IJAAR PUBLISHING®

International Journal of Research in Education and Sustainable Development | ISSN: **2782-7666**
Vol. 3, Issue 7 (July, 2023) | www.ijaar.org/ijresd

APPLICATION OF ITEM RESPONSE THEORY IN THE VALIDATION OF BASIC SCIENCE TEST OF DELTA STATE BASIC EDUCATION CERTIFICATE EXAMINATION

BY

OYIBORHORO, Akponemubemu Venture[1], Prof. J. N. ODILI[2] and Prof. P. U. OSADEBE[3]
1,2&3 Department of Guidance and Counseling (Measurement and Evaluation), Delta State University, Abraka.

ABSTRACT

The purpose of this study is to apply item response theory of measurement to validate the Basic Science Test of the Delta State Basic Education Certificate Examination. Four research questions guided the study. The study adopted the multiple triangulation research design. The population of the study comprised all public Upper basic school students in Delta State, Nigeria. The sample size comprised 1,000 JSS 3 students in public secondary schools in Delta State. The sampling technique that was used for the study is proportionate stratified, simple random and convenience sampling techniques. The main instrument that was used for the study is the Basic Science Test (BST) that is under validation. Another instrument that was used is the Physical and Health Education Test (PHET) that was used to establish the concurrent validity of the Basic Science Test. The a, b and c parameters item response theory dichotomous models were used to answer research questions 1, 2 and 3 respectively. Pearson product moment correlation coefficient was used to answer research question 4. The findings of the study revealed that majority of the items in the test are moderately difficulty and partially fall in line with the ability level of the students for whom the test was made; that majority of the items in the test discriminated well between higher achievers and low achievers; that majority of the items are not guessable; and that the BST has a very high concurrent validity. Based on the findings, it can be concluded that IRT and CTT can play a complementary role in test development. The study recommended that the validate Basic Science Test should be added to the already existing item pool.

Keywords: IRT, concurrent validity, Test

Published *30/06/2023*
Oyiborhoro, A.Venture, Odili, J. N. and Osadebe, P. U. (2023). Application Of Item Response Theory In The Validation Of Basic Science Test Of Delta State Basic Education Certificate Examination. *International Journal of Research in Education and Sustainable Development*, 3(7), 1-13.

INTRODUCTION

Background to the study

The revised basic education curriculum has grouped Basic Education in Nigeria into three, the Lower Basic, starting at Basic 1 and ending at Basic 3; the Middle Basic, starting at Basic 4 and ending at Basic 6; and the Upper Basic, starting at Junior Secondary 1 and ending at Junior Secondary 3. The upper basic class concludes the completion of Basic Education in the country. In other words, those who graduate at this level are deemed to have obtained Basic Education in Nigeria.

The objectives of the Upper Basic Education, as enshrined in the National Policy on Education (Federal Government of Nigeria, FGN, 2014), include to develop interest in science and technology, acquire basic knowledge and skills in science and technology, apply their scientific and technological knowledge and skills to meet societal needs, take advantage of the numerous career opportunities offered by science and technology and be come prepared for further studies in science and technology.

To achieve the above objectives, several subjects are offered at the Upper Basic Education level. One of such subjects is Basic Science which was in traduced as a core subject to introduce students to the world of science and to prepare them for higher education in science and technology. It is a course of study which is devised and presented in such a way that students gain the concept of the fundamental unity of science, the commonality of approach to problem of scientific nature and help students to gain an understanding of the roles and functions of science in everyday life and the world in which they live.

Agbo (as cited in Sambo, Kukwi, Eggari & Mahmuda, 2014) stated that basic science is the bedrock to advance studies in science, technology and engineering. The implementation of the basic science programme started in July, 2006 with the appropriation of the UBE fund to the Universal Basic Education Commission (UBEC) and subsequent disbursement to States. Its main objectives include to prepare students to acquire adequate laboratory and field skills; inculcation of meaningful and relevant knowledge in basic science; and the ability to apply scientific knowledge to everyday life in matters of personal and community health, and agriculture, reasonable and functional scientific attitudes.

