ISSN 2705-2559
E-ISSN 2705-2567

ITEM RESPONSE THEORY ESTIMATION OF A SENIOR SECONDARY SCHOOL CERTIFICATE PHYSICS PRACTICAL EXAMINATION

BY
Oluseyi Peter Adeoye: Department of Educational Foundations, Faculty of Education, Obafemi Awolowo University, Ile-Ife; E-mail: seyipeter680@gmail.com
&
Alaba Adeyemi Adediwura: Department of Educational Foundations, Faculty of Education, Obafemi Awolowo University, Ile-Ife; E-mail: yemtoy20002000@gmail.com

Abstract
The research was conducted using a descriptive survey research design. The study's population were all Osun State senior secondary school three Physics students. Six hundred and thirty-one senior secondary school Physics students were chosen using a multistage sampling technique for the study sample. NECO 2018 physics practical examination was used to collect relevant data for the study. The -2loglikelihood chi-square value, Generalized Partial Credit Model, Standard Error of Estimate (SE), standardized fit index (Zh) and limited-information fit measure were used to analyze the collected data. The results showed that only five out of the 37 the tasks required to judge the proficiency of students in Physics practical were of adequate level of difficulty. However, 28 out of the 37 tasks required to judge the proficiency of students in Physics practical has high discrimination parameter estimates that were greater than the benchmark of 0.35. The result also showed that out of the 37 tasks used in assessing the proficiency of students in Physics practical by NECO, 34 items fitted the generalized partial credit model used in the calibration of the test. The study concluded that the psychometric properties of 2018 NECO physics practical examination items were adequate but with deficiency in the difficulty level. The study therefore concluded that information provided by IRT model on the estimates of the test were dependable and can be reliably used in the evaluation of the quality of Physics practical test items.
Keywords: Item Difficulty, Item Discrimination, Item Response Theory, Item Fit Item Parameter

Introduction
Physics is a branch of science that aims to comprehend the material universe. In order to achieve this understanding, Physicists use experiments to question nature. These experiments are hands-on research that aims to test existing ideas and point to more powerful theories. However, experiments are not only essential in expanding our knowledge of our universe, but play a key role in the teaching of Physics (Chris & Vollmer, 2006). Physics' general objective is to use observation, experimentation, and theoretical formulation to understand and explain numerous physical processes that occur in nature and in the laboratory. The motion of planets around the sun, evaporation of water, sound output from a tuning fork, refraction of light, attraction of iron by magnets, discharge of an electrical capacitor, and decay of the pi meson are all instances of physical processes (Agrawal & Menon, 2010). As a result, relevant Physics learning is attained by careful structuring of Physics curricula and their connections to environmental challenges. The critical objective of Physics in secondary school is the acquisition of manipulative or practical skills. These skills enable students to apply their Physics knowledge to everyday challenges. Furthermore, these skills contribute in the comprehension of Physics concepts and their application in daily life (Zdenek & Hana, 2008).

Practical work in the school laboratory not only helps students learn Physics ideas and rules, but it also allows them to practice and manipulating laboratory equipment and apparatus (Abrahams, 2011). Students improve their practical abilities and learn to acquire more trustworthy and valid data by doing practical work on a regular basis. Thus, students' practical skills must be examined on a regular basis to see if they have acquired the abilities needed to execute practical tasks and handle data or information. Physics practical comprises two options

ISSN 2705-2559
E-ISSN 2705-2567

(Alternative A and B) and the questions are organized into three categories: mechanics, light, and electricity, according to the National Examinations Council timetable (2018). Laboratory sessions, which should ideally follow at the end of each module, are, however, sparsely conducted for secondary school students (NECO Chief Examiners Reports 2016, 2017, 2018). This is either due to under-equipped laboratories in most public and private schools, or to teachers' lack of interest or laziness (Ariyo, 2006; Kuti, 2012). Furthermore, the time allotted to Physics on the school timetable (three periods of 40-minute in each week) is insufficient to accommodate laboratory sessions (Chukwuneye, 2011). Chukwuneye (2012) stated that many Physics teachers are more concerned with finishing the scheme of work rather than the quality of their instruction. Therefore, students are deprived of adequate opportunities to develop skills and appropriate attitudes to scientific activity and exploration. Without a doubt, in most Nigerian secondary schools, the lack of a functional Physics laboratory and inadequate equipment or apparatus for Physics practical is hindering laboratory operations, which may be contributing to students' poor performance in Physics.

