

Analysis of Past West Africa Examination Council / Senior Secondary Certificate Examination Physics Essay Test Items Parameters: Application of IRT Polytomous Model

Ethe Nathaniel^{1*} Dennis Edougha² Godwin Odjegba³

1.Department of Educational Measurement and Evaluation, Delta State College of Education, Mosogar, Nigeria

2.Department of Educational Psychology, Delta State College of Education, Mosogar, Nigeria

3.Department of Agricultural Science Education, Delta State College of Education, Mosogar, Nigeria

Abstract

The study aimed at determining the quality of the WAEC SSCE Physics essay test items for 2015, 2016 and 2017, if they are acceptable when the item response theory (IRT) Polytomous model is applied. A descriptive survey research design was adopted and sampled 1000 senior secondary school students from the 27,615 students who registered for the 2018 senior school certificate physics examination in Delta State, Nigeria. Findings revealed that most of the May/June WAEC SSCE Physics essay test items for year 2015, 2016 and 2017 are not acceptable when the item response theory (IRT) Polytomous model is applied among as presented in this study. It was therefore recommended that test developers should always ensure proper test items validation before use.

Keywords: Parameters, IRT, Polytomous Model, Test Items, WAEC, Physics, Essay Test.

DOI: 10.7176/JEP/10-17-04

Publication date: June 30th 2019

1. Introduction

West African Examination Council (WAEC) saddled with the responsibility of conducting examination related issues, in particular the senior secondary school examination (SSCE) in her curriculum/syllabus identified physics as one of the core science subjects which is the study of matter, energy and their interactions. Technological development encompasses the application of the principles of physics and it describes how the natural world works through applied Mathematical formulas. The Senior School Certificate Examination papers are made up of two parts, objectives and essay test items. While the objective test items constitute twenty-five percent (25%), the essay test items constitute seventy-five percent (75%) of the total marks earned by the examinees. There is an outcry in the West African Examination Council report on poor performance especially in the science related subjects. In view of this, researchers have been making effort to proffer solution to the problem. Some of the studies focused on the teachers as a factor, others are on students study habit, learning environment and so on. Few or no studies have been carried out on the nature and quality of the test items administered on the students. Where such studies are carried out they dwelled on the objective test items which constitute 25% of the examination. Two major theories have been developed and available for psychometricians in determining the quality and characteristics (items parameters) of test items that are administered to examinees. They are classical test theory (CTT) and item response theory (IRT). While CTT has dominated and remained a leading framework for analyzing and developing standardized test items, it lacks the ability for analyzing essay test items. To this end, a modern theory, known as the Item Response Theory (IRT) was developed to correct the limitations of the CTT. Item Response Theory has two categories of model – dichotomous and polytomous model. While the dichotomous models deals with objective test items, the polytomous model is use for essay test items. The Polytomous model was adopted for this study as the study is aimed at dealing with essay test items.

Item Response Theory (IRT) has become a contemporary issue in psychological and educational assessment in recent times. IRT as described “New psychometric term” is popular in psychological testing, test manuals and journals and very soon psychologists who do not know about item response theory will be a minority and very much at a disadvantage in their research (Hambleton, 2004, Navas, 2001) reported that in Spain today there are several books as general testing practices in relation to IRT. Han & Hambleton (2007) stated that item response theory is a popular and valuable framework for modelling educational and psychological tests data, due to its attractive properties such as the invariance of item and examinee parameter values being reported on a common scale. The theory is built on three basic assumptions, first is that the items are measuring a single latent variable ranging from - to + i.e. unidimensional trait. This assumption refers to the fact that is assumed to be a unidimensional random variable and not a multidimensional random vector. The trait is measurable on a scale i.e. the mere existence of a test assumes this, typically set to a standard scale with a mean of 0.0 and a standard deviation of 1.0. The second assumption is “local independence” which means that items are not related except for the fact that they measure the same trait which is equivalent to the assumption of unidimensionality. Thirdly, that the item response function gives the probability that a person with a given ability level will answer correctly

(Hambleton, Robin & Xing, 2000; Bechger, Gunter, Huub & Beguin, 2003; Henrikson, Sundstrom & Wilberg, 2004 and Omorogiuwa, 2009). Interest in IRT is growing rapidly due to the expended uses of educational and psychological assessments and the corresponding needs to have test items that are valid and reliable. Item response theory which is also known as latent trait theory, strong score theory or modern mental test theory is conceptualized as a paradigm for the design, analysis and scoring of tests, questionnaires and other similar instruments measuring abilities, attitudes or other variable. Mobbing (2004) reported that until recent time, IRT was referred to as item characteristic curve theory. The theory is based on the application of related mathematical models to testing data and as such regarded as superior to classical test theory, those who are involve in high stakes tests development see it as a preferred method such as the Graduate Record Examination (GRE) and Graduate Management Admission Test (GMAT). The name item response theory is due to its focus on test item compared to classical test theory that focus on test-level, it actually models the response of an examinee of a given ability to each item in the test.