The achievement of the above objectives depends on the three pillars of education, which are the teacher, the student and the subject matter. Teachers are charged with the responsibility of teaching the subject to the students who are expected to pay attention and learn the contents of the subject matter. This is why during and at the end of a school term, the teacher is expected to assess the extent to which learning objectives have been achieved by the students. The aim is to ascertain areas of strength and weaknesses of the students either for certification, promotion, placement or formation. In doing this, the teacher makes use of test.

A test is an assessment tool used to obtain information about individuals in various behavioural traits. Different tests can be used to assess students in Basic Science in the three domains of learning; the cognitive, affective and psychomotor domains of learning. These tests include achievement, aptitude, 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.

Achievement tests which are the focus of this study are of two major forms such as teacher-made-tests and standardized tests. The teacher-made tests are those constructed by the teacher, who in most cases is not skilled in the art of test construction. These tests are, therefore, characterized by items that are not valid and reliable. These items are often poorly constructed in such a way that they lack both content and construct validity and they are often not reliable (Osadebe & Jessa, 2018).

The standardized achievement tests, on the other hand, are tests which have undergone serious psychometric evaluation and assessment. These tests are usually constructed by experts based on the principles of test construction. The tests are often used by examination bodies to conduct tests with which the summative assessment of the students is carried out. Standardized achievement tests follow rigorous item analysis and assessment, before being deployed for testing.

The Basic Science Test (BST) used by the Delta State Basic Education Certificate Examination (BECE) is an example of a standardized achievement test. It is used by the examination body to assess the extent to which the students have achieved the objectives of the Basic Science Curriculum. It is usually taken at the end of Upper Basic Education for certification and promotion purpose. Those who pass the test, alongside other subjects, are promoted to the senior secondary school stage of education.

As a result of the importance of the Basic Science Test to scientific and technological development, there is a need to ascertain its psychometric properties in terms of validity and reliability. For test items to achieve its aim, the test items must meet up with the theoretical scale for item selection using the item parameters.

Item parameters are statistical indicators that define the quality of an item in the instrument employed (Orheruata, 2015). These statistical indicators are item difficulty, discrimination and guessing parameters. Item difficulty parameter (b) refers to the examinee's ability level at which approximately half of the examinees are likely to answer a particular item correctly. Item discrimination parameter (a) describes the strength of an item discriminating between examinee with trait level (θ) below and above the threshold (b) while guessing parameter (c) is the probability of getting the item correct by guessing alone (Embretson & Reise, 2010).A well-developed test needed to have itsitem parameters in conformity with the theoretical scale for item selectionusing test theories.

There are two test theories used in determining test item parameters.They are the Classical Test Theory (CTT) and the Item Response Theory. Classical Test Theory (CTT) is abody of related psychometric theory that predicts outcomes of psychologicaltesting such as difficulty of items or the ability of candidates. The theoryactually deals with the effect of both unsystematic and systematic influence onthe observed test scores. Classical Test Theory is based on the decompositionof observed scores into true and error scores. The observedscore changes as the amount of random (unsystematic) error changes.

Theproblem with Classical Test Theory estimators has to do with circulardependency, which refers to a situation where (a) The person statistic (i.e., observed score) is (item) sample dependent, and (b) the item statistics (i.e., item difficulty and item discrimination) are (examinee) sample dependent. It is also on record that the Classical Test Theory estimators ofitem difficulty and item discrimination indexes are not generalized acrosspopulations. Classical Test Theory is limited in the comparison of performanceof different examinees. The examinees must be given either the same orparallel items. Embretson and Reise (2010)

reported thatClassical Test Theory provides no bases for determining how an examineemight perform when confronted with a test item and that CTT assumes that themeasurement error is the same for all examinees. These limitations have led toa new measurement theory which is Item Response Theory (IRT).

Item response theory is one of the statistical frameworks that generatea mathematical function to describe the relationship between studentperformance in a test and ability or trait level. Its procedure improves psychometric methodology and assessment instruments. It provides meaningful information about examinee when its methodology is focused onthe relationship between each individual item and the underlying (latent) traitor ability assessed by the instrument. A group of items responded to by a groupof examinees are used to estimate the item parameters in order to discover anitem's measurement qualities. A test developed using IRT provide informationabout an item at its difficulty level, discrimination level, and guessing level forefficient procedure for estimating item parameters.