The psychometric estimation of a test would imply analysing such constituents of psychometrics as (i) Validity – whether a test tests the measurement (ii) Reliability – this is consistency in measuring what it intends to measure (iii) Difficulty index or conversely easiness index (iv) Discrimination index –how sharply does the test distinguish between low and high ability students. it is therefore pertinent that the psychometric properties of practical tests by the examination body such as NECO should be determined using polytomous IRT models, this will indicate the overall quality of assessment/ test conducted by the examination bodies in practical and it will go a long way towards enlisting confidence or otherwise in the examining bodies. The estimation of practical physics test items is the assessment of psychometric properties of physics practical items which is very important because most studies always focused on dichotomous models with few studies on practical aspect of test which usually required the use of polytomous models. It is on this basis that the study estimated psychometric parameters of Physics practical test items in Senior Secondary School Certificate Examination conducted by National Examinations Council among students in Osun state.

Purpose of the Study

The broad objective of this study is to carry out item analysis of a NECO senior secondary school certificate practical Physics examination with the aid of IRT. However, the specific objectives of the study are to:

i. determination of the IRT model that is most suitable for the calibration of the NECO test;
ii. estimate the level of difficulty and discrimination of the Senior Secondary Certificate Examination physics practical test items conducted;
iii. establish items of the SSCE Physics practical test that fit the IRT model;

Research Questions

In order to carry out this study, the following research questions were raised;

i. Which of the IRT models is most suitable for the calibration of the NECO Physics practical test?
ii. What is the level of difficulty of the Senior Secondary Certificate Examination physics practical test items conducted by NECO?
iii. What is the level of discrimination of the SSCE Physics practical test items of NECO?
iv. How many items of the SSCE Physics practical test fit the IRT model?

Methodology

The research was conducted using a descriptive survey research design. The study's population was all Osun State senior secondary school three Physics students. Six hundred and thirty-one senior secondary school Physics students were chosen using a multistage sampling technique for the study sample. Five Local Government Areas (LGAs) were chosen at random from each of Osun State's three senatorial districts. Stratified random sampling was used to choose four schools from each LGA, with school ownership as the basis for stratification. The study employed a 444-student intact class of physics students from each of the selected schools as a sample. The items from the 2018 NECO physics practical examination were used to collect relevant data for the study. The -

ISSN 2705-2559
E-ISSN 2705-2567

2loglikelihood chi-square value, Generalized Partial Credit Model, Standard Error of Estimate (SE), standardized fit index (Zh), limited-information fit measure, and empirical reliability were used to analyze the collected data.

Results

Research Question One: Which of the IRT models is most suitable for the calibration of the NECO Physics practical test?

To answer this question, two stages were involved in the assessment of model-data fit: Fitting the data test to the two available IRT models for essay test which are Partial credit model (PCM) and generalized partial credit model (GPCM) and thereafter, the fitness of the two models to the data set are compared. The model that produced the best fit to the data is adjudged the model that fit the data. To achieve this feat, several measures apply. According to Oguoma, Metibemu and Okoye (2016); (Finch & French (2015), prominent among the measures include Chi-square difference test and use of information indices. Information indices are simply measures of variance not explained by a model, with an added penalty for model complexity. Among the most popular of these indices are the Akaike information criterion (AIC; Akaike, 1973), the Bayesian information criterion (BIC; Schwarz, 1978), and the sample-size-adjusted BIC (SBIC; Enders & Tofighi, 2008). These information indices are computed using the -2loglikelihood chi-square value and is interpreted such that the model with the lower value exhibits a better fit to the data. In addition, the chi-square and likelihood ratio goodness of fit tests the null hypothesis that two nested models provide the same fit to a set of data. A statistically significant likelihood indicates a difference in the models under examination. Table 1 presents the result of the model-data fit assessment.