The purpose of item response theory is to overcome the shortcomings of classical test theory by providing a reporting scale on which examinee ability i.e. the construct measured by the test, is independent of the particular choice of test items that are administered (Hambleton, 2004). Literature, pointed out that IRT began between 1940s and 1950s as a goal of psychometrician and became reality 30 years later (Lord, 1980 and Hambleton, 2004). The pioneering work of IRT as a theory occurred during the 1950 and 1960 by three Educational Testing Service Psychometrician, they are Frederie M. Lord, a Danish Mathematician Georg Rasch and Austrian sociologist Paul Lazarsfeld, they both pursued parallel research independently. Though, IRT is regarded as a modern psychometric theory, the concept and methodology has been developed for over three quarters of a century. Reeve (2000) reported that Thurstone L.L. laid down the conceptual foundation in his paper, entitled “A Method of Scaling Psychological and Educational Tests” in the year 1925. He later dropped the idea in measurement to pursue the development of multiple factor analysis. In this 21st century, IRT has a wide range of application such as item analysis, items selection, test publishing, reporting large scale testing, test development, test scores equating, the study of item bias, and computerized adaptive testing.

2. Problem

Poor academic performance had been a perennial problem in West African Examination Council in relation to science subjects. Could this poor performance be a function of the item parameters? The annual releases of Senior Secondary School Certificate Examination (SSCE) results have been justifying the problematic nature and generalization of the poor performance in different science school subjects. Considering the importance of physics for the technological advancement of societal needs, the continued poor performance calls for worry. Standardized testing has been the most common methodology, yet the validity and credibility of the expanded range of contemporary assessment techniques have been called into questions. In view of this, test developers are basically concerned about the quality of test items and how examinees respond to them when constructing/selecting test items. As earlier noted, there are two theory available for psychometrician for validating, analyzing and standardizing test items. Umobong (2004), Omorogiuwa (2009) and Ethe (2012) noted that examination bodies in Nigeria based their items analysis on classical test theory (CTT). A standardized examination like this, is supposed to possess acceptable item parameters, as it is true that test items parameters is a function in students’ academic achievement. Are the West African Examination Council Physics essay test items parameters acceptable when the IRT polytomous model is applied? It is against this background, this study “Analysis of Past West African Examination Council (WAEC), Senior School Certificate Examination (SSCE) Physics Essay Test items parameters: Application of IRT polytomous model”.

2.1 Objective of the Study

The aim of this study is to analyse the quality of the WAEC SSCE Physics essay test items. Specifically, is to determine if the difficulty and discriminating parameters of the test items for the year under review are acceptable/satisfactory using the item response theory polytomous model.).

2.2 Research Questions

1. How many items of the May/June WAEC SSCE Physics Test Items are acceptable for year 2015, 2016 and 2017.
2. Are there differences in the number of acceptable items of the May/June WAEC SSCE Physics Essay Test Items are acceptable for year 2015, 2016 and 2017?
3. Are there differences in the May/June WAEC SSCE Physics Essay Test Items Difficulty Parameters between year 2015, 2016 and 2017?
4. Are there differences in the May/June WAEC SSCE Physics Essay Test Items Discrimination Parameters between year 2015, 2016 and 2017?

2.3. Hypotheses

H01: There is no significant difference in the number of acceptable items of the May/June WAEC SSCE Physics Essay Test Items for year 2015, 2016 and 2017.

H02: There is no significant difference in the May/June WAEC SSCE Physics Essay Test Items Difficulty Parameters between year 2015, 2016 and 2017.

H03: There is no significant difference in the May/June WAEC SSCE Physics Essay Test Items Discriminating Parameters between year 2015, 2016 and 2017.

