurement Invariance for Applied Research*. Cambridge University Press. 2021.

[18] A. Gyamfi and R. Acquaye, "Parameters and models of item response theory (IRT): a review of literature," *Acta Educationis Generalis*, 13, no. 3, pp. 68–78, 2023, doi: 10.2478/atd-2023-0022.

[19] K. Sijtsma and L. A. van der Ark, "Advances in nonparametric item response theory for scale construction in quality-of-life research," *Quality of Life Research*, vol. 31, no. 1, pp. 1–9, Jan. 2022, doi: 10.1007/s11136-021-03022-w.

[20] Z. T. Goodman, K. R. Timpano, M. M. Llabre, and S. A. Bainter, "Revisiting the factor structure and construct validity of the cognitive failures questionnaire," *Psychological Assessment*, vol. 34, no. 7, pp. 671–683, 2022, doi: 10.1037/pas0001127.

[21] Ş. E. Mermer, "Educational implications of comparing unidimensional and multidimensional item response theories," *Pegem Journal of Education and Instruction*, vol. 14, no. 3, pp. 103–116, 2024.

[22] P. J. Kpolovie and C. O. Emekene, "Psychometric advent of advanced progressive matrices–smart version (APM-SV) for use in Nigeria," *European Journal of Statistics and Probability*, vol. 4, no. 3, pp. 20–60, 2016.

[23] S. E. Stemler and A. Naples, "Rasch measurement v. item response theory: knowing when to cross the line," *Practical Assessment, Research and Evaluation*, vol. 26, pp. 1–16, 2021, doi: 10.7275/v2gd-4441.

[24] P. J. Kpolovie, *Test measurement and evaluation in education*. Owerri: Springfield, 2014.

[25] S. Noventa, S. Ye, A. Kelava, and A. Spoto, "On the Identifiability of 3- and 4-parameter item response theory models from the perspective of knowledge space theory," *Psychometrika*, vol. 89, no. 2, pp. 486–516, 2024, doi: 10.1007/s11336-024-09950-z.

[26] K. T. Han, "Maximum likelihood score estimation method with fences for short-length tests and computerized adaptive tests," *Applied Psychological Measurement*, vol. 40, no. 4, pp. 289–301, 2016, doi: 10.1177/0146621616631317.

[27] R. E. Eghan, E. Osei-Sarpong, G. E. Awashie, R. N. Borkor, E. Yaokumah, and A. A. N'ganomah, "Item response theory for trait assessment in randomized item pool for computer based test," *Scientific African*, vol. 31, 2026, doi: 10.1016/j.sciaf.2026.e03226.

[28] D. Harris, "Educational measurement issues and practice: comparison of 1-, 2-, and 3- parameter IRT models," *Educational Measurement: Issues and Practice*, vol. 8. no. 1, pp. 35-41, 2005, doi: 10.1111/j.1745-3992.1989.tb00313.x.

[29] B. Akindele, "The development of an item bank for selection tests into Nigerian universities: an exploratory study," University of Ibadan.
[30] F. B. Baker and S. H. Kim. *The basics of item response theory using R*. New York, NY: Springer. 2017.

