

Development and Validation of Scientific Literacy Achievement Test to Assess Senior Secondary School Students' Literacy Acquisition in Physics

A.A Adeleke^{1*} E.O Joshua² (PhD)

1. Department of Teacher Education, University of Ibadan, Nigeria

2. Department of Physics, University of Ibadan, Nigeria

* E-mail of the corresponding author: aadeleke614@stu.ui.edu.ng

Abstract

Physics literacy plays a crucial part in global technological development as several aspects of science and technology apply concepts and principles of physics in their operations. However, the acquisition of scientific literacy in physics in our society today is not encouraging enough to the desirable standard. Therefore, this study focuses on development and validation of scientific literacy achievement test to assess senior secondary school students' literacy acquisition in physics. The study adopted descriptive survey research design in which a total of two hundred and six (206) students from 8 secondary schools in Ibadan North and Oorelope local government areas of Oyo state were assessed. The schools were selected using proportional random sampling techniques. The achievement test titled, Physics-Scientific Literacy Achievement Test (P-SLAT) that had been validated by experts and tested to have Cronbach's alpha and Split-half reliability coefficient of 0.82 and 0.80 respectively. The result of the administered test was collated and coded for analysis, which was carried out with the aid of SPSS 17 package and Microsoft word excel while both descriptive, such as simple percentage and charts were adopted. The result of the research findings shows that. The overall acquisition of scientific literacy in physics among the students is average. The achievement test used has an average reliability coefficient of 0.70 based on Cronbach's alpha, Kuder-Richardson (KR20) and Split-Half method. Based on the three psychometric properties (reliability, item difficulty and discrimination) considered in item analysis of the achievement test developed in the study, eighteen (18) items which is equivalent to 72% of the total items form the validated achievement test. These findings provide vivid evidence for recommendations for various examining bodies by subjecting multiple choice items to psychometric tests to ascertain and improve items' quality in order to obtain a reliable and valid test result. It was also suggested for prospective researchers in the same or related field of study that this research should be carried out in all secondary schools in Oyo state in order to have a more generalised result on the level of secondary school students' acquisition of scientific literacy in physics.

Keywords: Scientific literacy, Item analysis, difficulty index, discrimination index, reliability coefficient.

1. Introduction

Science is made up of huge number of fields such as Physics, Chemistry, Biology and lots more. According to Folajimi and Emmanuel (1999), in recent times, at the secondary education level, the integrated science curriculum is designed to introduce all the students to the whole of science which would provide relevant foundation on which the senior secondary will build careers in science or science related disciplines, while the senior secondary level pupils are introduced to different disciplines. In order to realize this, emphasis on school science has been shifted to realization of scientific literacy acquisition.

Among the two broad purposes of science education is to promote scientific literacy among citizens on matters directly affecting their own lives and the society so that they can make decisions based on information and understanding. This is essential for the sustainable development of a modern, technological society. The second acclaimed purpose of science education is to build up technological capability by equipping the future workforce with essential science-based knowledge and skills, and by preparing students for scientific disciplines in higher education and science-related careers. Given the potential benefits, the provision of quality science education to all children will have far reaching consequences on a country's development prospect.

Akindehin (1996) identified and measured acquisition of scientific literacy based on the following five (5) skills which he considered as scales relevant for measuring scientific literacy in physics: The Science Knowledge I (SKI), Science Knowledge II (SKII), Science Knowledge III (SKIII), Science Habit of the Mind (SHM), Solving Problems and Conducting Inquiries (SPC). The objectives measured by these skills respectively include information about nature phenomena that are usually emphasized in physics lessons; fundamental concepts, principles and theories in physics; nature of science, its characteristics such as empirical and theoretical methods and philosophy that distinguish science from other human activities; ability to think scientifically and the inclination to do so beyond the confines of the science classroom; and students interaction with the natural world.

As a matter of fact, the acquisition of scientific literacy in school science teaching is one of the most

widely accepted educational goals. The world over has its aims at making the learners able to cope and contribute to scientific enterprises from day-to-day in a world dominated by science and technology (Folajimi and Emmanuel, 1999).

According to Halliday, Resnick and Walker (2008) Physics, like all other sciences is based on experimental observations and quantitative measurements. The main objective of physics is to find the limited number of fundamental laws that govern natural phenomena and to use them to develop theories that can predict the results of future experiments. The fundamental laws used in developing theories are expressed in the language of mathematics, the tool that provides a bridge between theory and experiment. When a discrepancy between theory and experiment arises, new theories must be formulated to remove the discrepancy. Many times, a theory is satisfactory only under limited conditions; a more general theory might be satisfactory without such limitations.

Classical physics, according to Halliday, Resnick and Walker (2008) means all of the physics developed before 1900, includes the theories, concepts, laws, and experiments in classical mechanics, thermodynamics, and electromagnetism. Important contributions to classical physics were provided by Newton, who developed classical mechanics as a systematic theory and was one of the originators of calculus as a mathematical tool. Major developments in mechanics continued in the 18th century, but the fields of thermodynamics and electricity and magnetism were not developed until the latter part of the 19th century, principally because before that time the apparatus for controlled experiments was either too crude or unavailable.

A new era in physics, usually referred to as *modern physics*, began near the end of the 19th century. Modern physics developed mainly because of the discovery that many physical phenomena could not be explained by classical physics. The two most important developments in modern physics were the theories of relativity and quantum mechanics. Einstein's theory of relativity revolutionized the traditional concepts of space, time, and energy; quantum mechanics, which applies to both the microscopic and macroscopic worlds, was originally formulated by a number of distinguished scientists to provide descriptions of physical phenomena at the atomic level (Halliday, Resnick and Walker, 2008).

Scientists constantly work to improving our understanding of phenomena and fundamental laws, and new discoveries are made every day. In many research areas, a great deal of overlap exists between physics, chemistry, geology, and biology, as well as engineering. Some of the most notable developments are; numerous space missions and the landing of astronauts on the Moon, micro circuitry and high-speed computers, and sophisticated imaging techniques used in scientific research and medicine.

It was further emphasized by Halliday, Resnick and Walker that the impact which such developments and discoveries had on our society has indeed been great, and it is very likely that future discoveries and developments will be just as exciting, challenging and of great benefit to humanity if scientific knowledge of physics is improved.

Physics literacy is crucial for effective living in the modern age of science and technology. Given its application in industry and many other professions, it is necessary that every student is given an opportunity to acquire some of its concepts, principles and well-grounded skills. Unfortunately, the teaching and learning of physics has been fraught with challenges which prevent many students from acquisition of appropriate scientific skills generally and that of physics in particular. Despite the fact that the philosophy, objectives and content of physics curriculum have been adjudged by professionals in the field to be satisfactory enough, yet its implementation has fallen short of expectation because of the lack of sufficient number of qualified physics teachers among other factors which affect acquisition of proper scientific skill teaching in order to enhance meaningful learning whereas the nature of the subject thereby appearing to evoke difficulty.

Valladolid (2010) emphasised that in the new framework of instruction, student is the primary agent of learning. The framework includes the following areas: the students' learning process, the teacher's work, the students' learning environment, and assessment of students learning. This framework emphasizes the interrelationship between students' learning and assessment which can be verified through achievement test or other form of assessment procedures.

A test refers to a set of questions or items designed and to be responded to by one or a group of individuals within a specified period of time. Tests are used mainly as measures of achievement, to show how much an individual has learnt or not learnt in an aspect of instruction to which he has been exposed. It is therefore a major instrument in the process of evaluation. It then makes it imperative for us as a teacher or teacher trainer to be conversant with all forms of test and testing procedures (Aina, 2006).