3. Methods

Descriptive survey research design was adopted for the study. A descriptive survey study seeks or uses the sample data of an investigation to document, describe and explain what is in existence or non-existence or the present status of a phenomenon being investigated (Ali, 2006). Descriptive survey studies have been known to be very useful as a basis for collecting and documenting information for institutional policy formulation or systems-wide improvement and management decision support system.

The population of the study comprised the 27,615 Senior Secondary School (SSS) III who will register to write 2018 SSCE Physics subject in Delta State. A total of 1000 physics students were sampled from public and private senior secondary schools across five (5) local government area in the Delta State. A stratified and simple random sampling procedure was used to select schools and students for the study.

The instrument used for the study is the past May/June WAEC SSCE Physics essay test items (questions) papers for year 2015, 2016 and 2017 which was adopted and reproduced and administered to the sampled students. Due to the nature of the examination as a perceived standardized test, the researchers assume it to be a valid instrument. However, the instrument was first subjected to Pearson Product Moment Correlation Coefficient method of reliability estimate to establish the coefficient. To ensure a proper and reliable data, the instrument was reproduced and administered to the sampled students with the permission of the school principals of the various selected schools. This was done using train research assistance (the subject teachers). The students were informed about the examination three weeks before the instrument was administered to them.

Obtain responses from the students was scored using the WAEC scoring guide. Results were first subjected to xcalibre software for test items calibration where acceptable items were automatically selected and their parameters generated for further statistical analysis. Frequency counts, percentages, ANOVA, chi-square (χ^2) test statistical analysis method were used to answer research questions as well as to test formulated hypotheses.

4. Results

4.1 Research Question One

How many items of the May/June WAEC SSCE Physics essay test items are acceptable for year 2015, 2016 and 2017.

Table I: Number/percentages of the May/June WAEC SSCE Physics Essay Test Items that are acceptable for year 2015, 2016 and 2017

Year	No. of Items/Percentages
2015	7 (58.33%)
2016	8 (66.67%)
2017	5 (41.67%)

Results in Table I shows the Number and percentages of the May/June WAEC SSCE Physics Essay Test Items that are acceptable for year 2015, 2016 and 2017 when the item response theory (IRT) polytomous model is applied. Results revealed that seven (7), eight (8) and five (5) items are acceptable. This representing 58.33%, 66.67% and 41.67% respectively of the May/June WAEC SSCE Physics Essay Test Items

4.2 Hypotheses Testing

A total of three hypotheses were formulated and tested as presented below:

H₀₁: There is no significant difference in the number of acceptable items of the May/June WAEC SSCE Physics essay test items for year 2015, 2016 and 2017.

Table II: Chi-square (χ^2) test on the Number of Acceptable Items of the May/June WAEC SSCE Physics Essay Test Items for Year 2015, 2016 and 2017

Years	Observed N	Expected N	df	chi-square	P-value	Decision
2015	7	6.7	2	0.700	0.705	NS
2016	8	6.7				
2017	5	6.7				

$P > 0.05$

Result in Table II revealed the chi-square (χ^2) test value of 0.700, testing at 0.05 alpha (α) level of significance with a degree of freedom (df) of 2, P-value of 0.705. Since the P-value (0.705) exceeded the alpha

(0.05) level of significance, hypothesis one therefore not rejected at $P > 0.05$, that there is no significance differences in the number of acceptable items of the May/June WAEC SSCE Physics Essay Test items for year 2015, 2016 and 2017.

H₀₂: There is no significant difference in the May/June WAEC SSCE Physics essay test items difficulty parameters between year 2015, 2016 and 2017.

Table III: Analysis of Variance of the May/June WAEC SSCE Physics Essay Test Items Difficulty Parameters between 2015, 2016 and 2017

Source	Sum of Square	df	Mean Square	F	P-value	Decision
Between Groups	0.778	2	0.389			
Within Groups	10.557	12	0.621	0.627	0.546	NS
		19				
Total	11.335	19				

$P > 0.05$

Result in Table 3 revealed the analysis of variance (ANOVA) value of 0.627, testing at 0.05 alpha (α) level of significance with degree of freedom (2,17), P-value of 0.546. Since the P-value (0.546) exceed the alpha (0.05) hypothesis two is not rejected at $P > 0.05$, that there is no significant difference in the May/June WAEC SSCE Physics Essay Test items difficulty parameters between year 2015, 2016 and 2017.

H₀₃: There is no significant difference in the May/June WAEC SSCE Physics essay test items discriminating parameters between year 2015, 2016 and 2017.