APPENDIX

Table 1. a-parameter index for the SSAT

Item	Value	Standard error	Z	Remark
SSAT40	9.620	12.574	.765	Satisfactory (no revision required)
SSAT56	4.999	.710	7.045	Satisfactory (no revision required)
SSAT88	4.819	.539	8.934	Satisfactory (no revision required)
SSAT55	3.829	.507	7.549	Satisfactory (no revision required)
SSAT68	3.529	.518	6.811	Satisfactory (no revision required)
SSAT23	3.429	.703	4.875	Satisfactory (no revision required)
SSAT48	3.031	.502	6.039	Satisfactory (no revision required)
SSAT78	3.029	.588	5.153	Satisfactory (no revision required)
SSAT61	2.900	.386	7.511	Satisfactory (no revision required)
SSAT77	2.874	.341	8.420	Satisfactory (no revision required)
SSAT4	2.797	.584	4.788	Satisfactory (no revision required)
SSAT46	2.774	.710	3.909	Satisfactory (no revision required)
SSAT70	2.763	.382	7.235	Satisfactory (no revision required)
SSAT5	2.705	.406	6.656	Satisfactory (no revision required)
SSAT94	2.653	.481	5.515	Satisfactory (no revision required)
SSAT85	2.627	.318	8.262	Satisfactory (no revision required)
SSAT62	2.623	.308	8.526	Satisfactory (no revision required)
SSAT50	2.599	.350	7.426	Satisfactory (no revision required)
SSAT57	2.556	.269	9.490	Satisfactory (no revision required)
SSAT72	2.503	.260	9.625	Satisfactory (no revision required)
SSAT96	2.498	.263	9.489	Satisfactory (no revision required)
SSAT83	2.463	.328	7.514	Satisfactory (no revision required)
SSAT74	2.408	.303	7.947	Satisfactory (no revision required)
SSAT67	2.373	.283	8.372	Satisfactory (no revision required)
SSAT81	2.360	.334	7.062	Satisfactory (no revision required)
SSAT24	2.349	.401	5.860	Satisfactory (no revision required)
SSAT41	2.310	.268	8.635	Satisfactory (no revision required)
SSAT59	2.273	.348	6.528	Satisfactory (no revision required)
SSAT97	2.256	.560	4.030	Satisfactory (no revision required)
SSAT52	2.178	.292	7.452	Satisfactory (no revision required)
SSAT63	2.150	.219	9.817	Satisfactory (no revision required)
SSAT45	2.124	.609	3.486	Satisfactory (no revision required)
SSAT80	2.084	.235	8.879	Satisfactory (no revision required)
SSAT98	2.046	.254	8.069	Satisfactory (no revision required)
SSAT36	2.021	.236	8.546	Satisfactory (no revision required)
SSAT99	1.996	.312	6.386	Satisfactory (no revision required)
SSAT39	1.949	.276	7.054	Satisfactory (no revision required)
SSAT35	1.917	.223	8.592	Satisfactory (no revision required)
SSAT93	1.899	.229	8.308	Satisfactory (no revision required)
SSAT92	1.872	.270	6.941	Satisfactory (no revision required)
SSAT71	1.817	.271	6.700	Satisfactory (no revision required)
SSAT82	1.788	.262	6.822	Satisfactory (no revision required)
SSAT18	1.751	.379	4.620	Satisfactory (no revision required)
SSAT60	1.744	.274	6.370	Satisfactory (no revision required)
SSAT47	1.699	.385	4.411	Good (little or no revision required)
SSAT73	1.671	.193	8.665	Good (little or no revision required)
SSAT27	1.666	.444	3.751	Good (little or no revision required)
SSAT19	1.609	.355	4.537	Good (little or no revision required)
SSAT49	1.591	.246	6.481	Good (little or no revision required)
SSAT91	1.547	.225	6.867	Good (little or no revision required)
SSAT53	1.513	.687	2.203	Good (little or no revision required)
SSAT44	1.505	.118	12.708	Good (little or no revision required)
SSAT95	1.477	.236	6.269	Good (little or no revision required)
SSAT11	1.476	.129	11.462	Good (little or no revision required)
SSAT29	1.435	.120	11.921	Good (little or no revision required)
SSAT16	1.433	.126	11.415	Good (little or no revision required)
SSAT2	1.408	.145	9.698	Good (little or no revision required)
SSAT33	1.407	.116	12.113	Good (little or no revision required)
SSAT30	1.389	.289	4.813	Good (little or no revision required)
SSAT79	1.388	.228	6.086	Good (little or no revision required)
SSAT66	1.385	.231	6.006	Good (little or no revision required)
SSAT6	1.373	.191	7.190	Good (little or no revision required)
SSAT89	1.363	.193	7.078	Good (little or no revision required)

Table 1. a-parameter index for the SSAT (continue)