According to Charles (2007), the importance of tests in educational settings need not be over-emphasized as tests are used to improve an individual, inform him of how he is faring in a learning situation and to develop his intellect. Achievement tests are most useful in educational settings for analysing causes of academic failure, evaluating learning outcomes, curriculum evaluation, and motivation of learners as well as vocational counselling. Charles further argued that whenever we want to evaluate students against criteria that go beyond a single teacher's classroom and those teachers' conceptions of what should be taught. There is need for

a standardized test in addition to having norms - that is, bases for interpreting scores in terms of other students' scores because standardized tests are usually more carefully constructed than teacher made tests and the construction is carried out by experts, using the technical, statistical and research knowledge of the testing field. Hence, standardized test are different from conventional school (teacher-made) tests because they have norms, better constructed by experts, accompanied by manuals and have good psychometric properties.

Standardized tests usually are accompanied by detailed instructions about how the test should be administered - the exact directions to be given to students, the time limits (if any) and the ways in which the teacher should handle any special problem that may arise. All these essential information are contained in an accompanying booklet called the *test manual*. Based on the above argument one can deduce that standardized test is significantly different from the teacher-made test in terms of the content measured, student preparation, quality of the items, administration, scoring as well as interpretation.

Therefore, the functions of standardized test are those roles which it plays in a wide range of situations in educational, social, industrial or clinical contexts. Hence, one should however note that the usage of standardized tests overlap considerably since we can use one type of test in a number of situations perhaps a number of tests can be used for one particular situation.

The types of the teacher-made achievement test used in the Nigerian educational system are broadly categorised into two: the essay and objective type. These two broad categories are usually conveniently compared on the following parameters: ability being measured, coverage of the test, ease of preparing, scoring and motivation of learners. The steps in constructing achievement tests are diverse but sequential. The first which is a statement of objective must be done in terms of student behaviour, with an active verb, that is observable and must be realistic. Other steps include the specification of content to be covered, development of the test blueprint in which the content should be restricted to the topic areas covered and constructing achievement tests, which include: generation of items, item analysis, establishment of psychometric properties and norms. The test blue print is a two-fold table that shows the distribution of the items in a test, covering the course content and tapping the scientific skills to be measured. To make the blue print an effective guide to enable one ascertain content coverage, the type of items desired, the length of the test and the difficulty level of the items.

In constructing teacher-made tests, some guidelines are general to teacher-made tests like essay and objectives test items which include drafting the exercises well in advance, having the items critiqued by a colleague, generation of surplus items and keep the test blue-print as a constant guide. The items should be distributed across all cognitive skills: knowledge, comprehension, application analysis, synthesis and evaluation or other specific skills to be measured. This is done with due cognizance of the Bloom's Taxonomy of Educational Objectives which encourage test constructors to tap a range of cognitive skills from knowledge through synthesis to evaluation (Charles, 2007). Test scoring is the process of obtaining data from testees' protocols. Scoring can be done by Hand or by Machine while interpretation is the process of organizing data or test scores in order to help the process of making deductions by adopting the micro or macro model of test interpretation.

The setting of examinations in various subjects by an examination board external to the school is common around the world. The examinations are taken at the end of a course, or school curriculum, and determine success or failure for future advancement and are often referred to as once only, high stakes, examinations. These can involve objective testing, often multiple choice (MCQ) items, and/or more subjective questions where students construct the responses themselves. Where MCQ items are involved, the scoring is dichotomous, undertaken by machine and reliability is less of a concern than validity of the items chosen. As it is generally accepted that valid instruments are important, essay-type questions still remain an obvious choice for evaluation of knowledge. Unfortunately, allocating marks for essay-type, student free-response questions can be very unreliable, unless marking systems are well developed (Baird, Greatorex, & Bell, 2004).

Charles (2007) opined that a number of guidelines apply to all the formats. Among these are; commensurate difficulty level of items, consistency of the correctness of answer tapping of salient content area and independence of each item. Like all measuring instruments, achievement test is supposed to reveal the best possible results. This can be attained by ensuring that certain conditions are involved with the testing process at their optimum. Choices of hall and advance preparation are some of these factors as the venue must be environmentally conducive to elicit the desired response. Advance preparation in terms of the test materials is essential.

Based on the above premise, the role of test in evaluation cannot be brushed aside. It is a major instrument for evaluation. Because, from testing results data are obtained which are formulated into measurements while the judgment or decisions made from such formulations and other measured observation produces evaluation. Therefore, test and testing processes cannot be by-passed in the process of evaluation.

The psychometric properties of a test refer to the qualities expected of a test and basic desirable qualities that a test must possess to be deemed satisfactory. These qualities referred to as the psychometric

properties of a test hinge on the validity and reliability of the test. Validity and reliability are important psychometric properties in the literature of educational measurement and evaluation with the former being the more important as reliability is an estimate. The relationship between validity and reliability is that, a test may be reliable but not valid. But a test may not be valid unless it is reliable. Hence, a test must possess both in order to be deemed satisfactory for testing purposes.

Validity refers to the extent to which a test measures what it is supposed or purported to measure operationally, the extent to which the results of an evaluation programme serve the particular uses for which they are intended. In other words, it is the degree to which correct inferences may be drawn from test scores. If the results are to be used to describe pupil achievement, one would like them to represent the specific achievement we wish to describe, and to represent nothing else. The desires for valid achievement test in this regard are like that of the lawyer in the court room who wants the truth, the whole truth and nothing but the truth. Hence, validity is a matter of degrees which does not exist in an all-or-none basis. Thus, validity is best considered in terms of categories that specify degree such as high validity, moderate validity or low validity (Charles, 2007).

A test is said to be valid to the degree that it measures what it claims to measure, or, in the case of predictive validity, to the extent that it predicts accurately such types of behaviour as academic and job success/failure, as well as stability or instability under stress. Tests are often validated by correlating test scores against some outside criteria, which may be scores on tests of accepted validity, successful performance or behaviour, or the expert judgment of recognized authorities (John W. Best, James V. Kahn, 1959). Psychometricians believe that there is so much problem in establishing validity that the following characteristics may be used to describe validity types; Concurrent, Criterion-related, Congruent, Discriminant, Construct, Ecological, Content, Face Convergent Factorial, Predictive, Intrinsic et cetera (Charles, 2007).

Validity is the consistency of test scores and the extent which tests measure what they are originally designed for. Very many types of these qualities abound, and they can all be estimated to ensure that tests possess the qualities that will make them provide us with dependable scores. Reliability is usually taken to mean the reproducibility of scores on another occasion. Reliability has been defined by (Berkowitz, Wolkowitz, Firch, & Kopriva, 2000) as the degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable and repeatable for an individual test taker. An index of reliability of a test for a given population can be taken to be the ratio of true score variance to observed score variance (Dimitrov, 2002). True scores cannot be directly determined and hence the reliability is an estimate. And of course, this assumes that the scores were obtained from a test that was seen as sufficiently valid for the reliability of the results to have any meaning at all. A very reliable Physics test would be totally out-of-place as a reliable instrument for the determination of history ability! But if validity is at an acceptable and interpretable level, then reliability is an important consideration and challenge facing all examination boards as well as educational evaluators (Brooks, 2004).