Table IV: Analysis of variance of the May/June WAEC SSCE Physics Essay Test Items Discriminating Parameters between year 2015, 2016 and 2017

Source	Sum of Square	df	Mean Square	F	P-value	Decision
Between Groups	1.195	2	0.597			
Within Groups	14.111	17	0.830	0.720	0.501	NS
Total	15.306	19				

$P > 0.05$

Result in Table 4 revealed the analysis of variance (ANOVA) value of 0.720, testing at 0.05 alpha (α) level of significance with degree of freedom (2,17), P-value of 0.501. Since the P-value (0.501) exceed the alpha (0.05) hypothesis three is not reject at $P > 0.05$, that there is no significant difference in the May/June WAEC SSCE Physics Essay Test items discriminating parameters between 2015, 2016 and 2017.

4.3 Discussion of Findings

This study used Item Response Theory (IRT) Polytomous model statistics to generate the May/June West Africa Examination Council (WAEC) Senior School Certificate Examination (SSCE) Physics Essay Test Items discriminating and difficulty parameters as well as items that are acceptable. Results presented in Table one revealed the number of items that were accepted at good items of the 12 Physics essay test items presented to the students for year 2015, 2016 and 2017 respectively. It was observed that out of the 12 Physics essay test items, 7, 8 and 5 items are acceptable this represent 58.33%, 66.67% and 41.67% for year 2015, 2016 and 2017 respectively; an important examination like this as standardized test items, it would have been expected a near 100% of acceptable test items. However, this is not surprising as the findings of this study agreed with the finding of Osunde and Ethe, 2003 and Ethe, 2012 who reported that SSCE Mathematics multiple choice test items presented for the students were of high difficulty level. The subsequent report by this examination body of poor performance in science related subject also confirmed the findings of this study.

Results of the study further show that there is no significant difference in the May/June WAEC SSCE Physics essay test items difficulty and discriminating parameters between 2015, 2016 and 2017. The analysis of variance for both the difficulty and discriminating parameters led to the acceptance of hypotheses two and three at 0.05 alpha () level of significance. This finding confirmed that truly the test items are difficulty for the students as well as not discriminating well which support the results presented in Table one, that most of the May/June WAEC SSCE Physics essay test items for 2015, 2016 and 2017 are not acceptable. Though, this could imply either too difficult items or cheap items for the students.

5. Conclusion

Arising from the findings of this study the conclusion is that:

- (i) Most of the May/June WAEC SSCE Physics essay test items are not acceptable.
- (ii) There is no significant difference in the number of acceptable items of the May/June WAEC SSCE Physics essay test items for year 2015, 2016 and 2017.
- (iii) There is no significant difference in the May/June WAEC SSCE Physics essay test items difficulty

parameters for year 2015, 2016 and 2017.

- (iv) There is no significant difference in the May/June WAEC SSCE Physics essay test items discriminating parameters for year 2015, 2016 and 2017.

6. Recommendation

From the findings and conclusion of the study, the researchers therefore recommended that test developers should always ensure a proper validation of test items before use.

7. Acknowledgement

We wish to acknowledge TETFUND Nigeria for their sponsorship and support towards the completion of this project through the IBR grant 2015/16 for COEM, Nigeria