Table 1: Model-data Fit Assessment

Model	AIC	SABIC	BIC	logLik	X^2	df	p
PCM	33906.62	34017.32	34293.54	-16866.31	1832.385	36	0.000
GPCM	32146.24	32302.74	32693.25	-15950.12			

Table 1 presents the model-data fit assessment, showing the IRT model that is best for the calibration of the practical test. The table shows that when the fitness of Generalized partial credit model (GPCM) and Partial credit Model (PCM) to the data were compared, the result showed that the GPCM had AIC = 32146.24, SABIC = 32302.74, BIC = 32693.25 values that were respectively lesser than the AIC = 33906.62, SABIC = 34017.32, BIC = 32693.25 values of the PCM. In addition, the Likelihood ratio test that GPCM fitted the data better than PCM was statistically significant (χ^2 (59) = 1832.385, $p < 0.05$). These results showed that the GPCM model fitted the data better than the PCM model. Thus, the test was calibrated using Generalized Partial Credit Model and then the difficulty parameters of the items were extracted.

Research Question Two: What is the level of difficulty of the item of the NECO physics practical test items?

To answer this research question, calibration of the test data with the determined model and extraction of the difficulty parameter of the items was carried out. The result is presented in Table 2

Table2: Difficulty Parameters of the NECO SSSCE Practical Test Items

S/N	Item	b1	b2	b3	b4	b5	b6	Remark
1	Q1_ai	-2.51						Good
2	Q1_aii	-0.81	-4.02					Poor
3	Q1_aiii	-2.49	7.96					Poor
4	Q1_aiv	0.43	-4.95					Poor
5	Q1_av	-2.26	11.43					Poor
6	Q1_avi	3.48	-7.96					Poor
7	Q1_avii	-2.78						Good

ISSN 2705-2559
E-ISSN 2705-2567

8	Q1_Graph	-0.6	-1.02	1.41	0.34	-0.32	1.87	Poor
9	Q1_Slope	0.05	0.16					Good
10	Q1_Evaluation	0.72	3.14					Poor
11	Q1_Precaution	0.44	-0.34					Poor
12	Q1_bi	7.9	-3.18					Poor
13	Q1_bii	6.91	0.23					Poor
14	Q2_ai	0.79	-3	-2.21				Poor
15	Q2_aii	-0.43	-2.81	8.45				Poor
16	Q2_aiii	-0.84	-1.3					Poor
17	Q2_aiv	-0.38	-2.11					Poor
18	Q2_av	-1.25	7.57					Poor
19	Q2_avi	-1.83	6.11					Poor
20	Q2_Graph	-0.23	-1.07	2.56	0.1	-0.97	1.7	Poor
21	Q2_Slope	1.02	-0.2					Poor
22	Q2_Precaution	0.41	-0.4					Poor
23	Q2_bi	2.36	-1.55					Poor
24	Q2_bii	3.82	-1.88	5.03				Poor
25	Q3_ai	9.27						Poor
26	Q3_aii	835	-2463.37					Poor
27	Q3_aiii	14.66	-25.62					Poor
28	Q3_aiv	-1.49	25.54					Poor
29	Q3_av	16.62	-61.21					Poor
30	Q3_avi	-23.18	107.62					Poor
31	Q3_Graph	-0.14	-0.92	1.25	1.44	-0.3	1.31	Poor
32	Q3_Slope	0.9	0.17	2.22				Poor
33	Q3_Intercept	0.97						Good
34	Q3_Evaluation	0.71	2.07					Good
35	Q3_Precaution	2.21	-1.32					Poor
36	Q3_bi	1.48	-0.83					Poor
37	Q3_bii	3.36	-1.56					Poor

