



Application of Classical Test Theory in the Development, Validation, and Standardization of Multiple-Choice Basic Science Test Items for Secondary School Teachers' Use

Ajayi, O. A. and Salam, R. I.

Department of Counselling Psychology & Educational Foundations, Tai Solarin Federal University of Education, Ijagun, Nigeria

Corresponding email: ajayioa@tasued.edu.ng

Abstract

Development of standardized achievement tests in Basic Science requires a rigorous process of item analysis to ensure validity, reliability, and fairness. This is in addition to poor teachers' expertise in test development. This study employed classical test theory as its measurement framework with instrumentation research design of survey type. Five (5) research questions were raised while a pool of 120 multiple-choice tests in JSS Basic Science formed the research instrument. Multistage sampling technique was adopted in this study. The population of this study was made up of all the junior secondary school students in Lagos Central Senatorial District of Lagos State. Multistage sampling techniques was adopted in this study and (3) three local government were selected out of the five-local government in Lagos central senatorial district. A total number of three local governments, three public junior secondary school and 510 junior secondary school students constitute the sample for this study. The data obtained were analysed using item analysis, Kuder Richardson (20) and test norms. Results showed that out of the 120 pool of test items, only 63 were good items based on based on the criteria of difficulty index ranging from 40 to 60% and positive discrimination power. Part of the findings showed that the good items yielded a high reliability coefficient of 0.73 after dropping 23 items during the reliability test. The selected 40 items were administered to 120 JSS1 to 3 students for the establishment of test norms such as standard score, age and grade equivalent norms. The researchers' concluded that the selected standardized test items at the end of the study not only possess good psychometric properties but also has good test quality in terms of test usability and ability to reduce guessing. It is recommended that the standardized test items should be looked upon as a model achievement test in Basic Science for junior secondary schools and regular sensitization workshop, seminars and conferences should be organized for teachers in order for them to be acquainted with requisite techniques needed for construction of valid assessment instruments.

Keywords: Classical Test Theory, Item analysis, Validation, Multiple-choice test, Basic Science

1. Introduction

It is a common knowledge that teachers are the implementers of the curriculum. It is through teaching of various subjects by the teachers that the task of curriculum implementation is done. This actually occurs during teaching - learning process. The teacher actually has a set of behavioural objectives which the students have to attain by the end of the lesson. By the end of the lesson, the teacher evaluates the students through some questions or tests to know the extent of attainment of the behavioural objectives. One of the tests use by the teachers to evaluate the level of attainment of the behavioural objectives by the students is the Objective Questions Test. According to Egbule (2012), objectives test is a well-structured test item in which the testee or student is required to identify or select the correct option from a given set of alternatives. He added that test is a question with three

or more alternatives out of which one is correct, the more the alternatives the more reliable to a point. He emphasized that the choice of the number of alternatives depends on the age and ability level of the examinees. They could be designed to assess objectives at any level of the domain.

The quality of a test given by a teacher is closely linked with its ability to provide the kind of information needed regarding students' performances. A well written test allows the teacher to accurately and consistently measure students' mastery of specific contents taught in class. Results of such tests allow teachers to measure to some degree, how effective their instruction has been. Conversely, poorly designed test items can lead to inaccurate measurements of learning and provide false information regarding students' performance as well as instructional effectiveness (Nworgu, 2012). Any item answered correctly or incorrectly because of extraneous factors in the item results in misleading feedback to both examinee and examiner (Frey, 2007). The aftermath effect of such could lead to poor students' performance in science, and thus could hinder learning of science in students. Since the field of science is sacrosanct to

Ajayi, O. A. and Salam, R. I., (2025). Application of Classical Test Theory in the Development, Validation, and Standardization of Multiple-Choice Basic Science Test Items for Secondary School Teachers' Use. *Journal of Specialised and Professional Education (JOSPED)*, SP: Nov, 2025, pp. 30-41.

©COSPED, Special Edition, November, 2025

nations development, there is therefore need for quality student's assessment in science subjects.

Many nations develop because of the dominant role played by science. Science has been instrumental to the solution of most socio-economic problems such as hunger, unemployment and population explosion among others that are facing nations like Nigeria (Afolabi, 2009). However, Nigeria has been making frantic efforts to create scientific and technological awareness in her citizenry (Adeyemi, 2007). All over the world, there has been an increased search for means of improving and promoting teaching and learning, especially, the teaching and learning of Basic Science and other Sciences. It is believed that the search for improving and promoting teaching and learning of Basic Science and other science subjects will promote better understanding of science concepts and encourage greater scientific and technological advancements (Adams, 2008). This search has led to lots of research done by Science Educators towards making Science Education better, especially in the area of teaching students with the notion of discovering, engaging in critical thinking, questioning and developing problem solving skills. Therefore, the curriculum for Science and Technology should be developed to educate science – literate students to enable them inquire and solve problems facing them. Many developed and developing countries of the world have introduced many educational reforms especially in the field of Science and Technology. Some countries like Japan, America and Britain have achieved good results and have become self-reliant. Unfortunately, in Nigeria, despite all the reforms and innovations, such as the 6-3-3-4 system of education, the Universal Basic Education (UBE), the introduction of guidance and counselling in schools, expanded curriculum, and adoption of learning intervention programmes in schools, yet, not much success have been recorded particularly in Science Education, and has a result, it has hindered the goal of producing a self-reliant nation (Jegade, 2014).