In statistics, the Kuder-Richardson Formula 20 (KR-20) first published by Kuder, G.F. & Richardson, M.W. in 1937 is a measure of internal consistency (reliability) for measures with dichotomous choices. Kuder-Richardson formula has two versions (KR-20 and KR-21) for achievement and psychological test items respectively. These reliability techniques were named after its two developers (Kuder and Richardson). The formula for estimating reliability is given by:

$$KR20 = \frac{n}{n-1} \left(\frac{SD^2 - \sum PQ}{SD^2} \right)$$

Where n = number of test items, SD^2 = variance of scores on the test [square of the SD (standard deviation)], P = proportion of testees who responded correctly, Q = proportion of testees who responded wrongly.

Unlike KR-20; KR-21 formula assumed that all items comprise equal difficulty, so it is peculiar to psychological test reliability estimation. The formula for calculating KR21 is given by:

$$KR21 = \frac{n}{n-1} \left(1 - \frac{M - \frac{(M)^2}{n}}{SD^2} \right)$$

Where n = number of items in the test, M = mean of the scores on the test, SD^2 = variances of scores on the test [square of the SD (standard deviation)] (Wiseman, 1999 in Sabri, 2013).

Lee Joseph Cronbach (1916-2011) is a prominent American Psychologist and Mathematician who developed one of the current and most acceptable methods of determining (internal consistency) reliability of measuring instruments such as an achievement, aptitude, personality, attitude test, other forms of tests used in science and social science research. This competition on reliability developed and came out with by Lee Joseph Cronbach in 1951 is popularly called **Cronbach's alpha reliability test**. The reliability analysis of items can be carried out conveniently and easily on the popularly known Statistical Package for Social Science (SPSS) within a very short period of time after the appropriate coding and entering of data has been done. Other methods of determining reliability include; split half method and test-retest correlation method.

Adeleke (2010) opined that the greater the number of similar items, the greater the internal consistency. That's why a researcher sometimes gets very long scales asking a question a myriad of different ways, so that addition of more items results to a higher Cronbach's alpha coefficient. He added that alpha of 0.80 is considered as a reasonable benchmark. Cronbach's alpha reliability coefficient is given by the formula;

$$\alpha = \frac{K}{K-1} \left(1 - \frac{\sum_{i=1}^K \sigma_i^2}{\sigma_x^2} \right)$$

Where α = estimate of Cronbach's alpha reliability, K = Number of items and σ^2 = Variance, σ = Standard Deviation.

The implications of item analysis are significant for achievement testing, for example item difficulty and discrimination index. The difficulty index is important because it helps to know how well students have acquired the knowledge they are supposed to. The difficulty of a test actually varies with its use. A routine end of term test needs not to be difficult. But a competitive screening examination should be considerably of higher difficulty level. The range of normal item difficulty index (P) is given by the following inequality expression ($0.4 \leq P \leq 0.7$) = ($40\% \leq P\% \leq 70\%$) which implies that items with difficulty index below 0.4, (0.4 to 0.7), and above 0.7 is said to be difficult, moderate and simple for the target population respectively as represented in the figure below.

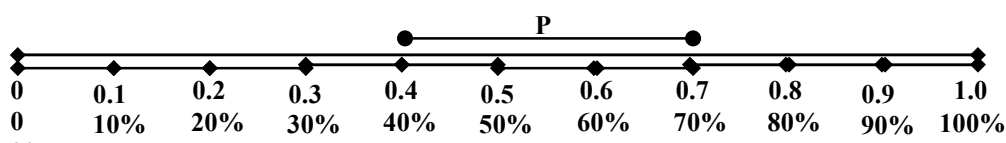


Figure 1: Number line showing the range of normal item difficulty index

Sabri (2013) also reported that based on recommendations by Instructional Assessment Resources (IAR), test items were classified into three categories in terms of level of difficulty as Easy (Above 0.90), Moderate (0.20-0.90) and Difficult (below 0.20).

The discrimination index is used to measure the extent to which the test has distinguishes between higher and lower students' achiever groups. Nenty (2000) in Ojerinde (2013) opined that it indicates the degree to which examinees' response to an item varies with, or relates to their trait level or ability. Charles (2007) also noted that discriminative index helps to remove or eliminate indiscriminate items since they serve little purpose in the test while the distractor analysis helps us to eliminate the distractors which are ineffective. Ojerinde (2013) further showed that in test development and item evaluation some items do have negative discrimination indexes and this would have accounted for the low reliability encountered. Therefore, such items should be modified, dropped or replaced. He further stressed that result has indicated that the Item Response Theory (IRT) method used in analysing the Pre-test data on UTME Physics was more reliable than the Classical Test Theory (CTT) approach.

Matlock-Hetzel (1997) also clarified two ways of determining discriminating power of test items, to be exact the discrimination index and discrimination coefficient. Matlock-Hetzel (1997), Si-Mui and Rasiah (2006), Mitra, Nagaraja, Ponnudurai and Judson (2009) and Boopathiraj and Chellamani (2013) defines item discrimination as a measure used to discriminate between students in the top with that of the low group who obtained the correct responses. Fundamentally, the discrimination index differentiates students who are knowledgeable and those who are not, meticulously revealing top scorers and low scorers achievement in each item. The value of discrimination index ranges between -1.0 to 1.0.

The challenges faced by scientific literacy assessment in general and physics literacy in particular among other things include the poor acquisition of scientific literacy among secondary school students which scholars have attributed to many reasons such as poor testing procedures with items that are deficient of essential psychometric properties. Thus, in an attempt to address these infringing issues, a need is felt to develop and validate a scientific literacy achievement test to assess Senior Secondary School Students' acquisition of scientific literacy in physics. The development of such assessment test would be coupled with appropriate Test Blue-Print (table of specification) in order to find out to what extent has the five (5) aforementioned scientific literacy skills identified by Akindehin (1996) been acquired by senior secondary students preparing for their Senior School Certificate Examination. Also, it should be noted that because physics is taught at senior secondary level of education which actually marks the foundation of physics knowledge in the first instance as a complement and to build on the already built background during the integrated (now basic) science that serves as introduction to some physics concepts, knowledge and skills. Therefore, this achievement test will give us a picture and provide answer to the extent of achievement of the following objectives given by Nigeria Educational Research and Development Council (NERDC, 2009) in the Nigeria Physics Curriculum document for Senior Secondary Schools.

1. provision of basic literacy in physics for functional living in the society;
2. acquisition of basic concepts and principles of physics as a preparation for further studies;
3. acquisition of essential scientific skills and attitudes as a preparation for technological application of physics; and
4. stimulate and enhance creativity.

2. Purpose of the study

The purpose of the study was to develop and validate scientific literacy achievement test to assess senior secondary school students' literacy in physics, carry out item analysis (difficulty and discrimination index) and estimate the reliability coefficient of the achievement test.

3. Research questions

The following research questions will be answered in the course of this study.

R_{Q1}: What is the overall acquisition of scientific literacy among the senior secondary school students?

R_{Q2}: What is the reliability of the scientific literacy achievement test items?

R_{Q3}: What is the difficulty index of the scientific literacy achievement test items?

R_{Q4}: What is the discrimination index of the scientific literacy achievement test items?

4. Research Method

4.1 Research design

The study is a descriptive survey research designed to develop and validate scientific literacy achievement test that will be apt to assess senior secondary school students' acquisition of scientific literacy in physics. The descriptive survey research method is considered appropriate for this study as it would help to obtain first-hand information regarding the students' level of acquisition of scientific literacy in physics.

The population comprised of science students offering physics at the senior secondary school level selected from private and public secondary schools in Oyo state. The students in the final class, senior secondary school three (SS 3), were used because they had completed the senior secondary two and were therefore familiar with the course content at the same time preparing for their Senior School Certificate Examination (SSCE).