Table 2 shows the difficulty parameters of the NECO Physics practical test. The difficulty parameters of polytomous test items of easy type are referred to as step difficulty. The step difficulties indicate the point on the ability metric at which the probabilities for adjacent categories is equal (De Ayala, 2009) or point of intersection of two adjacent categories on the ability scale. The average of these step difficulties is the overall difficulty of the items on a test (Paek and Cole, 2020). The categories represent the arrays of scores obtained on the items. In the table there are six step difficulties (b1, b2, b3, b4, b5 and b6) for the items. While some items have only one step difficulty, some have two, others have three and so on up to the highest category, 6. With b1, b2, b3, b4, b5 and b6 represents the step difficulty for score 0 and 1 (i.e., the point of intersection of categories 0 and 1); score 1 and 2, score 2 and 3, score 3 and 4, score 4 and 5 and score 5 and 6 respectively. The last column of the table is the remark. This column presents the evaluation of the difficulty parameter. Items with increasing step difficulties

ISSN 2705-2559
E-ISSN 2705-2567

within -3 to 3 are considered good. The table shows that item 1 has one step difficulty which is b1 = -2.51, which implies that a score of zero and one is equal for examinees with -2.51 level of proficiency in practical physics and thus, the overall difficulty was -2.51. The result showed that a score of 0 and 1 are equally likely for 50% of the examinees with -2.51 proficiency in practical physics. The result showed that 50% of the examinees with -2.51 or greater ability in Physics practical will be able to move from a score of 0 to 1. The implication of the result for this item showed that the item functioned well. As at least half of the examinees with even low ability level will be able to transit from scoring zero to scoring 1.

For item 2, the table shows that the item has two step difficulties. They are: $b_{0\,\&\,1} = -0.81$, $b_{1\&2} = -4.02$. The estimates show that a score of 0 and 1 are equally likely for 50% of the examinees with -0.81 proficiency in practical physics and a score of 1 and 2 are equally likely for 50% of the examinees with -4.02 proficiency in practical physics. The result showed that a score of 0 and 1 are equally likely for 50% of the examinees with -0.81 proficiency in practical physics and a score of 0 and 1 are equally likely for 50% of the examinees with -4.02 proficiency in practical physics. This showed that showed that 50% of the examinees with -0.81 or greater ability in Physics practical will be able to move from a score of 0 to 1 and 50% of those examinees with -4.02 or greater ability in Physics practical will be able to move from a score of 1 to 2. The implication of the result for the item is that the ability required by 50% of the examinees to move from a scoring 0 to scoring 1 was even greater than the ability it required by 50% of the examinees to move from scoring 1 to scoring 2 on the item. This showed that the item was not functioning as expected. As it was observed that examinees on higher ability level had tendency of scoring the in low category scores while examinees with low ability had tendency of scoring high category scores. For item 3, the table shows that the item has two step difficulties. They are: $b_{0\,\&\,1} = -2.49$, $b_{1\&2} = 7.96$. The estimates show that a score of 0 and 1 are equally likely for 50% of the examinees with -2.49 proficiency in practical physics and a score of 1 and 2 are equally likely for 50% of the examinees with 7.96 proficiency in practical physics. The result showed that a score of 0 and 1 are equally likely for 50% of the examinees with -2.49 proficiency in practical physics and a score of 0 and 1 are equally likely for 50% of the examinees with 7.96 proficiency in practical physics. This showed that 50% of the examinees with -2.49 or greater ability in Physics practical will be able to move from a score of 0 to 1 and 50% of those examinees with 7.96 or greater ability in Physics practical will be able to move from a score of 1 to 2.