The students' poor performance in Basic Science has drawn attention of researchers and curriculum planners towards Basic Science as a subject in the school curriculum (Kareem, 2013). The desire to know the causes of the poor performance in Basic Science has been the focus of researchers for some time now. Researchers such as (Ahmed, 2008) had shifted the blame of students' poor performance in Basic Science on teacher laxity, students poor study habits, parents' poor attitude to their children education and so on without thinking on the properties of tests such as difficulty indices and discrimination power of test items. It is in view of the forgoing that this study therefore becomes imperative. This study therefore seeks to construct, administer, validate and standardized pool of multiple-choice tests items in Basic Science with a view to select good items that can be useful for both internal and external assessments in Basic Science.

1.1 Statement of the problem

Statistical report indicates a steady decline in the achievement of junior senior secondary students in Basic Science. This poor students' achievement in Basic Science has been partly traced to badly constructed test items. As a result of these, student achievement in Basic Science at external examination is very poor. It is in view of the forgoing that this study therefore becomes imperative. This study therefore seeks to construct, validate and standardise a pool of multiple-choice tests items in Basic Science with a view to select standardized items that can be suitable for both internal and external assessments in junior secondary school Basic Science.

1.2 Purpose of the study

The main purpose of this study is to develop, administer and standardized pool of multiple-choice test items in junior secondary school Basic Science, while the specific purposes are to:

1. Determine the proportion of good items suitable for selection of the achievement test in Basic Science?
2. Identify the proportion of the items with positive discriminating power?
3. Determine the proportion of the items with difficulty index range of 40 – 60%
4. How reliable is the selected achievement test?
5. To what extent does students' performance in the standardized Basic Science test vary on the bases of test norms?

1.3 Research questions

The following research questions are germane for this study:

1. What is the proportion of good items selected for the achievement test?
2. What is the proportion of the items with positive discriminating power?
3. What is the proportion of the items with difficulty index range of 40 – 60%?
4. How reliable is the selected achievement test?
5. To what extent does students' performance in the standardized Basic Science test vary on the bases of test norms?

2. Methodology

This study adopted instrumentation research design of survey type. The population of this study is made up of all the junior secondary school students in Lagos Central Senatorial District of Lagos State. Multistage sampling technique was adopted in this study. At first, three local government were selected out of the five-

local government in Lagos central senatorial district. Secondly, each of the local government selected was used for each stage of the study. In the first local government selected (Mushin local government), one public junior secondary school and 360 junior secondary school students were selected across J.S. 1 to 3 for the study. Thirdly, in the second local government (Surulere), only one school and 30 students were selected and were used to determine the reliability of the instruments. Lastly, in the third local government selected (Isolo), one junior secondary school and 120 randomly selected junior secondary school students were selected and was used for the last stage of the study (standardization and establishment of test norms). In all, a total number of three local governments, three public junior secondary school and 510 junior secondary school students constitute the sample for this study. Ballot method without replacement was the sampling techniques adopted in selecting students across each school and classes.

2.1 Instrument for Data Collection

The instrument used for the data collection is the pool of 120 multiple choice test items in Junior Secondary school Basic Science. The test items were constructed by the researcher and were based on 1st and 2nd term J.S. 1 to J.S.3 Basic Science scheme of work. The researcher strictly follows guidelines for constructing multiple choice test items while constructing the pool of test items in Basic Science. Also, the distribution of the test items according to the content and the levels of cognitive domain are shown in the Table 1.

2.2 Validation of the Instrument

Content validity was adopted in validating the research instruments. Similarly, to ensure reliability of the achievement test, the selected good items after the item analysis stage were trial tested on 30 students that are parallel to the sample population and a reliability coefficient of 0.730 was obtained using Kuder Richardson (20) technique.

2.3 Data Analysis Technique

In order to determine the difficulty level, discriminating power, selection of standardized tests items from the pool of 120 items and establishment of test norms, the administered test items were however subjected to item analysis, reliability analysis and test norms analysis. The item analysis was performed electronically using Microsoft excel statistics, while the reliability estimates was obtained using Kuder Richardson (20) techniques. Furthermore, establishments of test norms were performed electronically using SPSS.