The research sample used was made up of two hundred and six (206) students proportionally drawn from 8 (eight) secondary schools from Ibadan-North and Oorelope local government areas of Oyo state. The schools are strictly those that offered sciences particularly physics at the senior secondary school level.

4.2 Instrument

The research instrument titled "Physics - Scientific Literacy Achievement Test (P-SLAT)" was used for this study. As shown in the table 1 (Test Blue-Print or Table of Specification) below, the instrument consists of 25 (twenty-five) objective test items developed by the researcher were designed with four options lettered A-D out of which only one was the correct as the acceptable answer based on all the six (6) themes of physics identified in the Nigeria Senior Secondary School Physics Curriculum and the five (5) scales identified by Akindehin (1996) for measuring the acquisition of scientific literacy.

Table 1: Test Blue-Print (Table of Specification) for the Physics - Scientific Literacy Achievement Test

S/N	CONTENT AREA	SCIENTIFIC LITERACY OBJECTIVES					TOTAL
		SKI	SKII	SKIII	SHM	SPC	
1.	Interaction of Matter, Space and Time	1	1	-	1	-	3
2.	Conservation Principles	1	1	1	1	1	5
3.	Waves: Motion without Material Transfer	1	1	1	1	1	5
4.	Fields at Rest and in Motion	1	1	1	-	1	4
5.	Energy Quantization and duality of Matter	1	-	1	1	1	4
6.	Physics in Technology	-	1	1	1	1	4
TOTAL		5	5	5	5	5	25

Scales for measuring Scientific Literacy Acquisition in Physics

Akindehin (1996) identified five (5) scales for measuring the acquisition of scientific literacy. The scales and the corresponding scientific literacy objective measured by each scale are shown in the Table 2 below.

Table 2: Descriptive, Information for Scientific Literacy Achievement Test

S/N	SCALE NAME	SCIENTIFIC LITERACY OBJECTIVES MEASURED
1.	Science Knowledge I (SKI)	Information about nature phenomena that are usually emphasized in physics lessons.
2.	Science Knowledge II (SKII)	Fundamental concepts, principles and theories in physics.
3.	Science Knowledge III (SKIII)	Nature of science, its characteristics such as empirical and theoretical methods and philosophy that distinguish science from other human activities.
4.	Science Habit of the Mind (SHM)	Ability to think scientifically and the inclination to do so beyond the confines of the science classroom.
5.	Solving Problems and Conducting Inquiries (SPC)	Students' interaction with the natural world.

Source: Akindehin, 1996

In the Table 3 below, sample items from each of the above scales are presented as used in the Physics - Scientific Literacy Achievement Test (P - SLAT) developed for validation in this study.

Table 3: Sample Items for Each of the Scientific Literacy Scales

S/N	SCALE NAME	SAMPLE ITEM
1.	Science Knowledge I (SKI)	A wave travelling from water to glass suffers a change in its speed at the common boundary. Which of the following properties explains this observation? (A) Dispersion (B) Refraction (C) Interference (D) Diffraction
2.	Science Knowledge II (SKII)	Laws of refraction state the following except; (A) the incident ray, the refracted ray and the normal ray at the point of incidence all lie in the same plane (B) $[\sin i / \sin r]$ is constant for the media concerned (C) the incident angle and the reflected angle are equal (D) refractive index $(n) = \text{speed of light in vacuum} / \text{speed of light in the medium}$
3.	Science Knowledge III (SKIII)	The sagging of overhead electrical cables is the consequence of (A) linear expansivity (B) superficial expansivity (C) cubic expansivity (D) conductivity
4.	Science Habit of the Mind (SHM)	Which of the following reasons explains why the walls and ceilings of a standard concert hall are usually covered with perforated pads? To (A) increase the intensity of sound waves (B) increase the loudness of the sound waves (C) reduce the effect of reverberation of sound waves (D) decrease the frequency of sound waves.
5.	Solving Problems and Conducting Inquiries (SPC)	When the pressure of a fixed mass of gas in a bicycle pump is doubled at constant temperature, the volume of the gas in the pump is (A) doubled (B) halved (C) increased four times (D) unchanged.

Source: Akindehin, 1996

4.3 Procedures

The researcher visited the public and private secondary schools within Ibadan North and Oorelope local government areas and administered the achievement tests to the Senior Secondary three (3) physics students after which permission have been sought from the school principal and the physics teachers in charge. The students were instructed to attempt all the question items to the best of their knowledge under the supervision of the teachers in charge, researcher and his research assistant while the attempted copies of the achievement test items were collected back immediately.

5. Results

5.1 Research question 1: What is the overall acquisition of scientific literacy in physics among senior secondary school students?

Table 4: Statistical summary of the students' overall scientific literacy acquired

Mean	Median	Mode	Range	Variance	Std. Deviation	Total Item	Total No. of Students
12.7	13.0	14.0	19.0	15.3	3.9	25	206

Table 4 above presents the statistical summary of the students' score in the achievement test. The summary reveals as a measure of central tendency to describe the students' score that, the mean (average), median, mode is 12.7, 13.0 and 14.0 respectively. By measure of dispersion, the range, variance and standard deviation is 19.0, 15.3 and 3.9 respectively.

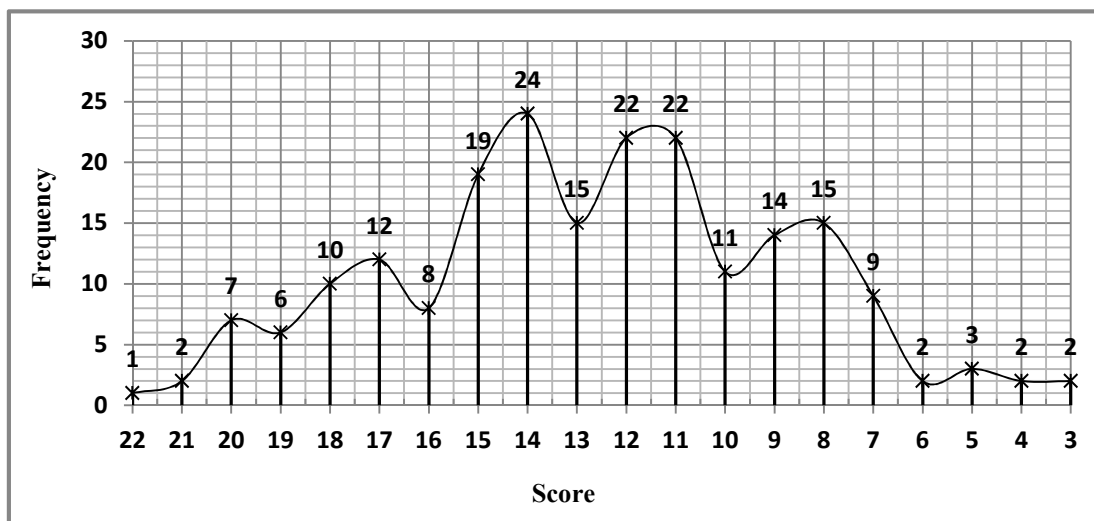


Figure 2: Graphical illustration of distribution of students' score in the achievement test

The graphical illustration shown in figure 2 above presents the distribution students' score (performance) as a measure of overall acquisition of scientific literacy in physics among senior secondary school students based on the assessment carried out by adopting the developed test, Physics-Scientific Literacy Achievement Test (P-SLAT) comprising of twenty-five (25) question items, hence the maximum obtainable students' score is 25. The graph further reveals that the highest and lowest score is 22 and 3 respectively.