One of the problems observed in the Basic Science Test is that it is developed with the use of Classical Test Theory (CTT). The Ministry of Basic and Secondary Education Examinations and Standards (as cited in Ogugo & Lotobi, 2019) stated that the Basic Education Certificate Examination body use the classical test theory to estimate the item parameter of their tests. This is despite the shortcomings of classical test theory, as stated above. There is therefore the need to validate the test using Item Response Theory. The reason for the application of item response theory is because, the International Test Commission recommended that IRT be used for the proper description and evaluation of existing and widely used tests (Muñiz, 2011). In view of this background, the aim of this study is to apply the Item Response Theory of measure to validate the Basic Science Test of the Delta State Basic Education Certificate Examination.

Statement of the Problem

As laudable as the objectives of basic science appear, one is not sure whether these objectives have been achieved or achievable as documented evidences have shown that the performance of students in science-related subjects is on the decline. Over the years, the Basic Education Certificate Examination Chief examiner's report in Delta State has been showing a downward trend in the performance of students in Basic Science. In order to overcome this issue of poor performances, the government has been organizing workshop and seminars to improve on teachers' professional skills most especially in the area of teaching methodology but the problem remain the same.

The persistent issue of poor performances has shown that the teaching methodology could not be the sole factor responsible for students' poor performances. Also, it was observed that the examination body has been using the Classical Test Theory (CTT) in the development of their tests, including the Basic Science Test. With the numerous shortcomings of the CTT, such as circular dependency, CTT estimators of item difficulty and item discrimination indexes not being generalized across populations, limitation in comparison of performance of different examinees and providing no basis for determining how an examinee might perform when confronted with a test item. These limitations have led to a new measurement theory which is Item Response Theory (IRT).

The researcher believes there is a need to validate the test using a more robust and contemporary theory of measurement. Hopefully, this will help finetune the test to function more in testing the performance of the testees. In view of this, the problem of this study,

therefore, is to what extent will the Basic Science Test be considered valid when its parameters are estimated with the use of Item Response Theory?

Research Questions

The following research questions guided the study:

1. What is the b-parameter index for each item of the 2021 Basic Science Test?
2. What is the a-parameter index for each item of the 2021 Basic Science Test?
3. What is the c-parameter index for each item of the 2021 Basic Science Test?
4. What is the concurrent validity of the 2021 Basic Science Test with other science-related subject such as Physical and Health Education Test?

Purpose of the Study

The purpose of this study is to apply item response theory of measurement to validate the Basic Science Test of the Delta State Basic Education Certificate Examination. The study basically:

1. examined the b-parameter index for each item of the 2021 Basic Science Test;
2. found out the a-parameter index for each item of the 2021 Basic Science Test;
3. determined the c-parameter index for each item of the 2021 Basic Science Test; and
4. showed the concurrent validity of the 2021 Basic Science Test with other science-related subject such as Physical and Health Education Test.

Scope and Delimitation of the Study

The focus of this study was on the application of item response theory of measurement to the validation of the Basic Science Test of the Delta State Basic Education Certificate Examination. The test was validated in terms of its b-parameter index, a-parameter index, c-parameter index and concurrent validity with Physical and Health Education Test. The study was delimited to public upper basic schools in Delta State.

RESEARCH METHOD AND PROCEDURE

The study adopted the multiple triangulation research design. According to Kpolovie (2016), this design allows for multi-method approach in studying psychometric properties of an instrument. Since the objective of the study was to validate an existing instrument, this design became appropriate.

Population of the Study

The population of the study comprised all public Upper basic school students in Delta State, Nigeria. From the information obtained from Ministry of Basic and Secondary Education, Delta State has a total of 475 public secondary schools and 140,959JSS 3 students in the 2021/2022 academic session.

Sample and Sampling Techniques

The sample size of the study comprised 1,000 JSS 3 students in public secondary schools in Delta State. The choice of the sample size is based on the recommendation of Lord (1968), that a minimum of 1,000 examinees are required to estimate an *a*-parameter with high accuracy, while 100 examinees for b-parameter and 200 for c-parameter. The sampling technique that was used for the study is proportionate stratified, simple random and convenience sampling techniques. The reason of this choice of sampling technique is because all the local government areas that were used in the study do not have equal number

of students. Hence, there was the need to represent the population of students in each local government area. In doing this, the researcher estimated the percentage of the sample size in relation to the overall size, which resulted in 0.71%.