3. Results

Research Question One: What is the proportion of good items selected for the achievement test? This research question was answered using item analysis and the results is presented as shown in Table 2.

Results of item analysis as shown in Table 1 and in

line with psychometric experts suggested that items with difficulty index ranging from 40 – 60% are good items, out of Pool of 120 test items analyzed, only 63 are good items to be selected, while the remaining (57) items were either too difficult or too easy, hence they were discarded.

Research Question Two: What is the proportion of the items with positive discriminating power?

This research question was answered using item analysis and the result is found in Table 1. From table 1, it could be observed that out of the pool of 120 multiple choice test in Basic Science 112 items has positive discriminating power while the remaining 8 items have negative discriminating power.

Research Question Three: What is the proportion of the items with difficulty index range of 40 – 60%?

This research question was also answered using item analysis and the result is found in Table 1. From Table 1 above, it could be observed that out of the pool of 120 multiple choice test in Basic Science only 63 items have difficulty index range of 40 – 60%, hence were regarded as good items, while the remaining 57 items have difficulty index below or above 40 – 60%.

Research Question Four: How reliable is the selected good achievement test? This research question was also answered using Kuder – Richardson (20) techniques and the results in presented in Table 2.

At the first reliability test carried out on the 63 tests items, reliability estimate of 0.003 was obtained which is extremely low reliability index. Upon deletion of 23 items, a higher reliability estimates of 0.730. Conclusively, 40 items out of the 63 good items are highly reliable and poses good internal consistency as shown in the table above. Inferentially 40 good items from the pool of 120 test items poses good psychometric properties of test in terms of item analysis, validity, reliability and useability. However, the forty (40) valid and reliable items were therefore regarded as standardized test items and were further used for establishment of test norms.

Research Question Five: To what extent does students' performance in the standardized Basic Science test vary on the bases of test norms?

This research question was also answered electronically using rank, percentile rank, standard scores, age and grade equivalent norms with SPSS as presented in Table 3.

The results of table 5 shows variation in individual and group performance on the basis of standard scores, age and grade equivalent norms. The standard scores such as the Z-score, T-score and percentile rank shows the performance of each student and position among their peers. For example, the percentile rank shows the

Table 1: Test Blue Print on 120 Pool of Test Items in J.S.S. Basic Science

S/N.	Content	Knowledge (37.5%)	Comprehension (33.3%)	Application (29.2%)	Analysis (0%)	Synthesis (0%)	Evaluation (0%)	Total (100%)
1	Living and non- living things 22.5%	9	10	8				27
2	Environmental pollution 5%	2	2	2				6
3	Reproduction 11.7%	6	4	4				14
4	Sexually transmitted infection 7.5%	3	3	3				9
5	Energy 12.5%	6	5	4				15
6	Habitat 5.8%	3	2	2				7
7	family life education 5.8%	3	2	2				7
8	kinetic theory of matter 11.7%	5	5	4				14
9	Genetics 6.7%	3	3	2				8
10	Erosion and flooding 10.8%	5	4	4				13
TOTAL		45	40	35	-	-	-	120

Table 2: Item analysis table for calculating discrimination and difficulty level

ITEM ANALYSIS TABLE FOR CALCULATING DISCRIMINATION AND DIFFICULTY LEVEL							
FOR 120 MULTIPLE CHOICE TEST ITEMS IN JSSCE BASIC SCIENCE							
ITEM NO	RANK	RH=97	RL=97	D=RH – RL	RH+RL	P=RH+RL/n*100	COMMENTS
1	28	85	37	48	122	62.9	Not Good Item
2	20	92	40	52	132	68.0	Difficult item
3	48	94	57	37	151	77.8	Difficult item
4	90	71	53	18	124	63.9	Not Good Item
5	71	74	46	28	120	61.9	Not Good Item
6	30	77	31	46	108	55.7	Good Item
7	120	15	59	-44	74	38.1	Poor Item
8	75	59	34	25	93	47.9	Good Item
9	10	95	39	56	134	69.1	Difficult item
10	63	53	22	31	75	38.7	Poor Item
11	90	45	27	18	72	37.1	Poor Item
12	77	62	38	24	100	51.5	Good Item
13	97	36	20	16	56	28.9	Poor Item
14	59	55	23	32	78	40.2	Good Item
15	48	60	23	37	83	42.8	Good Item
16	41	64	25	39	89	45.9	Good Item
17	23	70	21	49	91	46.9	Good Item
18	80	89	66	23	155	79.9	Difficult item
19	45	75	37	38	112	57.7	Good Item