References

- Ali, A. (2006). *Conducting Research in Education and the Social Sciences*. Enugu: Tashiwa Networks Ltd.
- Bechger, T., Gunter, M., Huub, H. & Beguin, A. (2003). Using Classical Test Theory in Combination with Item Response Theory. *Applied Psychological Measurement*, 27(5). 319-334.
- Ethe, N. (2012). *Comparative Analysis of WAEC and NECO SSCE Mathematics Multiple Choice Test Items: Application of Item Response Theory*. Ph.D. Thesis submitted to School of Postgraduate Studies, University of Benin, Nigeria.
- Hambleton, R.K. (2004). Theory, Methods and Practices in Testing for the 21st Century. *Psicothema*, 16(4), 696-701. Retrieved from <http://www.psicothema.com>
- Hambleton, R.K., Robin, F. & Xing, D. (2000). Item Response Models for the Analysis of Educational and Psychological Test Data. In H. Tinsley & S. Brown (eds.). *Handbook of Applied Multivariate Statistics and Modelling*. San Diego, CA: Academic Press.
- Han, K.T. & Hambleton, R.K. (2007). *User's Manual for Wingen: Window Software that Generate IRT Model Parameters and Item Responses*. Amherst, MA: University of Massachusetts, Centre for Educational Assessment.
- Henrikson, W., Sunstrom, A. & Wiberg, M. (2004). *The Swedish Driving-License Test: A Summary of Studies from the Department of Educational Measurement, Umea University, Umea*.
- Lord, F.M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. New Jersey: Lawrence Erlbaum Associates Publishers.
- Navas, M.J. (2001). *Psychological Measurement Instrument I: Construction and Administration*. In M.J. Navas (Ed.) *Methods, Design and Research Techniques in Psychology*. Madrid: UNED.
- Omorogiuwa, K.O. (2009). *An Empirical Comparison of the Classical Test Theory and Item Response Theory in the Selection of Physics Achievement Test Items*. Ph.D. Thesis submitted to School of Postgraduate Studies, University of Benin, Nigeria.
- Umobong, M.E. (2004). *Item Response Theory: Introducing Objectivity into Educational Measurement*. In Afemikhe, O.A. and Adewale, J.G. (eds.) *Issues in Educational: Measurement and Evaluation in Nigeria, Nigeria, Educational Research and Study Group, Institute of Education, University of Ibadan*.

Nathaniel Ethe (Chief Lecturer), holds a Ph.D. in Educational Measurement and Evaluation from University of Benin, Benin City, Nigeria in 2012. He equally obtained Masters Degree in Education (M.Ed) in Educational Measurement and Evaluation and B.Sc. (Ed) in Education/Mathematics. His teaching and research career begins at the Delta State College of Education, Mosogar in 2004. He has taught various courses in the field of Education. Dr. Ethe has carried out several research studies and publication in learned journals (Local and International), Chapters in Books etc. He was an Acting Dean, School of Education, Head of Department (severally), Director, Professional Diploma in Education (PDE), Director, Quality Assurance Unit, Head and member of different committee and currently Director, University of Benin Affiliate Degree Programme, in the Delta State College of Education, Mosogar, Delta State. He was appointed to monitor the nationwide assessment of the Federal Teachers Scheme in 2010. Dr. Ethe is a member of different professional bodies.

Mr. Dennis E. Edougha, holds a Master degree in Educational Psychology (2001) of the Ludwig-Maximilians University (LMU) Munich, Germany. and a B.Sc. (Hons) degree in General and Applied Psychology (1994) of the University of Jos, Nigeria. He has taught various courses in the field of Psychology. He is a Senior Lecturer and Head of Counselling Unit and the current Head of Educational Psychology Department, Delta State College of Education, Mosogar. He is a member of Nigerian Society for Educational Psychologists (NISEP).

Mr. Godwin Odjegba, holds a first degree in Agricultural Education from Delta State University, Abraka in 1998. He is currently doing his Master in Instructional Technology in University of Benin, Benin City, Nigeria. He is the pioneer HOD of the Agricultural Education Department. He has taught several courses in the Department. He is a member of Teaching Practice Committee, also the College/Community Liaison Officer. Mr. Godwin

Odjegba, has written several articles and published in several journals. He is a member of several professional bodies, including Nigeria Vocational Association (NVA), Teachers Registration Council of Nigeria (TRCN).

Table 1: May/June WAEC SSCE Physics essay test calibrated items Parameters for year 2015

S/M	Item ID	a	B
1	ITEM02	1.727	-0.118
2	ITEM04	1.533	-0.879
3	ITEM05	2.498	-0.632
4	ITEM06	3.518	0.893
5	ITEM07	2.951	-0.871
6	ITEM09	2.872	-0.863
7	ITEM11	2.951	-0.871

Table 2: May/June WAEC SSCE Physics essay test calibrated items Parameters for year 2016

S/N	Item ID	a	b
1	ITEM02	2.928	-0.728
2	ITEM03	1.465	-0.641
3	ITEM04	3.980	0.909
4	ITEM06	3.212	0.973
5	ITEM09	0.625	0.954
6	ITEM10	3.229	0.970
7	ITEM11	1.150	-0.980
8	ITEM06	3.418	0.823

Table 3: May/June WAEC SSCE Physics essay test calibrated items Parameters for year 2017

S/N	Item ID	a	b
1	ITEM01	2.117	0.394
2	ITEM03	1.197	-0.152
3	ITEM04	2.198	0.399
4	ITEM07	2.117	0.400
5	ITEM10	3.117	-0.161