Research Instrument

The main instrument that was used for the study is the Basic Science Test (BST) that is under validation. Another instrument that was used is the Physical and Health Education Test (PHET) that was used to establish the concurrent validity of the Basic Science Test. The two tests were obtained from the Exams and Standard in Delta State. The tests contain 60 multiple choice questions having five options; one key and four distracters.

Method of Data Collection

The researcher recruited and trained five research assistants who accompanied him to administer the instrument. Prior to the administration of the test, the researcher visited the principals and took permission from them to conduct the study in their schools.

Method of Data Analysis

Various statistical tools were used to analyse the data. The a, b and c parameters item response theory dichotomous models were used to answer research question 1, 2 and 3 while research question 4 was answered usingPearson's product moment correlation coefficient respectively.

PRESENTATION OF DATA AND DISCUSSION OF FINDINGS

Research Question 1: What is the b-parameter index for each item of the 2021 Basic Science Test?

Table 1: b-Parameter Index for the BST

Item	Value	Standard Error	Z	Remark
BST56	34.435	144.889	.238	Very Difficulty
BST28	3.775	2.708	1.394	Very Difficulty
BST43	3.477	1.847	1.883	Very Difficulty
BST2	3.122	2.095	1.490	Very Difficulty
BST17	2.206	.764	2.889	Very Difficulty
BST45	2.180	.623	3.500	Very Difficulty
BST9	1.967	.421	4.667	Difficult
BST48	1.840	.906	2.032	Difficult
BST59	1.509	.839	1.799	Difficult
BST52	1.438	.859	1.673	Difficult
BST23	1.305			Difficult
BST14	1.269	.257	4.930	Difficult
BST11	1.257	11.899	.106	Difficult
BST58	1.181	.156	7.554	Difficult
BST50	1.178	.796	1.480	Difficult
BST12	1.161	.131	8.862	Difficult
BST55	1.044	.161	6.495	Difficult
BST24	1.029	.117	8.829	Difficult
BST22	.987	.233	4.234	Moderately Difficult
BST1	.974	.236	4.133	Moderately Difficult
BST38	.915	.205	4.470	Moderately Difficult
BST41	.907	.117	7.728	Moderately Difficult

BST25	.895	.136	6.563	Moderately Difficult
BST40	.880	.093	9.449	Moderately Difficult
BST4	.857	.117	7.300	Moderately Difficult
BST49	.799	.142	5.632	Moderately Difficult
BST30	.721	.096	7.547	Moderately Difficult
BST51	.690	2.745	.251	Moderately Difficult
BST39	.659	.112	5.858	Moderately Difficult
BST7	.650	.255	2.545	Moderately Difficult
BST54	.630	.094	6.716	Moderately Difficult
BST6	.618	.171	3.609	Moderately Difficult
BST8	.602	.381	1.581	Moderately Difficult
BST33	.556	.151	3.675	Moderately Difficult
BST16	.471	.146	3.219	Moderately Difficult
BST53	.462	.155	2.987	Moderately Difficult
BST44	.449	.086	5.254	Moderately Difficult
BST31	.317	.099	3.203	Moderately Difficult
BST36	.276	.201	1.373	Moderately Difficult
BST35	.186	.211	.883	Moderately Difficult
BST18	.180	.153	1.179	Moderately Difficult
BST47	.094	.210	.446	Moderately Difficult
BST32	.028	.169	.167	Moderately Difficult
BST19	.008	.109	.078	Moderately Difficult
BST13	-.045	.124	-.361	Easy
BST60	-.136	.374	-.363	Easy
BST34	-.155	.167	-.931	Easy
BST37	-.166	.159	-1.045	Easy
BST10	-.226	.341	-.661	Easy
BST3	-.303	.142	-2.128	Easy
BST57	-.351	.331	-1.059	Easy
BST20	-.517	.162	-3.202	Easy
BST21	-.899	.208	-4.328	Easy
BST15	-.975	.221	-4.418	Easy
BST26	-1.697	.310	-5.474	Easy
BST5	-2.411	.718	-3.357	Very Easy
BST46	-2.577	.485	-5.314	Very Easy
BST27	-4.143	4.449	-.931	Very Easy
BST29	-14.761	28.268	-.522	Very Easy
BST42	-24.382	26.736	-.622	Very Easy

As shown in Table 1, the item difficulty index ranged from -24.382 to 34.435, with a higher index indicating a very difficult item and a lower index indicating a very easy item. From the result, 6 items are very difficult, 12 are difficult, 26 are Moderately Difficult, 11 are easy while 5 items are very easy.