20	94	22	5	17	27	13.9	Poor Item
21	41	67	28	39	95	49.0	Good Item
22	4	87	27	60	114	58.8	Good Item
23	48	70	33	37	103	53.1	Good Item
24	74	49	23	26	72	37.1	Poor Item
25	113	70	72	-2	142	73.2	Difficult item
26	33	72	28	44	100	51.5	Good Item
27	68	57	27	30	84	43.3	Good Item
28	119	18	27	-9	45	23.2	Poor Item
29	80	47	24	23	71	36.6	Poor Item
30	63	60	29	31	89	45.9	Good Item
31	83	45	23	22	68	35.1	Poor Item
32	54	63	29	34	92	47.4	Good Item
33	69	97	68	29	165	85.1	Difficult item
34	69	63	34	29	97	50.0	Good Item
35	38	68	26	42	94	48.5	Good Item
36	41	84	45	39	129	66.5	Difficult item
37	115	37	41	-4	78	40.2	Good Item
38	116	22	27	-5	49	25.3	Poor Item
39	63	53	22	31	75	38.7	Poor Item
40	3	86	25	61	111	57.2	Good Item
41	5	97	38	59	135	69.6	Difficult item
42	15	81	28	53	109	56.2	Good Item
43	12	86	31	55	117	60.3	Not Good Item
44	59	47	15	32	62	32.0	Poor Item
5	83	48	26	22	74	38.1	Poor Item
46	35	96	53	43	149	76.8	Difficult item
47	38	80	38	42	118	60.8	Not Good Item
48	12	87	32	55	119	61.3	Not Good Item
49	41	87	48	39	135	69.6	Difficult item
50	83	55	33	22	88	45.4	Good Item
51	54	82	48	34	130	67.0	Difficult item
52	100	49	37	12	86	44.3	Good Item
53	80	52	29	23	81	41.8	Good Item
54	48	61	24	37	85	43.8	Good Item
55	22	82	32	50	114	58.8	Good Item

56	86	44	23	21	67	34.5	Poor Item
57	48	68	31	37	99	51.0	Good Item
58	59	60	28	32	88	45.4	Good Item
59	106	30	23	7	53	27.3	Poor Item
60	45	62	24	38	86	44.3	Good Item
61	23	74	25	49	99	51.0	Good Item
62	87	55	35	20	90	46.4	Good Item
63	113	53	55	-2	108	55.7	Good Item
64	100	38	26	12	64	33.0	Poor Item
65	28	87	39	48	126	64.9	Not Good Item
66	111	37	37	0	74	38.1	Poor Item
67	118	16	23	-7	39	20.1	Poor Item
68	116	36	41	-5	77	39.7	Good Item
69	54	68	34	34	102	52.6	Good Item
70	98	54	39	15	93	47.9	Good Item
71	72	69	42	27	111	57.2	Good Item
72	20	79	27	52	106	54.6	Good Item
73	1	72	10	62	82	42.3	Good Item
74	53	64	29	35	93	47.9	Good Item
75	63	56	25	31	81	41.8	Good Item
76	5	87	28	59	115	59.3	Good Item
77	23	79	30	49	109	56.2	Good Item
78	105	42	34	8	76	39.2	Good Item
79	58	66	33	33	99	51.0	Good Item
80	38	74	32	42	106	54.6	Good Item
81	32	88	43	45	131	67.5	Difficult item
82	14	80	26	54	106	54.6	Good Item
83	15	88	35	53	123	63.4	Not Good Item
84	33	81	37	44	118	60.8	Not Good Item
85	63	55	24	31	79	40.7	Good Item
86	54	54	20	34	74	38.1	Poor Item
87	110	24	23	1	47	24.2	Poor Item
88	106	31	24	7	55	28.4	Poor Item
89	100	46	34	12	80	41.2	Good Item
90	94	50	33	17	83	42.8	Good Item
91	103	54	43	11	97	50.0	Good Item

92	15	93	40	53	133	68.6	Difficult item
93	87	54	34	20	88	45.4	Good Item
94	15	80	27	53	107	55.2	Good Item
95	5	93	34	59	127	65.5	Difficult item
96	99	53	39	14	92	47.4	Good Item
97	30	89	43	46	132	68.0	Difficult item
98	15	94	41	53	135	69.6	Difficult item
99	75	60	35	25	95	49.0	Good Item
100	1	92	30	62	122	62.9	Not Good Item
101	10	92	36	56	128	66.0	Difficult item
102	45	69	31	38	100	51.5	Good Item
103	77	37	13	24	50	25.8	Poor Item
104	23	80	31	49	111	57.2	Good Item
105	23	83	34	49	117	60.3	Not Good Item
106	77	59	35	24	94	48.5	Good Item
107	8	91	33	58	124	63.9	Not Good Item
108	9	75	18	57	93	47.9	Good Item
109	111	32	32	0	64	33.0	Poor Item
110	59	65	33	32	98	50.5	Good Item
111	104	37	28	9	65	33.5	Poor Item
112	90	49	31	18	80	41.2	Good Item
113	72	82	55	27	137	70.6	Difficult item
114	35	75	32	43	107	55.2	Good Item
115	109	31	29	2	60	30.9	Poor Item
116	35	70	27	43	97	50.0	Good Item
117	90	45	27	18	72	37.1	Poor Item
118	89	49	30	19	79	40.7	Good Item
119	108	53	49	4	102	52.6	Good Item
120	94	45	28	17	73	37.6	Poor Item