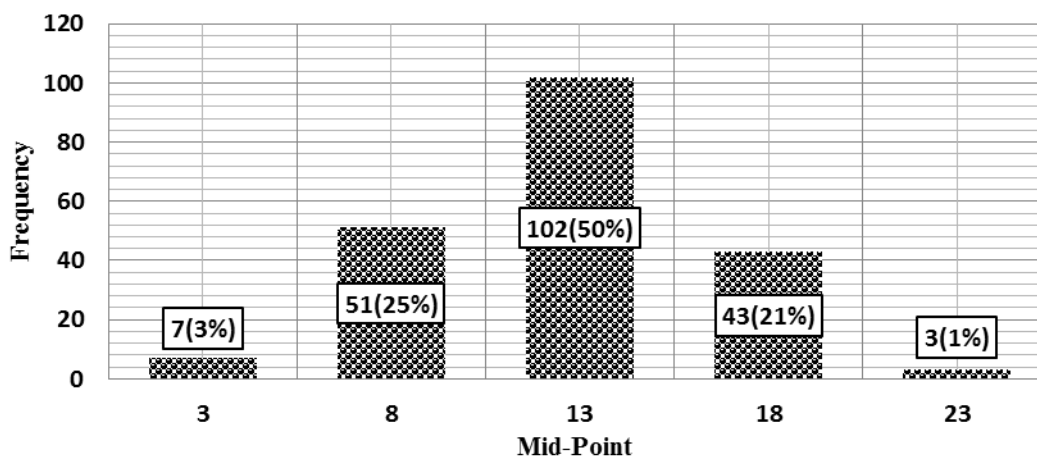


Figure 3: Bar chart illustrating the distribution of students' score in the achievement test

Figure 3 above reveals that the mean, median and mode of the students' score fall in to the same score interval. It can be deduced that the overall acquisition of scientific literacy in physics is average because the score interval (11-15) where the students' mean, median and modal score falls to, which has a mid-point value of 13 and has accounted for the highest frequency count of 102 students which is equivalent to 50% of the total number of 206 students assessed is the study. It also depicts that 58 students which is equivalent to 28% of the total students,

score below the average group while a frequency count of 46 students which is equivalent to 22% of the total students has accounted for those that score above the average.

Hence, this finding is justified to conclude that the overall acquisition of scientific literacy in physics among senior secondary school students is average.

5.2 Research question 2: What is the reliability of the scientific literacy achievement test items?

Table 5: Reliability coefficient of items under each scale and overall items of the physics literacy test

SCALES	Reliability Methods		
	Cronbach's alpha	Kuder-Richardson (KR20)	Split-Half
SKI	0.353	0.358	0.374
SKII	0.570	0.573	0.556
SKIII	0.031	0.037	0.045
SHM	0.277	0.282	0.220
SPC	0.443	0.447	0.327
Overall	0.682	0.684	0.601
Average	0.70		

Figure 4: Graphical illustration of the overall reliability coefficient of the achievement test items based on the three methods of reliability estimates

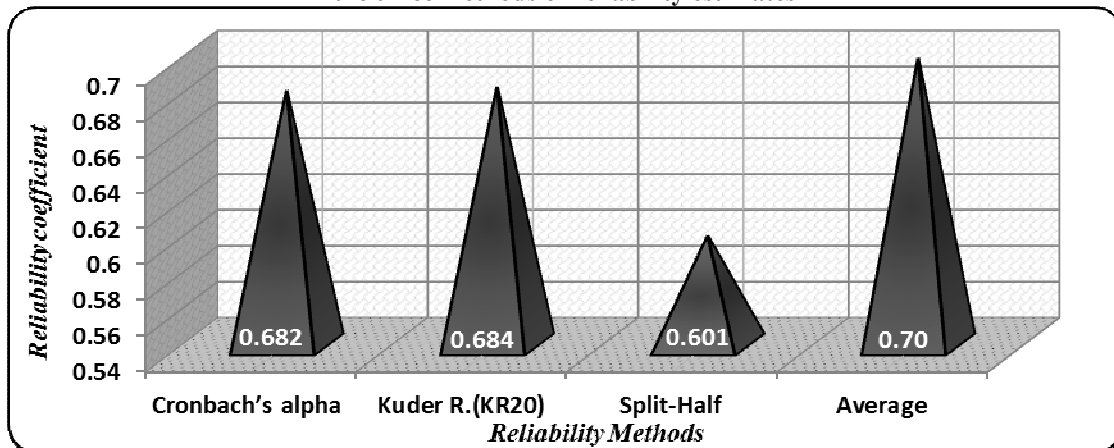


Table 5 and figure 4 above presents the corresponding reliability coefficients of all the five scales in the literacy achievement test as estimated by Cronbach's alpha, Kuder-Richardson (KR20) and Split-Half method as 0.682, 0.684 and 0.601 respectively, while the average reliability based on the three methods for the Literacy Achievement Test developed in this study is approximately 0.70.

Figure 5: Graphical illustration of the reliability coefficients of each scale of the achievement test based on the three methods of reliability estimates

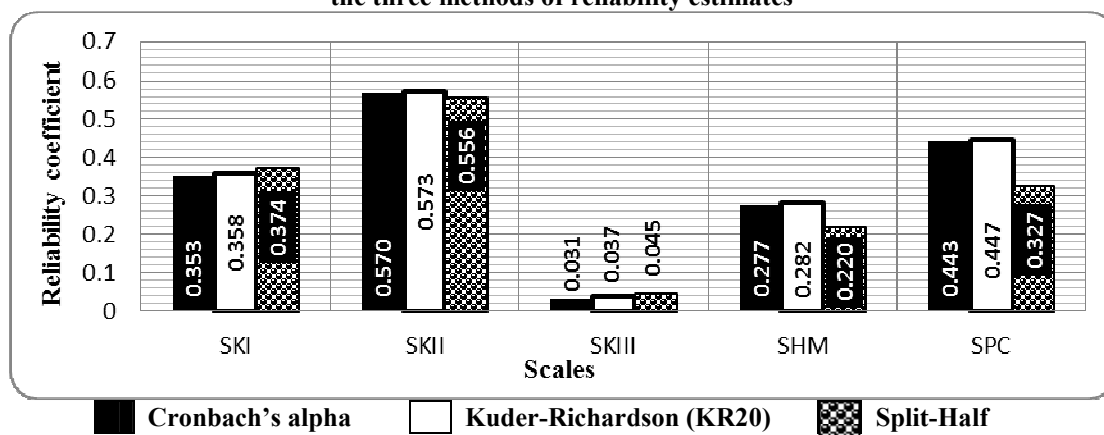


Figure 5 above presents the graphical illustration of the distribution of the corresponding reliability coefficients

of each of the five scales in the literacy achievement test in accordance with Cronbach's alpha, Kuder-Richardson (KR20) and Split-Half methods respectively.