Research Question 2: What is the a-parameter index for each item of the 2021 Basic Science Test?

Table 2: a-Parameter Index for the BST

Item	Value	Standard Error	Z	Remark
BST51	29.435	6889.605	.004	Satisfactory (No Revision Required)
BST11	25.822	2989.887	.009	Satisfactory (No Revision Required)
BST23	20.252	2635.736	.007	Satisfactory (No Revision Required)
BST52	16.308	179.723	.091	Satisfactory (No Revision Required)
BST41	5.087	1.947	2.613	Satisfactory (No Revision Required)
BST44	4.157	1.013	4.102	Satisfactory (No Revision Required)
BST58	4.096	2.779	1.474	Satisfactory (No Revision Required)
BST31	4.078	1.000	4.076	Satisfactory (No Revision Required)
BST12	4.007	1.827	2.193	Satisfactory (No Revision Required)
BST30	3.993	1.171	3.410	Satisfactory (No Revision Required)
BST24	3.760	1.475	2.549	Satisfactory (No Revision Required)
BST40	3.723	.883	4.216	Satisfactory (No Revision Required)
BST53	3.625	1.323	2.739	Satisfactory (No Revision Required)
BST4	3.190	.919	3.471	Satisfactory (No Revision Required)
BST6	3.076	1.331	2.310	Satisfactory (No Revision Required)
BST33	2.935	1.022	2.870	Satisfactory (No Revision Required)
BST54	2.874	.597	4.812	Satisfactory (No Revision Required)
BST39	2.798	.698	4.008	Satisfactory (No Revision Required)
BST25	2.601	.840	3.097	Satisfactory (No Revision Required)
BST37	2.549	.617	4.135	Satisfactory (No Revision Required)
BST49	2.495	.667	3.742	Satisfactory (No Revision Required)
BST16	2.423	.660	3.670	Satisfactory (No Revision Required)
BST47	2.246	.702	3.198	Satisfactory (No Revision Required)
BST14	2.176	.995	2.187	Satisfactory (No Revision Required)
BST9	2.116	.992	2.133	Satisfactory (No Revision Required)
BST55	2.056	.713	2.884	Satisfactory (No Revision Required)
BST19	1.945	.316	6.161	Satisfactory (No Revision Required)
BST7	1.845	.802	2.301	Satisfactory (No Revision Required)
BST22	1.800	.882	2.041	Satisfactory (No Revision Required)
BST60	1.766	.651	2.715	Satisfactory (No Revision Required)
BST38	1.670	.585	2.857	Good (Little or no Revision Required)
BST13	1.585	.266	5.965	Good (Little or no Revision Required)
BST8	1.553	.794	1.956	Good (Little or no Revision Required)
BST1	1.503	.855	1.758	Good (Little or no Revision Required)
BST35	1.444	.370	3.900	Good (Little or no Revision Required)
BST45	1.430	.778	1.837	Good (Little or no Revision Required)
BST3	1.409	.246	5.739	Good (Little or no Revision Required)
BST20	1.286	.231	5.569	Moderate (Little or no Revision Required)
BST10	1.206	.314	3.841	Moderate (Little or no Revision Required)
BST26	1.185	.267	4.431	Moderate (Little or no Revision Required)
BST18	1.146	.218	5.264	Moderate (Little or no Revision Required)
BST21	1.126	.221	5.103	Moderate (Little or no Revision Required)
BST15	1.095	.219	4.999	Moderate (Little or no Revision Required)
BST34	1.055	.206	5.115	Moderate (Little or no Revision Required)
BST32	1.004	.201	4.990	Moderate (Little or no Revision Required)
BST36	.830	.192	4.324	Moderate (Little or no Revision Required)
BST50	.805	.845	.952	Moderate (Little or no Revision Required)
BST57	.717	.186	3.854	Moderate (Little or no Revision Required)