Table 3: Statistical Report Table

Total number of items	120
Number of good items	63
Number of poor/diff. items	57

Table 4: BSAT Reliability Statistics

KR-20	N of Items
.730	40

Table 5: Test Norms Analysis Table.

S/N	Class	AGE	BSI Score	Rank	Percentile Rank BSI	Z_score BSI	T_Score BSI	Grade Equivalent	Age Equivalent
1	JSS3	Above 15yrs	25	23	18.8	1.0	59.8	8.9	13.8
2	JSS3	Above 15yrs	24	26	21.7	.8	58.4	8.8	13.6
3	JSS3	Above 15yrs	23	30	24.6	.7	57.0	8.7	13.4
4	JSS3	Above 15yrs	19	49	40.4	.1	51.3	8.3	12.6
5	JSS3	Above 15yrs	17	62	51.7	-.2	48.5	8.1	12.2
6	JSS3	Above 15yrs	16	69	57.5	-.3	47.1	8.0	12.0
7	JSS3	Above 15yrs	15	77	63.8	-.4	45.6	7.9	11.8
8	JSS3	Above 15yrs	11	104	86.3	-1.0	40.0	7.5	11.0
9	JSS3	13-15yrs	36	1	.8	2.5	75.5	10.0	16.0
10	JSS3	13-15yrs	32	3	2.1	2.0	69.8	9.6	15.2
11	JSS3	13-15yrs	31	6	5.0	1.8	68.4	9.5	15.0
12	JSS3	13-15yrs	31	6	5.0	1.8	68.4	9.5	15.0
13	JSS3	13-15yrs	30	11	8.8	1.7	66.9	9.4	14.8
14	JSS3	13-15yrs	30	11	8.8	1.7	66.9	9.4	14.8
15	JSS3	13-15yrs	29	14	11.3	1.6	65.5	9.3	14.6
16	JSS3	13-15yrs	13	89	74.2	-.7	42.8	7.7	11.4
17	JSS3	13-15yrs	27	17	13.8	1.3	62.7	9.1	14.2
18	JSS3	13-15yrs	26	19	15.8	1.1	61.3	9.0	14.0
19	JSS3	13-15yrs	25	23	18.8	1.0	59.8	8.9	13.8
20	JSS3	13-15yrs	23	30	24.6	.7	57.0	8.7	13.4
21	JSS3	13-15yrs	22	34	27.9	.6	55.6	8.6	13.2
22	JSS3	13-15yrs	5	118	98.3	-1.9	31.4	6.9	9.8
23	JSS3	13-15yrs	18	55	45.8	.0	49.9	8.2	12.4
24	JSS3	13-15yrs	17	62	51.7	-.2	48.5	8.1	12.2
25	JSS3	13-15yrs	16	69	57.5	-.3	47.1	8.0	12.0
26	JSS3	13-15yrs	14	83	69.2	-.6	44.2	7.8	11.6
27	JSS3	13-15yrs	11	104	86.3	-1.0	40.0	7.5	11.0
28	JSS3	13-15yrs	10	109	90.4	-1.1	38.5	7.4	10.8
29	JSS3	13-15yrs	9	111	92.5	-1.3	37.1	7.3	10.6
30	JSS3	13-15yrs	30	11	8.8	1.7	66.9	9.4	14.8
31	JSS3	13-15yrs	27	17	13.8	1.3	62.7	9.1	14.2
32	JSS3	13-15yrs	26	19	15.8	1.1	61.3	9.0	14.0
33	JSS3	13-15yrs	25	23	18.8	1.0	59.8	8.9	13.8
34	JSS3	13-15yrs	24	26	21.7	.8	58.4	8.8	13.6
35	JSS3	13-15yrs	23	30	24.6	.7	57.0	8.7	13.4
36	JSS3	13-15yrs	22	34	27.9	.6	55.6	8.6	13.2
37	JSS3	13-15yrs	21	38	31.3	.4	54.2	8.5	13.0
38	JSS3	13-15yrs	18	55	45.8	.0	49.9	8.2	12.4
39	JSS3	13-15yrs	17	62	51.7	-.2	48.5	8.1	12.2
40	JSS3	13-15yrs	16	69	57.5	-.3	47.1	8.0	12.0
41	JSS2	Above 15yrs	20	43	35.4	.3	52.7	8.4	12.8
42	JSS2	Above 15yrs	18	55	45.8	.0	49.9	8.2	12.4