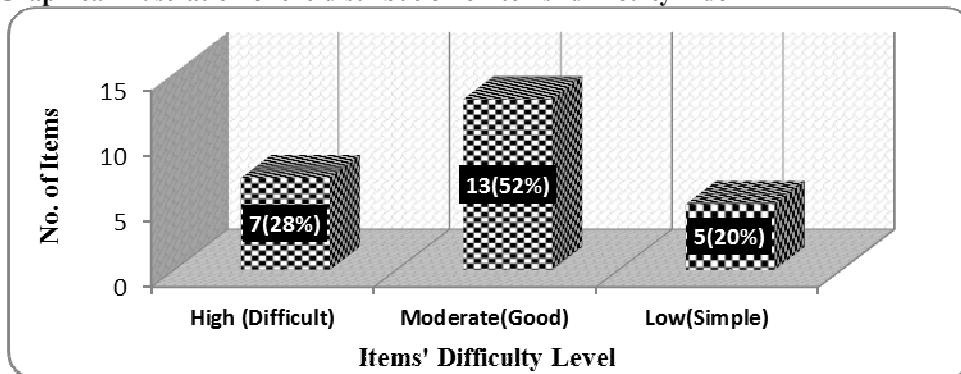
5.3 Research question 3: What is the difficulty index of the scientific literacy achievement test items?

Table 6: Distribution of items' difficulty index

Difficulty index range	No. of items	Percent	Cumulative percent	Difficulty level	Item evaluation
0.00 - 0.40	7	28.0	28.0	High	Difficult
0.41 - 0.70	13	52.0	80.0	Moderate	Good
above 0.7	5	20.0	100.0	Low	Simple
Total	25	100.0			

Table 6 above presents the distribution of difficulty index of all the twenty five (25) items in the achievement test vis-à-vis their respective index range, the corresponding number of items in various ranges and percentage of each difficulty level as well item evaluation. The table reveals that 7(28%), 13(52%) and 5(20%) items have their difficulty index between 0.00-0.40, 0.41-0.70 and above 0.7 and evaluated to be difficult, good and simple items respectively

Figure 6: Graphical illustration of the distribution of items' difficulty index



Graphical illustration in figure 6 above further shows that 7(28%) of the items have low difficulty index value hence they are regarded as being difficult items, 13 items which is equivalent to 52% of the total items have moderate difficulty index, so they are refer to as Good items, while 5 items which is equivalent to 20% of the total items have highest value of difficulty index, hence they are said to be simple items.

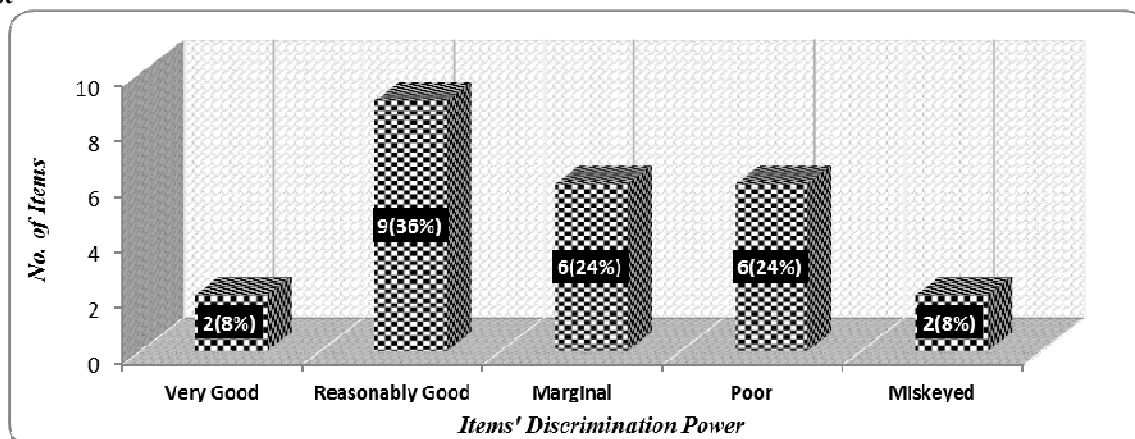
5.4 Research question 4: What is the discrimination index of the scientific literacy achievement test items?

Table 7: Distribution of discrimination index of the achievement test items

Discrimination index	No. of items	Valid percent	Cumulative percent	Item evaluation
above 0.40	2	8.0	8.0	Very Good
0.30-0.39	9	36.0	44.0	Reasonably Good
0.20-0.29	6	24.0	68.0	Marginal
Below 0.19	6	24.0	92.0	Poor
Negative value	2	8.0	100.0	Miskeyed
Total	25	100.0		

Table 7 presents the distribution of discrimination index of the achievement test items. It shows that 2(8%) items each have their discrimination index above 0.04 and as negative value, 6(24%) items each have discrimination index falling into the range of 0.20-0.29 and below 0.19 while 9(36%) items have their discrimination index between the ranges of 0.30-0.39.

Figure 7: Graphical illustration of the distribution of discrimination index of the items in the achievement test



The figure above shows the results of the item analysis in terms of distribution of discrimination power. It further reveals that 2(8%), 9(36%) and 6(24%) items are those that can be described in terms of discrimination power as being Very Good, Reasonably Good and Marginal, while 6 and 2 items which is equivalent to 24% and 8% of the total items are respectively regarded as being poor and Miskeyed. Therefore, none of the items have zero (0) discrimination value, thus, none of the items may be described as being “inadvertently Miskeyed” or “intrinsically ambiguous”.

6. Discussion

The findings in this study implies that the overall acquisition of scientific literacy in physics among senior secondary school students is average because the score interval (11-15) has accounted for the highest frequency count of 102 students which is equivalent to 50% of the total number of 206 students used for the study. The average acquisition of scientific literacy in physics among secondary school students is also supported by the fact that 58 students which is equivalent to 28% of the total students, score below the average group while a frequency count of 46 students which is equivalent to 22% of the total students has accounted for those that score above the average. Effective physics teaching is recognized as one of the major instrument for scientific literacy acquisition, which can be used to develop global, technological and scientific advancement. For this to be achieved, the issue of poor acquisition of scientific literacy in physics should be addressed in public and private schools in the nation at large because researches have proved that several factors such as insufficient and unqualified teachers as well as inadequate infrastructures among others constitute to the problems of poor acquisition of scientific literacy in physics.

Anakwe (2008) revealed in his study that students' academic achievement below average is linked with the student's loss of control and poor school adjustment found in their study habits. Anakwe supported the findings in this study that performance of students in different technical colleges is below expectation, where the mean and the standard deviation analysis of students' academic achievement indicated that the general proportion of the students had a high failure rate in the examination where majority of them were found to be 50 percent below average performance. In another support to the findings in this study is Akanbi, (2006) who reported that low output of public and private secondary schools has necessitated the argument for the Post University Matriculation Examinations conducted by most of the tertiary institutions in Nigeria. This is for the stakeholders to evaluate the students' feedback on what is supposed to have been taught in the classroom setting.

The findings in this study depicts that the reliability coefficient of the achievement test based on Cronbach's alpha, Kuder-Richardson (KR20) and Split Half method is 0.682, 0.684 and 0.601 respectively. The reliability coefficient of each of the five (5) scales is based on the three (3) stated methods of reliability estimate adopted. It was found out that each of the scales portray lesser reliability as reflected by the reliability coefficients because each scale of the test is made up of five questions each which is not enough to obtain a better reliability coefficient. Since, the average reliability of the three methods adopted is 0.70, the achievement is reliable enough and homogenous to access senior secondary school students acquisition of scientific literacy in physics.