BST17	.574	.206	2.784	Marginal (Revision Required)
BST48	.534	.434	1.232	Marginal (Revision Required)
BST43	.360	.193	1.860	Marginal (Revision Required)
BST59	.339	.164	2.067	Poor Item
BST2	.259	.169	1.527	Poor Item
BST28	.242	.171	1.413	Poor Item
BST56	.046	.173	-.625	Poor Item
BST29	-.098	.190	-.516	Poor Item
BST42	-.522	.182	-.525	Poor Item
BST27	-.940	1.447	-.650	Poor Item
BST5	-1.401	1.090	-1.285	Poor Item
BST46	-5.701	10.597	-.538	Poor Item

As shown in Table 3, the item discrimination index ranged from -5.701 to 29.435, with a higher index indicating a satisfactory item and a lower index indicating a poor item. From the result, 30 items are satisfactory, 7 are good, 11 items are moderate, 3 items are marginal while 9 items are Poor.

Research Question 3: What is the c-parameter index for each item of the 2021 Basic Science Test?

Table 3: c-Parameter Index for the BST

Item	**Value**	**Standard Error**	**Z**	**Remark**
BST8	.451	.104	4.330	Guessable
BST27	.446	.062	7.248	Guessable
BST11	.418	.040	10.369	Guessable
BST6	.413	.064	6.399	Guessable
BST58	.409	.052	7.903	Guessable
BST14	.388	.059	6.590	Guessable
BST41	.340	.048	7.014	Guessable
BST52	.283	.037	7.726	Guessable
BST5	.282	.061	4.648	Guessable
BST12	.257	.047	5.439	Guessable
BST22	.257	.082	3.123	Guessable
BST7	.248	.097	2.556	Not Guessable
BST53	.246	.067	3.679	Not Guessable
BST33	.245	.065	3.789	Not Guessable
BST30	.217	.045	4.790	Not Guessable
BST49	.217	.051	4.243	Not Guessable
BST4	.188	.046	4.051	Not Guessable
BST24	.183	.045	4.027	Not Guessable
BST60	.182	.159	1.143	Not Guessable
BST25	.167	.050	3.347	Not Guessable
BST38	.158	.069	2.287	Not Guessable
BST16	.148	.061	2.431	Not Guessable
BST51	.148	.035	4.208	Not Guessable
BST23	.146	.029	5.049	Not Guessable
BST55	.126	.051	2.485	Not Guessable
BST39	.120	.044	2.722	Not Guessable
BST40	.116	.034	3.376	Not Guessable

BST31	.099	.045	2.214	Not Guessable
BST47	.087	.098	.887	Not Guessable
BST42	.083	.020	4.049	Not Guessable
BST45	.081	.039	2.075	Not Guessable
BST9	.067	.027	2.532	Not Guessable
BST10	.067	.131	.511	Not Guessable
BST1	.057	.117	.491	Not Guessable
BST48	.054	.230	.236	Not Guessable
BST35	.049	.080	.614	Not Guessable
BST44	.048	.032	1.518	Not Guessable
BST54	.045	.028	1.604	Not Guessable
BST50	.044	.337	.131	Not Guessable
BST56	.036	.362	.152	Not Guessable
BST37	.034	.070	.486	Not Guessable
BST46	.014	.010	1.404	Not Guessable
BST29	.002	.101	.024	Not Guessable
BST57	.001	.073	.019	Not Guessable
BST2	.000	.017	.013	Not Guessable
BST3	.000	.000	.003	Not Guessable
BST13	.000	.000	.003	Not Guessable
BST15	.000	.004	.008	Not Guessable
BST17	.000	.003	.003	Not Guessable
BST18	.000	.000	.002	Not Guessable
BST19	.000	.001	.003	Not Guessable
BST20	.000	.001	.004	Not Guessable
BST21	.000	.002	.007	Not Guessable
BST26	.000	.009	.008	Not Guessable
BST28	.000	.006	.007	Not Guessable
BST32	.000	.000	.004	Not Guessable
BST34	.000	.002	.007	Not Guessable
BST36	.000	.003	.004	Not Guessable
BST43	.000	.002	.003	Not Guessable
BST59	.000	.008	.009	Not Guessable

As shown in Table 3, the item guessing index ranged from .000 to .451, with a higher index indicating a guessable item and a lower index indicating not guessable item. The suggested range should be 0.00 and 0.25. this is because a low examinee should have 1/4 = 0.25 chance of guessing the right answer on the BST, which includes 4 possibilities. The examinees will be guessing amongst the remaining three possibilities once the correct key has been extracted because c = 0.25 for this 4-alternative item. 11 items from the result had a guessing index above the recommended 0.25, indicating that they are able to be guessed. The guessing index for the remaining items (49 of the 60) ranged from 0.000 to 0.248, indicating that they are not guessable.