The table shows that in all only five (1, 7, 9, 33 and 34) out of the 37-task required by the students in the practical test were good. The remaining 32 (2, 3, 4, 5, 6, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 35, 36 and 37) tasks were poor level of difficulty. The result showed that almost all the tasks required to judge the proficiency of students in Physics practical were of inadequate level of difficulty.

Research Question Two: What is the level of discrimination of the NECO SSSCE Physics practical test items? To answer this research question, the discrimination parameters of the items were extracted from the calibrated test. Table 3 presents the discrimination parameters of the NECO physics practical test items.

Table 3: Difficulty Parameters of the NECO SSSCE Practical Test Items

S/N	Item	A	Remark	S/N	Item	a	Remark
1	Q1_ai	1.29	Good	20	Q2_Graph	1.37	Good
2	Q1_aii	0.52	Good	21	Q2_Slope	1.63	Good
3	Q1_aiii	0.74	Good	22	Q2_Precaution	1.55	Good
4	Q1_aiv	0.33	Poor	23	Q2_bi	1.23	Good
5	Q1_av	0.56	Good	24	Q2_bii	0.77	Good
6	Q1_avi	0.28	Poor	25	Q3_ai	-0.14	Poor
7	Q1_avii	0.57	Good	26	Q3_aii	0.00	Poor

ISSN 2705-2559
E-ISSN 2705-2567

8	Q1_Graph	1.75	Good	27	Q3_aiii	-0.11	Poor
9	Q1_Slope	3.68	Good	28	Q3_aiv	-0.05	Poor
10	Q1_Evaluation	3.00	Good	29	Q3_av	-0.07	Poor
11	Q1_Precaution	1.41	Good	30	Q3_avi	0.04	Good
12	Q1_bi	0.33	Poor	31	Q3_Graph	0.96	Good
13	Q1_bii	0.34	Poor	32	Q3_Slope	2.01	Good
14	Q2_ai	0.44	Good	33	Q3_Intercept	2.60	Good
15	Q2_aii	0.63	Good	34	Q3_Evaluation	4.42	Good
16	Q2_aiii	0.56	Good	35	Q3_Precaution	0.75	Good
17	Q2_aiv	0.49	Good	36	Q3_bi	1.21	Good
18	Q2_av	0.70	Good	37	Q3_bii	0.55	Good
19	Q2_avi	0.51	Good				

Table 3 showed the discrimination parameters of the tasks used in judging the performance of students in NECO Physics test. The table contains the number represented by S/N, the description of the test items represented by "Item", the discrimination parameter estimate represented by "a" and the judgment of the items' quality represented by "Remark". The quality of the items was judged based on the criteria established by (Baker, 2001; Hasmy, 2014). According to the authors the following benchmark: Very high discrimination $a \geq 1.7$; High discrimination $1.35 \leq a < 1.7$; Moderator discrimination $0.65 \leq a < 1.35$; Low discrimination $0.35 \leq a < 0.65$; and very low discrimination, $a < 0.35$. Based on the benchmark, discrimination estimate less than 0.35 is considered poor.Table 3 showed that item 1 recorded a discrimination index of 1.29. The result showed that the item adequately discriminates examinees with low ability in physics practical from those that are highly proficient in Physics practical. The table shows that while 28 (1, 2, 3, 5, 7, 8, 9, 10, 11, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 30, 31, 32, 33, 34, 35, 36 and 37) of the items had discrimination parameter estimates that were greater than 0.35, 9 items (4, 6, 12, 13, 25, 26, 27, 28 and 29) were of discrimination parameter less than 0.35. The result showed that most of the items were of good discrimination parameters. The implication of the result is that most of the NECO SSSCE practical Physics test items were able to discriminate examinees with low Physics practical from those with high Physics practical ability.

Research Question Four: How many items of the NECO SSSCE Physics practical test fit the IRT model?
Toanswer this research question, the item fit indices of the items were extracted from the calibrated NECO SSSCE Physics test and the result is presented in Table 4.