43	JSS2	13-15yrs	31	6	5.0	1.8	68.4	9.5	15.0
44	JSS2	13-15yrs	30	11	8.8	1.7	66.9	9.4	14.8
45	JSS2	13-15yrs	29	14	11.3	1.6	65.5	9.3	14.6
46	JSS2	13-15yrs	21	38	31.3	.4	54.2	8.5	13.0
47	JSS2	13-15yrs	20	43	35.4	.3	52.7	8.4	12.8
48	JSS2	13-15yrs	20	43	35.4	.3	52.7	8.4	12.8
49	JSS2	13-15yrs	19	49	40.4	.1	51.3	8.3	12.6
50	JSS2	13-15yrs	18	55	45.8	.0	49.9	8.2	12.4
51	JSS2	13-15yrs	17	62	51.7	-.2	48.5	8.1	12.2
52	JSS2	13-15yrs	17	62	51.7	-.2	48.5	8.1	12.2
53	JSS2	13-15yrs	16	69	57.5	-.3	47.1	8.0	12.0
54	JSS2	13-15yrs	15	77	63.8	-.4	45.6	7.9	11.8
55	JSS2	13-15yrs	11	104	86.3	-1.0	40.0	7.5	11.0
56	JSS2	13-15yrs	11	104	86.3	-1.0	40.0	7.5	11.0
57	JSS2	13-15yrs	8	114	95.0	-1.4	35.7	7.2	10.4
58	JSS2	13-15yrs	23	30	24.6	.7	57.0	8.7	13.4
59	JSS2	13-15yrs	22	34	27.9	.6	55.6	8.6	13.2
60	JSS2	13-15yrs	20	43	35.4	.3	52.7	8.4	12.8
61	JSS2	13-15yrs	19	49	40.4	.1	51.3	8.3	12.6
62	JSS2	13-15yrs	18	55	45.8	.0	49.9	8.2	12.4
63	JSS2	13-15yrs	17	62	51.7	-.2	48.5	8.1	12.2
64	JSS2	13-15yrs	15	77	63.8	-.4	45.6	7.9	11.8
65	JSS2	13-15yrs	15	77	63.8	-.4	45.6	7.9	11.8
66	JSS2	13-15yrs	14	83	69.2	-.6	44.2	7.8	11.6
67	JSS2	13-15yrs	13	89	74.2	-.7	42.8	7.7	11.4
68	JSS2	13-15yrs	12	96	80.0	-.9	41.4	7.6	11.2
69	JSS2	13-15yrs	11	104	86.3	-1.0	40.0	7.5	11.0
70	JSS2	13-15yrs	9	111	92.5	-1.3	37.1	7.3	10.6
71	JSS2	13-15yrs	31	6	5.0	1.8	68.4	9.5	15.0
72	JSS2	13-15yrs	24	26	21.7	.8	58.4	8.8	13.6
73	JSS2	13-15yrs	18	55	45.8	.0	49.9	8.2	12.4
74	JSS2	13-15yrs	20	43	35.4	.3	52.7	8.4	12.8
75	JSS2	13-15yrs	19	49	40.4	.1	51.3	8.3	12.6
76	JSS2	13-15yrs	16	69	57.5	-.3	47.1	8.0	12.0
77	JSS2	10-12yrs	16	69	57.5	-.3	47.1	8.0	12.0
78	JSS2	10-12yrs	15	77	63.8	-.4	45.6	7.9	11.8
79	JSS2	10-12yrs	13	89	74.2	-.7	42.8	7.7	11.4
80	JSS2	10-12yrs	12	96	80.0	-.9	41.4	7.6	11.2
81	JSS1	13-15yrs	32	3	2.1	2.0	69.8	9.6	15.2
82	JSS1	13-15yrs	26	19	15.8	1.1	61.3	9.0	14.0
83	JSS1	13-15yrs	25	23	18.8	1.0	59.8	8.9	13.8
84	JSS1	13-15yrs	22	34	27.9	.6	55.6	8.6	13.2
85	JSS1	13-15yrs	21	38	31.3	.4	54.2	8.5	13.0
86	JSS1	13-15yrs	20	43	35.4	.3	52.7	8.4	12.8
87	JSS1	13-15yrs	19	49	40.4	.1	51.3	8.3	12.6