Research Question 4: What is the concurrent validity of the 2021 Basic Science Test with other science-related subject such as Physical and Health Education Test?

Table 4: Concurrent validity of the BST

Correlations			
		PHE	BST
PHET	Person Correlation	1	.687**
	Sig. (2-tailed)		.000
	N	1000	1000
BST	Person Correlation	.687**	1
	Sig. (2-tailed)	.000	
	N	1000	1000

**. Correlation is significant at the 0.01 level (2-tailed).

The results have shown that the correlation coefficient between the two tests (PHE and BST) is 0.687, and therefore the Basic Science Test has a concurrent validity of 0.687. A correlation coefficient of 0.687 is a very strong correlation and a stunning indication of very high concurrent validity of the BST.

Discussion of Findings

b-Parameter Index for Each Item of the 2021 Basic Science Test

The first finding revealed that majority of the items in the test are moderately difficulty and partially fall in line with the ability level of the students for whom the test was made. However, the results showed that some of the questions had an index that was greater than the benchmark of -3.0 to 2.00, suggesting that the items are tough and very difficult, and as a result, they need to be modified before they can be used to evaluate achievement in Basic Science. The chance of a successful response equals c/2 +.50 at a certain point on the theta (θ) continuum, which is indicated by the b-parameter, which is also known as the item difficulty parameter. As a result, the b-parameter is in the center of the IRF and is where the slope is steepest, demonstrating the item's greatest ability to discriminate. The b parameter displays the examinee's value for which the item is appropriate because the Basic Science Exam is focused on the examinees selected from Delta State's public secondary school students. A score below -1.0 denotes an extremely easy object, while a value higher than 1.0 denotes a more challenging thing. According to Kpolovie and Emekene (2016), the difficulty index goes from -3.0 (extremely easy) to +3.0 in actuality, while in principle it ranges from negative to positive infinity (very difficult). The values of the b-parameter column for the Basic Science Test items ranged from -3.05 to 22.22, as can be seen by carefully examining the column. The last item, which has a b value of 22.22, is where the b-parameter reached its maximum degree of difficulty. The b-parameter is related to the traditional P statistic, according to Kpolovie and Emekene (2016), who observed a similar discovery. Items with low P values will tend to have higher (more positive) b-parameters, whereas items with high P values would tend to have lower (more negative) b-parameters.

a-Parameter Index for Each Item of the 2021 Basic Science Test

The second finding revealed that majority of the items in the test discriminated well between higher achievers and low achievers. However, the results showed that some items had indices below the standard of 0.34, indicating that they had poor, marginal, or moderate discriminatory indices and hence needed further work before they could be used. The item discrimination (slope) parameter, also known as the a-parameter, measures how well a response can distinguish between respondents with skills to the left of the item position and those with abilities to the right of the item location (Thorpe & Favia, 2012). When higher ability students have a larger probability of answering a test question correctly than lower ability students, the test item shows positive discrimination.When candidates with high

ability have a low likelihood of answering an item correctly and candidates with low ability have a higher probability of responding an item correctly, a test item has negative discrimination (a-values). A high level of discrimination means that the item makes a good distinction between people with low and high skill levels. The steepness of the item characteristics curve can be used to graphically represent a measure as a discrimination parameter (ICC). A good test item should typically have an item discrimination value (a-values) over 1; however, a-values above 0.75 are also sometimes appropriate.The aforementioned result is consistent with the research by Kpolovie and Emekene (2016), who employed item response theory to validate the Raven's Advanced Progressive Matrices (APM). The authors discovered that, according to the 3-Parameter Logistic IRT Model, all test items produce favorable results in terms of discrimination, difficulty, and guessing.