The average reliability coefficient obtained in this study is supported to be appropriate by Ehigie (2012) who stated that good tests have reliability coefficients which ranges from as low as 0.65 to above 0.90 (the theoretical maximum is 1). This is further supported by a commonly accepted rule of thumb which uphold that alpha of 0.6-0.7 indicates acceptable reliability. The findings in this study is also in consonance with Fraenkel and Wallen (2008) who reported that one should attempt to generate a KR20 reliability coefficient of 0.70 and

above to acquire reliable score which appears to be reliable thus revealing that this comprehensive test is a reasonably reliable instrument. The reason behind smaller reliability coefficient values of each scale compared to the overall reliability coefficient is because of the lesser number of items which is involved in each scale compared to the larger number of items used in the achievement test as a result of the integration of all the scales' items. Ehigie, (2012), also supports this findings and stated that "the more items there in a scale designed to measure a particular concept, the more reliable will the measurement (sum scale be)" which implies that, the larger the number of items in a test, the higher the chance of obtaining a better and more reasonably higher reliability coefficient for such a test. Hence, a test consisting of 40 items is more homogenous than a test consisting of 10 items. In the later there might be little or no relationship between the testees performance on different types of items. Unless the test items are highly homogenous, the Kuder Richardson coefficient will be lower than split-half reliability. But when highly heterogeneous, the split-half coefficient will be higher than KR coefficient. This proved the homogeneity of the test developed.

The closer value (equal approximation) of Cronbach's alpha and Kuder Richardson (KR20) reliability coefficient is also supported by Ehigie (2012) that Cronbach's alpha when computed for binary (e.g true or false) items, is identical to the so called Kuder-Richardson (KR20) formular of reliability for sum scales. Another reason that supported this is reported in Wikipedia (2014) that Kuder-Richardson (KR20) is a derivative of the Cronbach's alpha formula, in other words Cronbach's alpha is sophisticated version of Kuder-Richardson (KR20) formular.

All the items in the achievement test are categorised into three (3) based on the difficulty index range. The items with difficulty index range of 0.00-0.40, 0.41-0.70 and those whose difficulty indices are above 0.7 are regarded to have high, moderate and low difficulty index respectively. Hence, these items are described to be difficult, good and simple where 7(28%), 13(52%) and 5(20%) items fall into this category respectively.

The seven (7) difficult items include item 3, 11, 12, 13, 14, 19, and 22 whereas the 13 good items are item 2, 6, 7, 8, 9, 15, 16, 17, 18, 21, 23, 24, 25. The five (5) simple items which include item 1, 4, 5, 10 and 20 may be discarded such that remaining 80% of the total items which is equivalent to a numerical count of 20 items are suitable for assessment in physics achievement test as justified from the level difficulty criteria. Therefore, 20 items that are recommended in terms difficulty level include, item 2, 3, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24 and 25 form the items of the validated test with suitable psychometric property measured in term of difficulty index.

In agreement to this item analysis, Sabri (2013) reported that the results of students' achievement in comprehensive test were utilized to determine the quality of each particular item and item difficulty lends a hand in distinguishing easy item from difficult ones. Instructional Assessments Resources (IAR), (2013) insinuates the usage of easy question as warm up questions in assessing student mastery. The items that were classified as difficult items ought to be reconsidered in terms of language and content appropriateness. A low value of difficulty index may possibly indicate a Miskeyed item. Additionally, it may also indicate that the tested topic were inappropriate. In spite of this, difficult questions which results in frustration for some students function as challenge among the top student as evokes by Schreyer Institute for Teaching Excellence (2013).

The Miskeyed two (2) items and include item 14 and 19 may be discarded such that the remaining 92% of the total items that is equivalent to 23 items which include item 1-13, 15-18 and 20-25. These items are said to be suitable and form the items of the validated test with suitable psychometric property in terms of discrimination power. If the 6 poor items that is equivalent to 24% of the total items which include item 1, 3, 5, 11, 12, and 21 were to be discarded in addition to the two Mikeyed items (i.e. 14 and 19), it implies that a total of 8 items will be removed, which include item 1, 3, 5, 11, 12, 14, 19 and 21 and accounted for a total 32% of the total items in the achievement test. It implies that the following seventeen (17) items (item 2, 4, 6-10, 13, 15-18, 20 and 21-25) which is equivalent to 68% of the total items form the items of the validated test with suitable psychometric property in terms of discrimination power and were to be retained. Since, the poor items only have low discrimination power and show low discrimination between the upper and the lower groups i.e. they are neither Miskeyed nor inadvertently Miskeyed. So, there is no need to remove them but may be reframed or restructured in subsequent assessment.

The five categories of discrimination in terms of discrimination coefficient adopted in this study are based on recommendation by Ebel (1972) in Ovwigho (2013) and they include Very Good (above 0.40), Reasonably Good (0.30-0.39), Marginal (0.20-0.29), Poor (Below 0.19). Stanley and Hopkin, (1972) described the items with zero or negative discrimination value as being "inadvertently Miskeyed" or "intrinsically ambiguous". Since, none of the items have zero (0) as their discrimination value; hence none of the items may be described to be "inadvertently Miskeyed" or "intrinsically ambiguous". These items could only be revealed by item analysis of this type which is recommended to be rejected as part of subsequent test items.

The findings is also similar to a study conducted by El-Uri and Malas (2013) while analysing undergraduate examination in obstetrics and gynaecology, it was recommended that test items with negative discrimination coefficient poor and negative discrimination coefficient should be highlighted for reviewing

purpose. A poor discriminating power might signify confusing items which were ambiguously worded or indicates a Miskeyed item. Ultimately, items with negative coefficient should be removed from the comprehensive test. Si Mui & Rasiah and Matloct-Hetzel coincide in the reasoning of the negative value, they proposed that student in the low achievement group often make a guess in answering the easy question and by chance come up with the correct answer. Contradictory, students in the upper achievement group embark upon the easy question too vigilantly and end up choosing the wrong answer. In another support to the findings of this study, El-Uri and Malas, and Ovwigho put forward that items with negative discrimination coefficient should be eliminated from the test. The reason is that item with negative discrimination coefficient indicates students with low score got the item right and students with high score answer the item incorrectly.

7. Conclusion

This study was guided with four (4) research questions and one of the conclusions drawn is that, the overall acquisition of scientific literacy in physics among senior secondary school students could be described as being average, as this is evident from 50% of students which have their score clustered around average of the total students assessed. Another conclusion drawn is that the average reliability coefficient of the achievement test items is 0.70 as estimated from Cronbach's alpha, Kuder-Richardson (KR20) and Split-Half reliability estimate method while Split-half method yield a low reliability coefficient value compared to others. Removal of item 14 as recommended in the SPSS analysis reliability result will improve the test split half reliability coefficient from 0.601 to a better value of 0.709.

It could also be concluded from the study that twenty (20) items which is equivalent to 80% of the total items in the achievement test are psychometrically appropriate in terms of difficulty. Twenty-three (23) items which is equivalent to 92% of the total items are psychometrically appropriate in terms of item discrimination. It was also concluded based on these three psychometric properties; reliability, item difficulty and discrimination that, a total of eighteen (18) items which is equivalent to 72% of the total items form the validated version of Physics-Scientific Literacy Achievement Test (P-SLAT) developed in the study. These include, item 2, 3, 6- 9, 11-13, 15-18 and 21-25. Therefore, item 1, 4, 5, 10, 14, 19 and 20 were discarded and recommended to be subjected to reviewer while others that satisfy the criteria as earlier listed were retained in accordance with the following criteria:

- ✓ Difficulty index must be fall the range, 0.00-0.7.
- ✓ Discrimination power must be a positive.
- ✓ Recommendation from SPSS analysis on item removal that will mostly yield best reliability coefficient.