Item	Value	Standard error	Z	Remark
SSAT54	1.346	.196	6.856	Good (little or no revision required)
SSAT31	1.332	.214	6.212	Moderate (little or no revision required)
SSAT87	1.323	.280	4.721	Moderate (little or no revision required)
SSAT26	1.321	.190	6.954	Moderate (little or no revision required)
SSAT43	1.321	.252	5.245	Moderate (little or no revision required)
SSAT34	1.320	.271	4.874	Moderate (little or no revision required)
SSAT38	1.298	.487	2.666	Moderate (little or no revision required)
SSAT75	1.298	.253	5.136	Moderate (little or no revision required)
SSAT25	1.292	.311	4.153	Moderate (little or no revision required)
SSAT84	1.290	1.491	.865	Moderate (little or no revision required)
SSAT7	1.266	.242	5.228	Moderate (little or no revision required)
SSAT51	1.257	.255	4.933	Moderate (little or no revision required)
SSAT17	1.225	.107	11.419	Moderate (little or no revision required)
SSAT9	1.182	.104	11.317	Moderate (little or no revision required)
SSAT15	1.179	.121	9.766	Moderate (little or no revision required)
SSAT10	1.166	.102	11.465	Moderate (little or no revision required)
SSAT13	1.117	.098	11.348	Moderate (little or no revision required)
SSAT3	1.074	.107	10.027	Moderate (little or no revision required)
SSAT22	1.074	.201	5.355	Moderate (little or no revision required)
SSAT37	1.049	.196	5.363	Moderate (little or no revision required)
SSAT1	1.040	.313	3.322	Moderate (little or no revision required)
SSAT28	1.032	.095	10.883	Moderate (little or no revision required)
SSAT69	1.006	.205	4.914	Moderate (little or no revision required)
SSAT20	.929	.099	9.370	Moderate (little or no revision required)
SSAT21	.906	.203	4.469	Moderate (little or no revision required)
SSAT12	.800	.159	5.036	Moderate (little or no revision required)
SSAT100	.746	.224	3.336	Moderate (little or no revision required)
SSAT86	.671	.893	.751	Moderate (little or no revision required)
SSAT58	.624	.079	7.878	Marginal (needs revision)
SSAT32	.494	.075	6.594	Marginal (needs revision)
SSAT64	.453	.080	5.660	Marginal (needs revision)
SSAT8	.214	.074	2.897	Poor (should be eliminated or revised)
SSAT90	.166	.074	2.251	Poor (should be eliminated or revised)
SSAT14	.139	.068	2.051	Poor (should be eliminated or revised)
SSAT42	-.051	.068	-.757	Poor (should be eliminated or revised)
SSAT76	-.318	.116	-2.734	Poor (should be eliminated or revised)
SSAT65	-.987	.896	-1.101	Poor (should be eliminated or revised)

Table 3. c-parameter index for the SSAT

Item	Value	Standard error	Z	Remark
SSAT2	.000	.011	.009	Not guessable
SSAT3	.000	.004	.006	Not guessable
SSAT8	.000	.007	.018	Not guessable
SSAT9	.000	.000	.001	Not guessable
SSAT10	.000	.001	.003	Not guessable
SSAT11	.000	.000	.002	Not guessable
SSAT13	.000	.001	.003	Not guessable
SSAT14	.000	.021	.017	Not guessable
SSAT15	.000	.000	.002	Not guessable
SSAT16	.000	.000	.001	Not guessable
SSAT17	.000	.001	.004	Not guessable
SSAT20	.000	.001	.004	Not guessable
SSAT28	.000	.003	.007	Not guessable
SSAT29	.000	.000	.001	Not guessable
SSAT32	.000	.002	.005	Not guessable
SSAT33	.000	.000	.001	Not guessable
SSAT44	.000	.000	.003	Not guessable
SSAT58	.000	.002	.005	Not guessable
SSAT64	.000	.000	.002	Not guessable
SSAT76	.000	.000	.001	Not guessable
SSAT90	.001	.046	.025	Not guessable
SSAT73	.010	.048	.201	Not guessable
SSAT43	.011	.256	.042	Not guessable
SSAT42	.013	.272	.047	Not guessable
SSAT37	.024	.064	.379	Not guessable
SSAT22	.025	.103	.242	Not guessable
SSAT36	.031	.058	.535	Not guessable
SSAT6	.033	.083	.394	Not guessable

Note: criterion=c>0.20.

Table 3. c-parameter index for the SSAT (continue)