c-Parameter Index for Each Item of the 2021 Basic Science Test

The third finding revealed that majority of the items are not guessable. Meaning that the instrument is not subject to guessing tendency, as it meets the assumption of item response theory in terms of the c-parameter. The c-parameter is equal to the probability that an examinee will guess correctly and receive a correct response, which is endlessly low. As a result, c is also the IRF's lower asymptote. The position of a and b on the scale is influenced by the presence of a non-zero c parameter. For multiple-choice assessments, the c parameter should roughly equal 1 divided by the number of options.Hence, a low examinee should have 1/4 = 0.25 chance of finding the right answer for the BST that has four possibilities. The examinees will be guessing amongst the remaining three possibilities once the correct key has been extracted because c = 0.25 for this 4-alternative item.Consequently, the value of c will be significantly lower than 0.25 in areas where guessing is prevalent. A higher number indicates that there is significant evidence for guessing.The results are consistent with those of Ani (2014), who used item response theory to develop and validate a multiple-choice test in economics and discovered that, using a three-parameter model (3pl) model, 49 out of 50 items were reliable. The results were in line with those of Kpolovie and Emekene (2016), who used item response theory to validate advanced progressive matrices in Nigeria and discovered that all test items produced favorable statistics under the 3-Parameter Logistic IRT Model in terms of discrimination, difficulty, and guessing.

Concurrent Validity of the 2021 Basic Science Test with other Science-Related Subject such as Physical and Health Education Test

The fourth finding revealed that the BST has a very high concurrent validity. Concurrent validity of a test under validation or under development is simply established bycorrelating it with another test that validly and reliably measures the same trait or domain in the population that the test is being validated or developed for. It was for this reason that the PHET that validly and reliably measures students' science skills was simultaneously administered with the BST to the subjects of thecurrent study. The correlation coefficient between the PHET and BST is the concurrent validityof the BST. The correlation coefficient obtained between the Basic Science Test and Physical and Health Education test is 0.687, which indicate a very high concurrent validity. The above finding supports the work of Kpolovie and Emekene (2016), who used a similar approach to examine the concurrent validity of Advanced ProgressiveMatrices (APM). The author used Pearson product moment correlation coefficient to correlate the scores of APM and Culture Fair Intelligence Test (CFIT) and found a coefficient of 0.701, which indicate a high concurrent validity of the APM.

Conclusion

Based on the findings, it can be concluded that IRT can play a complementary role in tests development. The findings about the suitability of the 3-Parameter Logistics Model is an eye opener to the fact that a perfect fitstatistic can be achieved for the enhancement of the credibility of Basic Science Test, which was developed using CTT.

Recommendations

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

1. The validated Basic Science Test should be added to the already existing item pool;
2. All items having bad difficulty and discrimination should be modified so that they will be useful in the assessment of Upper Basic school students in Basic Science;

REFERENCES

Ani, E. N. (2014). *Application of item response theory in the development and validation of multiple choice test in economics.* (Master's thesis). University of Nigeria, Nsukka.

Embretson, S. E., & Reise, S. P. (2010). *Item Response Theory for Psychologists Multivariate Applications.* Book Series, Lawrence Erlbaum Associates Inc, Mahwah New Jersey.

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

Kpolovie, P. J.,& Emekene, C.O. (2016) Psychometric advent of advanced progressive matrices- smart version (APM-SV) for use in Nigeria. *European Journal of Statistics and Probability,* 4(3), 20-30.

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

Oguguo, B. C. E., & Lotobi, R. A. (2019). Parameters of Basic Science Test Item's of 2011 Basic Education Certificate Examination Using Item Response Theory (IRT) Approach in Delta State, Nigeria. *European Journal of Educational Sciences,* 6(1), 22-36.

Orheruata, M. U. (2015). Item Parameter Drift in Certificate Examinations and it's implication on Decision Making. *African Journal of Theory and Practice of Educational Assessment.* Educational Assessment 2, 98-105.

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

Sambo, M., Kukwi, I., Eggari, S., & Mahmuda, A. (2014). Assessment of the Implementation of Basic Science Programme in Upper basic school in Nasarawa. *Developing Country Studies,* 4(20).

Thorpe, G.L., & Favia, A. (2012). Data analysis using item response theory methodology: An introduction to selected programs and applications. *Psychology Faculty Scholarship*, 1-34.