Table 4: Model Fit of the NECO SSSCE Physics Practical

S/N	Item	S_X2	Df	P	S_X2/df	RMSEA	Remark
1	Q1_ai	35.057	24	0.068	1.46	0.027	Fit
2	Q1_aii	88.903	57	0.004	1.56	0.030	Fit
3	Q1_aiii	56.43	37	0.021	1.53	0.029	Fit
4	Q1_aiv	130.938	75	0.000	1.75	0.034	Fit
5	Q1_av	111.415	46	0.000	2.42	0.048	Fit
6	Q1_avi	110.817	58	0.000	1.91	0.038	Fit
7	Q1_avii	102.874	43	0.000	2.39	0.047	Fit
8	Q1_Gra	264.611	83	0.000	3.19	0.059	Misfit
9	Q1_Slo	147.862	40	0.000	3.70	0.065	Misfit

ISSN 2705-2559
E-ISSN 2705-2567

10	Q1_Eva	77.082	32	0.000	2.41	0.047	Fit
11	Q1_Prec	85.158	59	0.015	1.44	0.027	Fit
12	Q1_bi	70.251	50	0.031	1.41	0.025	Fit
13	Q1_bii	57.431	26	0.000	2.21	0.044	Fit
14	Q2_ai	114.091	80	0.007	1.43	0.026	Fit
15	Q2_aii	88.274	61	0.013	1.45	0.027	Fit
16	Q2_aiii	109.111	73	0.004	1.49	0.028	Fit
17	Q2_aiv	110.202	72	0.003	1.53	0.029	Fit
18	Q2_av	63.392	46	0.045	1.38	0.024	Fit
19	Q2_avi	86.225	49	0.001	1.76	0.035	Fit
20	Q2_Gra	111.518	76	0.005	1.47	0.027	Fit
21	Q2_Slo	114.977	49	0.000	2.35	0.046	Fit
22	Q2_Prec	75.981	55	0.032	1.38	0.025	Fit
23	Q2_bi	71.581	37	0.001	1.93	0.039	Fit
24	Q2_bii	57.61	41	0.044	1.41	0.025	Fit
25	Q3_ai	50.432	51	0.496	0.99	0.000	Fit
26	Q3_aii	89.077	50	0.001	1.78	0.035	Fit
27	Q3_aiii	49.1	44	0.276	1.12	0.014	Fit
28	Q3_aiv	105.593	89	0.111	1.19	0.017	Fit
29	Q3_av	75.83	51	0.014	1.49	0.028	Fit
30	Q3_avi	57.688	52	0.273	1.11	0.013	Fit
31	Q3_Gra	211.843	123	0.000	1.72	0.034	Fit
32	Q3_Slo	134.256	60	0.000	2.24	0.044	Fit
33	Q3_Inter	194.728	29	0.000	6.71	0.095	Misfit
34	Q3_Eva	54.82	27	0.001	2.03	0.040	Fit
35	Q3_Prec	133.396	65	0.000	2.05	0.041	Fit
36	Q3_bi	74.129	49	0.012	1.51	0.029	Fit
37	Q3_bii	104.599	67	0.002	1.56	0.030	Fit

Table 4 shows the item fit of the NECO Physics practical test to the generalized partial credit model that fitted the test data. On the table, the S-X2 represents the chi-square statistic, df is the degree of freedom, p-value is the measure of significance of the chi-square value, the S_X2/df is the ratio of degree of freedom to S_X2 and the remark is the qualitative description of the item fit. An item is considered fit when the p-value of S-X2 is greater than 0.05 and misfit or not fit when the p-value is less than 0.05. However, the result of S-X2 significance value are not always dependable. This is because it is affected by sample size. For accuracy the ratio of degree of freedom to S-X2 value and other measures such as root mean square error of approximation (RMSEA) have been suggested to overcome the weakness of the chi-square significance value (Kline, 2016). As a general rule, the ratio of degree of freedom to chi-square value less than 3:1 (Kline, 2016), values of RMSEA lower than or equal to 0.05 perfect but values equal to 0.08 are also acceptable (McDonald and Ho, 2002). As result of the criteria, the fitness of the items were judged based on the ratio of degree of freedom to chi-square value less than 3:1 and RMSEA less than or equal to 0.05. The table shows that out of the 37 tasks used in assessing the proficiency of students in Physics practical by NECO, 34 items fitted the generalized partial credit model used in the calibration