88	JSS1	10-12yrs	18	55	45.8	.0	49.9	8.2	12.4
89	JSS1	10-12yrs	17	62	51.7	-.2	48.5	8.1	12.2
90	JSS1	10-12yrs	15	77	63.8	-.4	45.6	7.9	11.8
91	JSS1	10-12yrs	28	15	12.5	1.4	64.1	9.2	14.4
92	JSS1	10-12yrs	15	77	63.8	-.4	45.6	7.9	11.8
93	JSS1	10-12yrs	15	77	63.8	-.4	45.6	7.9	11.8
94	JSS1	10-12yrs	14	83	69.2	-.6	44.2	7.8	11.6
95	JSS1	10-12yrs	14	83	69.2	-.6	44.2	7.8	11.6
96	JSS1	10-12yrs	14	83	69.2	-.6	44.2	7.8	11.6
97	JSS1	10-12yrs	13	89	74.2	-.7	42.8	7.7	11.4
98	JSS1	10-12yrs	13	89	74.2	-.7	42.8	7.7	11.4
99	JSS1	10-12yrs	13	89	74.2	-.7	42.8	7.7	11.4
100	JSS1	10-12yrs	13	89	74.2	-.7	42.8	7.7	11.4
101	JSS1	10-12yrs	12	96	80.0	-.9	41.4	7.6	11.2
102	JSS1	10-12yrs	12	96	80.0	-.9	41.4	7.6	11.2
103	JSS1	10-12yrs	12	96	80.0	-.9	41.4	7.6	11.2
104	JSS1	10-12yrs	12	96	80.0	-.9	41.4	7.6	11.2
105	JSS1	10-12yrs	11	104	86.3	-1.0	40.0	7.5	11.0
106	JSS1	10-12yrs	11	104	86.3	-1.0	40.0	7.5	11.0
107	JSS1	10-12yrs	19	49	40.4	.1	51.3	8.3	12.6
108	JSS1	10-12yrs	11	104	86.3	-1.0	40.0	7.5	11.0
109	JSS1	10-12yrs	31	6	5.0	1.8	68.4	9.5	15.0
110	JSS1	10-12yrs	21	38	31.3	.4	54.2	8.5	13.0
111	JSS1	10-12yrs	10	109	90.4	-1.1	38.5	7.4	10.8
112	JSS1	10-12yrs	9	111	92.5	-1.3	37.1	7.3	10.6
113	JSS1	10-12yrs	8	114	95.0	-1.4	35.7	7.2	10.4
114	JSS1	10-12yrs	7	116	96.7	-1.6	34.3	7.1	10.2
115	JSS1	Below 10yrs	6	117	97.5	-1.7	32.8	7.0	10.0
116	JSS1	Below 10yrs	4	119	99.2	-2.0	30.0	6.6	9.6
117	JSS1	Below 10yrs	3	120	100.0	-2.1	28.6	6.3	9.4
118	JSS1	10 - 13yrs	16	69	57.5	-.3	47.1	8.0	12.0
119	JSS1	10 - 13yrs	12	96	80.0	-.9	41.4	7.6	11.2
120	JSS1	10 - 13yrs	8	114	95.0	-1.4	35.7	7.2	10.4

position of each student among the 120 students the test was administered on. Looking at table 3 above, it could be observed that only a student had a rank of 1, percentile rank of 0.8% and had the highest standard score (t-score= 75.5), which implies that he or she performed well than the other 119 students in the test. For the grade equivalent, the results in table 3 above reveals that out of the 40 students sampled in J.S.3, only 1 (2.5%) had grade equivalent of 10.0 which implies that his/her performance is beyond its level and its corresponds with that of students in SS1, 11(27.5%) of them elicit performance that correspond with students in J.S.3 (Grade 9), 20 (50.0%) of them elicit

performance of students in J.S. 2 (Grade 8), 7 (17.5%) of the students elicit performance of students in J.S.1 (Grade 7), while the remaining 1(2.5%) elicit performance which correspond with student in primary 6. By implication, out of the forty (40) J.S. 3 students assess with the standardized test, only 12 qualifies to be in J.S.3 (Grade 9). For J.S.2, out of the 40 students sampled, 4 (10.0%) of the students' performed excellently and their performance correspond with students in J.S.3 (Grade 9), 22(55.0%) elicit performance that correspondent with their actual class (J.S.2/Grade 8), while the remaining 14(25.0%) of the students elicit performance of students in J.S.1 (Grade