8. Recommendations

Having undergone this study, efforts were suggested, necessitating coordination and collaboration between different stakeholders: teachers, teacher trainers, administrative staff, examination bodies and government towards improved and effective teaching of physics as well as development of valid and reliable tests items with good psychometric properties in order to ensure effective, valid and reliable assessment of general scientific literacy skills in general and physics in particular.

Based on the above premise, the following recommendations were made;

1. All multiple choice exams conducted in at all levels of education should be subjected to psychometric test for reliability, difficulty index, discrimination power and distractor analysis as this will go a long in improving and ascertaining the items' quality, validity and reliability of learning outcome.
2. There should be intervention by government and University Physics Departments in sponsor of development and provision for physics curriculum running that will lead students to develop 21st century scientific skills which will make them acquire values of good citizenship and contribute to technological and scientific development.
3. Government should help to provide necessary facilities and materials for effective physics teaching, such facilities and materials should include well equipped laboratories and well stocked libraries with current text books and internet facilities.
4. Educational institutes have a role to play in advancing more interest and awareness in the application of Item Response Theory (IRT) by making it core modules in undergraduate programmes and should sponsor graduate research and retraining of staff through local and international workshops and seminars on how to develop valid and reliable tests.
5. Government should sponsor and give scholarship opportunities to candidates willing to study physics education in Universities and to outstanding and well performed students' in physics as this will go in sustaining their interest towards scientific literacy acquisition in physics.

References

Adeleke, J.O. (2010). *The Basics of Research and Evaluation Tools*. Ikeja-Lagos: Somerest Ventures.

- Aina, M.E. (2006). Unpublished lecture note on Instructional Techniques and Methods in School Mathematics Curriculum. Victoria Island, Lagos: National Open University of Nigeria. Retrieved from: www.nou.edu.ng.
- Akanbi, S.T. (2006). Teacher as Determinants of Academic Performance of Students in Public Examinations, A Paper Presented at the WAEC Monthly Seminar, Held on Friday 31st June.
- Akindehin (1996)
- Anakwe, A.I. (2008). The Resilient Effects of Locus of Control and School Adjustment on Academic Performance of Senior Secondary-Student in Plateau State. *Nigerian Journal of Teacher Education and Teaching*. 5.2.
- Baird, J.A., Grotorex, J., and Bell, J. P. (2004). What Makes Marking Reliable? Experiments with UK Examinations. *Assessment in Education*. 11.3:331-348.
- Berkowitz, D., Wolkowitz, B., Firch, R., and Kopriva, R. (2000). The Use of Tests as Part of High-Stakes Decisions for Students: A Resource Guide for Educators and Policy-Makers. Washington DC: US Dept. of Education.
- Boopathiraj, C. and Chellamani, K. (2013). Analysis of Test Items on Difficulty Level and Discrimination Index in the Test for Research in Education. *International Journal of Social Science and Interdisciplinary Research*. 2.2.
- Brooks, V. (2004). Double Marking Revisited. *British Journal of Educational Studies*. 52.1:29-46.
- Charles, B.U.U. (2007). Measurement and Evaluation in Educational Practices. Ibadan: Distance Learning Centre, University of Ibadan.
- Dimitrov, D.M. (2002). Reliability: Arguments for Multiple Perspectives and Potential Problems with Generalization Across Studies. *Educational Psychological Measurement*, 62.5:783-801.
- Ehigie, B.O. (2012). *Psychological Tests and Testing (PSY 503)*. Ibadan: Distance Learning Centre, University of Ibadan.
- El-Uri, F.I. and Malas, N. 2013. Analysis of Use of a Single Best Answer Format in an Undergraduate Medical Examination. *Qatar Medical Journal*. 1:14 ISSN: 2201-6333 (Print) ISSN: 2201-6740 (Online). Retrieved From: www.ijern.com
- Folajimi, A. and Emmanuel, S. (1999). Development and Validation of Scientific Literacy Achievement Test to Assess Senior Secondary School Students' Acquisition of Scientific Literacy in Physics. *Journal of the Science Teachers Association of Nigeria*. 34.1&2:96-101.
- Fraenkel, J.R. and Wallen, N.E. (2008). *How to Design and Evaluate Research in Education (7th ed.)*. New York: McGraw-Hill.
- Halliday, D., Resnick, R. and Walker, J. (2008). *Fundamentals of Physics, 8th Edition*. United States: John Wiley & Sons, Inc. and Von Hoffmann Press.
- Instructional Assessment Resources. 2011. *Item Analysis*. Retrieved November 9, (2013). from University of Texas at Austin, Instructional Assessment Resources, IAR Web site: <http://www.utexas.edu/academic/ctl/assessment/iar/students/report/itemanalysis.php>
- John W.B., James, V.K. (1959). *Research in Education. 8th edition*. United States of America: Allyn and Bacon (A Viacom Company). Retrieved from: www.abacon.com.
- Matlock-Hetzel, S. (1997). Basic Concepts in Item and Test Analysis. Paper Presented at the Annual Meeting of the Southwest Educational Research Association, Austin.
- Mitra, N.K, Nagaraja, H.S, Ponnudurai, G. and Judson, J.P. (2009). The Levels of Difficulty and Discrimination Indices in Type A Multiple Choice Questions of Pre-clinical Semester 1 Multidisciplinary Summative Tests. *E-Journal of Science, Medicine and Education*. 3.1:2-7.
- Nigerian Educational Research and Development Council (NERDC), (2009). *Senior Secondary Education Physics Curriculum*. Federal Ministry of Education, Nigeria.
- Obemeata, J.O. 1991. A Re-appraisal of Secondary School Curriculum in Nigeria. Paper presented at the National School Curriculum Review Conference Proceedings, Kaduna 2nd-6th September. Lagos, Nigeria: Federal Ministry of Education Press.
- Ojerinde, D. 2013. Classical Test Theory (CTT) Vs Item Response Theory (IRT): An Evaluation of the Comparability of Item Analysis Results. Lecture Presentation at the Institute of Education, University of Ibadan. May 23.
- Ovwohio, B.O. 2013. Empirical Demonstration of Techniques for Computing the Discrimination Power of a Dichotomous Item Response Test. *IOSR Journal of Research and Method in Education*. 3.2:12-17.
- Sabri, S. (2013). Item Analysis of Student Comprehensive Test for Research in Teaching Beginner String Ensemble Using Model Based Teaching among Music Students in Public Universities. *International Journal of Education and Research*. 1.12.
- Schreyer Institute for Teaching Excellence, (2013). *Improve Multiple Choice Tests Using Item Analysis*. Retrieved November 28 from Pennsylvania State University, Schreyer Institute for Teaching Excellence Web site: <http://www.schreyerinstitute.psu.edu/Tools/ItemAnalysis/>

Sim, S.M and Rasiah, R.I. (2006). Relationship between Item Difficulty and Discrimination Indices in True/False Type Multiple Choice Questions of a Para-clinical Multidisciplinary Paper. *Ann Acad. Med Singapore*; 35:67-71.

Valladolid, V.C. (2010). An Evaluation of the Teaching-Learning Process: A Transformative Learning Set-up. *Educational Measurement and Evaluation Review*. 1:150-160

Wikipedia, (2014). The Free Encyclopedia, Retrieved on 17 May at 08:44am.

The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage:

<http://www.iiste.org>

CALL FOR JOURNAL PAPERS

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

Prospective authors of journals can find the submission instruction on the following page: <http://www.iiste.org/journals/> All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: <http://www.iiste.org/book/>

Academic conference: <http://www.iiste.org/conference/upcoming-conferences-call-for-paper/>

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digital Library, NewJour, Google Scholar