ISSN 2705-2559
E-ISSN 2705-2567

of the test, while only three tasks (8, 9 and 33) do not fit the IRT model adopted for the test calibration. The result showed that the test items fitted the generalized partial credit model. The implication of the result is that the information provided by IRT model on the estimates of the test were dependable and can be reliably used in the evaluation of the quality of the test.

Discussion

The study carried out item analysis of a senior secondary school certificate physics practical examination with the goal of establishing the psychometrics properties of the items. The study therefore estimated the items difficulty and discrimination indices as well as the fitness of the practical items into the IRT model. The difficulty parameters of polytomous test items of easy type are referred to as step difficulty. The step difficulties indicate the point on the ability metric at which the probabilities for adjacent categories is equal (De Ayala, 2009) or point of intersection of two adjacent categories on the ability scale. The average of these step difficulties is the overall difficulty of the items on a test (Paek and Cole, 2020). The categories represent the arrays of scores obtained on the items. In Table 2 there are six step difficulties (b1, b2, b3, b4, b5 and b6) for the items. While some items have only one step difficulty, some have two, others have three and so on up to the highest category, 6. With b1, b2, b3, b4, b5 and b6 representing the step difficulty for score 0 and 1 (i.e., the point of intersection of categories 0 and 1); score 1 and 2, score 2 and 3, score 3 and 4, score 4 and 5 and score 5 and 6 respectively. Items with increasing step difficulties within -3 to 3 are considered good.The findings indicated that all physics practical items required to judge the proficiency of students in physics practical were of inadequate level of difficulty.

The quality of item discrimination parameters of the tasks used in judging the performance of students in NECO Physics practical test was judged based on the criteria established by (Baker, 2001; Hasmy, 2014). According to the authors the following benchmark: Very high discrimination $a \geq 1.7$; High discrimination $1.35 \leq a < 1.7$; Moderator discrimination $0.65 \leq a < 1.35$; Low discrimination $0.35 \leq a < 0.65$; and very low discrimination, $a < 0.35$. Based on the benchmark, discrimination estimate less than 0.35 is considered poor. Findings showed that28of the items had discrimination parameter estimates that were greater than 0.35, nineitems were of discrimination parameter less than 0.35. The implication of the result is that most of the NECO SSSCE practical Physics test items were able to discriminates examinees with low Physics practical ability from those with high Physics practical ability.

The model that produced the best fit to the data is adjudged the model that fit the data. According to Oguoma, Metibemu and Okoye (2016); Finch and French (2015), prominent among the measures include Chi-square difference test and use of information indices. The item fit of the NECO Physics practical test was determined by the generalized partial credit model that fitted the test data. From the analysis it showed that the S- χ^2 represents the chi-square statistics, df is the degree of freedom, p-value is the measure of significance of the chi-square value, the S_X2/df is the ratio of degree of freedom to S_ χ^2 and the remark is the qualitative description of the item fit. An item is considered fit when the p-value of S- χ^2 is greater than 0.05 and misfit or not fit when the p-value is less than 0.05. However, the result of S- χ^2 significance value is not always dependable.

Conclusion

The study concluded that most of the tasks required to judge the proficiency of students in Physics practical were of inadequate difficulty level but with high discrimination power and that the information provided by IRT model on the estimates of the test were dependable and can be reliably used in the evaluation of the quality of Physics practical test items.