Item	Value	Standard error	Z	Remark
SSAT12	.039	.120	.327	Not guessable
SSAT63	.039	.033	1.159	Not guessable
SSAT72	.054	.031	1.732	Not guessable
SSAT85	.057	.021	2.721	Not guessable
SSAT26	.073	.113	.645	Not guessable
SSAT45	.073	.013	5.498	Not guessable
SSAT89	.073	.056	1.304	Not guessable
SSAT66	.076	.038	2.013	Not guessable
SSAT96	.077	.023	3.352	Not guessable
SSAT78	.086	.014	6.233	Not guessable
SSAT87	.089	.036	2.477	Not guessable
SSAT71	.091	.037	2.501	Not guessable
SSAT51	.093	.046	2.005	Not guessable
SSAT57	.103	.037	2.816	Not guessable
SSAT75	.103	.050	2.061	Not guessable
SSAT21	.105	.209	.504	Not guessable
SSAT88	.106	.019	5.701	Not guessable
SSAT100	.106	.100	1.057	Not guessable
SSAT69	.112	.051	2.174	Not guessable
SSAT50	.117	.039	3.043	Not guessable
SSAT65	.119	.019	6.388	Not guessable
SSAT95	.119	.028	4.184	Not guessable
SSAT35	.121	.058	2.090	Not guessable
SSAT79	.124	.102	1.222	Not guessable
SSAT91	.127	.065	1.951	Not guessable
SSAT99	.127	.022	5.714	Not guessable
SSAT67	.136	.051	2.675	Not guessable
SSAT83	.139	.022	6.428	Not guessable
SSAT93	.141	.043	3.247	Not guessable
SSAT38	.142	.026	5.405	Not guessable
SSAT84	.143	.019	7.722	Not guessable
SSAT40	.145	.013	11.331	Not guessable
SSAT80	.148	.038	3.892	Not guessable
SSAT94	.151	.021	7.142	Not guessable
SSAT5	.152	.022	6.987	Not guessable
SSAT46	.158	.018	8.941	Not guessable
SSAT54	.166	.074	2.233	Not guessable
SSAT62	.167	.044	3.846	Not guessable
SSAT86	.168	.044	3.857	Not guessable
SSAT41	.175	.037	4.709	Not guessable
SSAT77	.177	.033	5.430	Not guessable
SSAT98	.177	.046	3.878	Not guessable
SSAT7	.180	.116	1.556	Not guessable
SSAT55	.184	.040	4.637	Not guessable
SSAT59	.192	.029	6.588	Not guessable
SSAT74	.201	.040	5.042	Guessable
SSAT92	.204	.054	3.802	Guessable
SSAT34	.206	.122	1.681	Guessable
SSAT81	.207	.036	5.696	Guessable
SSAT82	.214	.071	3.039	Guessable
SSAT49	.217	.065	3.317	Guessable
SSAT56	.221	.038	5.755	Guessable
SSAT53	.222	.023	9.841	Guessable
SSAT31	.230	.104	2.215	Guessable
SSAT68	.234	.023	10.223	Guessable
SSAT30	.235	.060	3.949	Guessable
SSAT52	.237	.070	3.384	Guessable
SSAT18	.241	.030	8.066	Guessable
SSAT39	.262	.069	3.830	Guessable
SSAT48	.278	.034	8.125	Guessable
SSAT60	.282	.090	3.125	Guessable
SSAT61	.282	.052	5.385	Guessable
SSAT25	.292	.130	2.244	Guessable
SSAT23	.318	.026	12.157	Guessable
SSAT19	.329	.058	5.662	Guessable
SSAT47	.331	.038	8.835	Guessable
SSAT70	.336	.049	6.859	Guessable
SSAT24	.368	.055	6.651	Guessable
SSAT97	.373	.031	12.142	Guessable

Note: criterion=c>0.20.

Table 3. c-parameter index for the SSAT (continue)

Item	Value	Standard error	Z	Remark
SSAT1	.375	.260	1.441	Guessable
SSAT27	.408	.048	8.508	Guessable
SSAT4	.444	.035	12.597	Guessable

Note: criterion=c>0.20.

BIOGRAPHIES OF AUTHORS

Morrison O. Jessa is a psychometrician, scholar, and ICT expert with a Ph.D. in Measurement and Evaluation. He is a lecturer of Educational Measurement and Evaluation in the Department of Guidance and Counselling, Delta State University, Abraka Nigeria. He specializes in item response theory, classical test theory, psychometric data forensic, and quantitative research methods. His skills extend to program evaluation, statistical data analysis, and software for investigating test fraud (SIFT). He is also an expert in Oracle Database administration. He can be contacted at email: jessa.morrison@delsu.edu.ng.

Eric A. Oghounu is an associate professor of Guidance and Counselling and a lecturer in the Department of Guidance and Counselling in Delta State University, Abraka, Nigeria. He holds a Ph.D. degree, Master degree and Bachelor degree in Guidance and Counselling. His major area of specialization is behavior modification and personality assessment. He is a member of the Counselling Association of Nigeria (CASSON). He has published in many journals. He can be contacted at email: aeoghounu@delsu.edu.ng.

Patrick U. Osadebe is a professor of Educational Measurement and Evaluation in the Department of Guidance and Counselling, Delta State University Abraka Nigeria. He holds a Ph.D. degree in Educational Measurement and Evaluation. His area of specialization is in test construction and standardization. He can be contacted at email: osadebepu@delsu.edu.ng.