Table 6: Coding Criteria equivalent Norms grade and age

S/N	Grade Equivalents Coding Criteria		
	Raw Scores Range	Corresponding Grade/Class	Grade Coding
1	Below 6	Pry 6	6.0 – 6.9
2	6 – 15	J.S. 1	7.0 – 7.9
3	16 – 25	J.S.2	8.0 – 8.9
4	26 – 35	J.S. 3	9.0 – 9.9
5	Above 35	SS1	10.0
Age Equivalents Coding Criteria			
Raw Scores Range		Corresponding Age	Age Coding
1- 5		Below 10 years	9.0 – 9.8
6 – 10		10 years	10.0 -10.8
11 – 15		11 years	11.0 – 11.8
16 – 20		12 years	12.0 – 12.8
21 – 25		13 years	13.0 – 13.8
26 – 30		14 years	14.0 – 14.8
31 – 35		15 years	15.0 – 15.8
Above 35		16 years and above	16.0

7). By implication, out of the forty (40) J.S. 2 students examined with the standardized test, only 26 are qualified to be in J.S.2, while the remaining 14 are not qualified to be in J.S.2. Similarly, in J.S. 1, out of the 40 students assess, 4(10.0%) perform excellently and their performance correspond with student in J.S.3 and J.S. 2 respectively, 24(60.0%) of the students perform with their class performance range (Grade 7/J.S.1), while the remaining 2(5%) of the students elicit performance of students in primary 6 (Grade 6). By implication, it could be inferred that only 5% of students in J.S.1 is not qualified to be in grade level 7 (J.S.1) while the remaining 95% are qualified to be in J.S.1.

In the case of age equivalent norms, there exists variation in students' performance with reference to age. Where students who performed poorly in the test were regarded as students whose performance correspond with students within the age range of 10 years or below. Average students in the test were student's whose performance correspond with students within the age range of 13 – 15 years while students who performed excellent in the test were students whose performance correspond within the age range of 15 years and above.

However, it should be noted that both age and grade equivalents norms are not as good as the standard scores or percentile rank in explaining variation of students' performance but they are quite good in explaining how an individual student performed among his age or educational level status.

4. Conclusion

The findings of study established that the selected standardized test items at the end of the study not only

poses good psychometric properties but also poses good test quality in terms of test manual, usability and ability to reduce guessing. This suggests that the developed instrument is valid and reliable enough, to be used to measure learners' achievement in J.S.S Basic Science. The findings of this study also have educational implications for teachers, and others in education sector. The developed, validated and standardized instrument will be of immense benefit to teachers especially in the area of continuous assessment. The instrument will also serve as a model test to Basic Science teachers and stakeholders in unified and external examination in Junior Secondary school Basic Science.

6. Recommendations

Based on the findings of this research, the following recommendations are proffered.

7. Junior secondary School Basic Science teachers should make use of this standardized test in the assessment of students in both internal and external examination in Basic Science.
8. The standardized test items should be looked upon as a model achievement test in Basic Science for junior secondary schools.
9. Regular sensitization workshop, seminars and conferences should be organized for teachers in order for them to be acquainted with requisite techniques needed for construction of valid assessment instruments.
10. Educational inspectors should improve upon their duties by embarking on consistent, planned and objective inspection particularly in areas of classroom-based assessments.
11. Recruitment of measurement and evaluation experts in state ministry of education would

help foster quality of test items in unified examination in Lagos state.

References

- Adams, J. C. (2008). Effect of physical facilities and problem-based learning aid on knowledge acquisition, knowledge retention, and critical thinking ability of agriculture students in urban schools. *Doctoral dissertation*. University of Missouri, Columbia.
- Adeyemi, B. (2007). School related factors as correlates of pupils' achievement in social studies in South West Nigeria. *Electronic Journal of Research in Educational Psychology*, 8(1), 313-332.
- Afolabi, F. (2009). The Effects of inquiry-based and competitive learning strategies on academic performance of senior secondary school students in physics. *International Journal of Social and Management Sciences* 2, 4-10.
- Ahmed, A. S. (2008). Effect of test environment and duration on students' academic achievement in Biology. *Unpublished M.Ed. Dissertation*, University of Ilorin. 45 - 52.
- Egbule, J. F. (2012). *Fundamentals of Test, Measurement and Evaluation*. Owerri Nigeria, Whyte and Whyte Publishers.
- Frey, B. A. (2007). Knowledge of english language as a predictor of students' cognitive achievement in senior secondary school mathematics. *African Journal of Cross cultural Psychology and Sport Facilitation*, 9:80-84
- Jegede, S. A. (2014). Three Modes of Instruction and Gender on Learning Outcomes in Biology Science. *Journal of Education and Practice*, 5(23), 7 – 13.
- Kareem, A. D. (2013). Influence of Physical classroom environment on students' academic concentration and performance in Science. *Journal of Science Education*, 1(3), 18 21.
- Nworgu, B. G. (2012). *Educational measurement and evaluation: Theory and Practice*. Awka: Hallman Publishers.