Recommendations

The following recommendations were made from the study:

1. Physics practical items required to judge the proficiency of students in physics practical were of inadequate level of difficulty

ISSN 2705-2559
E-ISSN 2705-2567

2. Most of the NECO SSSCE practical Physics test items were able to discriminates examinees with low Physics practical ability from those with high Physics practical ability.
3. Information provided by IRT model on the estimates of the test were dependable and can be reliably used in the evaluation of the quality of the test

References

Abrahams, I. (2011). Practical work in secondary science: A minds-on approach. London; New York: Continuum International Publishing Group.

Abrahams, I. (2011). *Practical work in secondary science: A minds-on approach.*London; New York: Continuum International Publishing Group.

Agrawal, A. T & Meson, N. (2010). *Relative efficiency of test scores equating methods in the comparison of students' continuous assessment measures.* An unpublished Ph.D. Thesis Submitted to the Department of Science Education University of Nigeria, Nssuka.

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In: Petrov, B.N. and Csaki, F., Eds., *International Symposium on Information Theory*, 267-281.

Ariyo, A. O. (2006). *School and student factors as determinants of students' achievement at the secondary school level in Oyo State, Nigeria*. Unpublished Doctoral Thesis, University of Ibadan, Nigeria.

Baker, F. (2001). The Basics of Item Response Theory. *Eric Clearing House on Assessments and Evaluation.* University of Maryland College. Park M.D.

Chukwunenye, J. N. (2011). *Effects of two modes of computer simulated experiment on senior secondary school students' achievement and interest in practical physics*. Pre field seminar Paper, Department of Teacher Education, University of Ibadan, Nigeria.

Chukwunenye, J. N. (2012). *Effects of two modes of computer simulated experiment on senior secondary school students' achievement and interest in practical physics*. Pre field seminar Paper, Department of Teacher Education, University of Ibadan, Nigeria.

Chris, C. & Vollmer, M. (2006). "*Learning Physics from the Experiments," workshop report inInformal Learning and Public Understanding of Physics,Selected Contributions of the 3rd International Girep Seminar 2005,* edited by G.Planinicand A. Mohoric (Univ. of Ljubljana, Slovenia, 2006), pp. 185–190

De Ayala, R. J. (2009). *The theory and practice of item response theory*. 1st ed. New York, NY: The Guilford Press. 144-200.

Enders, C. K., & Tofighi, D. (2008). The impact of misspecifying class-specific residual variances in growth mixture models. *Structural Equation Modeling, 15*(1), 75–95.

Finch, W. H., French, B. F. (2015). *Latent variable modeling with R.* New York, NY: Routledge.

Hasmy, A. (2014). Compare unidimensional& multidimensional Rasch model for test with multidimensional construct and items local dependence. *Journal of Education and Learning,* 8(3), 187-194.

Kline, T. J. (2005). Classical test theory assumptions, equations, limitations, and item analyses. *Psychological testing*. T.J, Kline. Calgary, Canada: Sage Publications, Chapter 5, Pp. 91-106.

Kuti, J. B. (2012). *Effect of split attention multimedia principle and cognitive load on senior secondary school students' retention in Quantum Physics in Ogun State, Nigeria.* Unpublished Ph.D. thesis, University of Ibadan, Nigeria.

McDonald, R. P., & Ho, M. H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7*(1), 64–82. https://doi.org/10.1037/1082-989X.7.1.64

Oguoma, R. O., Metibemu, C. C. & Okoye, M. A. (2016). "An Assessment of the dimensionality of 2014 West African secondary school examination mathematics objective test scores in Imo State, Nigeria," African.*Journal of Theory and Practice of EducationalAssessment*, vol. 4, 18-33.

Paek, I., & Cole, K. (2019). *Using R for item response theory model applications*. New York, NY: Routledge.

Zdenek, K., & Hana, P. (2008). About the Project Education at the Secondary School at Czech Republic. *Lat. Am. J. Phys. Educ., Vol. 2(3):*212-213.