

Full Length Research Paper

Comparative analysis of psychometric properties of mathematics items constructed by Waec and Neco in Nigeria using item response theory approach

Aborisade, Olatunbosun James^{1*} and Fajobi, Olutoyin Olufunke²

¹Institute of Education, Faculty of Education, Ekiti State University, Ado-Ekiti, Nigeria.

²Department of Educational Foundation, Faculty of Education, Federal University, Oye-Ekiti, Nigeria.

Received 7 July, 2019; Accepted 5 December, 2019

West Africa Examination Council (WAEC) and National Examination Council (NECO) are the two major examination bodies saddled with the responsibility of awarding Senior Secondary School Certificate in Nigeria. This study examined the comparability of the psychometric properties of the items constructed by the two examination bodies using Item Response Theory (IRT) approach. Three parameters (difficulty, discriminating and distractor Indices) logistic model was adopted for the study. The study employed descriptive research design of the survey type. The population for the study comprised all Senior Secondary School Students who enrolled for 2019 WAEC and NECO examinations in South West, Nigeria. The sample for the study consisted of 1,200 Senior Secondary School Students selected using multistage sampling procedure. The instruments for the study were objective items constructed by the examination bodies. The findings of the study showed that the difficulty and guessing indices of the mathematics items constructed by the two examination bodies are comparable while the discriminating powers not comparable. It is, therefore, recommended that certificates issued by WAEC and NECO could be used for same purposes without any discrimination since the items constructed by the examination bodies are of comparable standard in terms of their qualities.

Key words: Difficulty levels, discriminating powers, guessing indices, Item Response Theory (IRT), West Africa Examination Council (WAEC), National Examination Council (NECO).

INTRODUCTION

West Africa Examination Council (WAEC) and National Examination Council (NECO) are the two major examination bodies in Nigeria that have mandate to award Senior School Certificate. WAEC was established in the year 1952 to conduct examinations for Anglophoric Countries of West Africa (Ghana, Sierra Leone, Liberia, Gambia and Nigeria) while NECO were established in

Nigeria in year 2000. The two examination bodies conduct parallel Senior School Certificate Examinations in various subjects, such as Mathematics, English Language ect cetera. However, at inception, there was criticism against NECO examinations, some say its questions are too tough than those of WAEC (Anigbo, 2018; Awogbemi et al., 2015). Since WAEC and NECO

*Corresponding author. E-mail:Olatunbosun.aborisade@eksu.edu.ng.

have similar syllabi and award certificate for same purposes, their examinations are expected to be of comparable standard in terms of the psychometric properties of the items constructed by the examination bodies. The set of students that took these examinations are expected to be of comparable ability because the students that took the examinations were exposed to the same course contents and curriculum within the same time frame. Therefore, they are expected to have equal possibility of success in NECO and WAEC examinations but this is not so in reality which shows that it is probable that some technical weaknesses exist in the test constructed by the examination bodies.

To this end, the study did a comparative analysis of the psychometric properties of items constructed by the examination bodies. Psychometric characteristics of examinations refer to certain attributes inherent in tests upon which an assessment of candidates is based. These characteristics include the difficulty index, the discriminating index, distractor index, validity and reliability indices of the test items. Items that are correctly answered by students do not convey any message about individual differences in performance. Item difficulty tries to estimate how easy or difficult is the item, the higher the value, the easier the item or lower the difficulty.

The ability of an item to discriminate between higher ability examinees and lower ability examinees is known as item discrimination. There are several methods being used to assess item discrimination. These include, finding the difference in the proportion of high and low achieving students who score the item correctly. Therefore, it is necessary to know the psychometric properties of examination items constructed by these examination bodies to ascertain its effectiveness.

Oshkosh (2002) stated that Item Analysis is probably the most important tool to increase test effectiveness, it is a scientific way of improving the quality of tests, and test items in an item bank. An item analysis provides three kinds of important information about the quality of test items. Item difficulty is a measure of whether an item was too easy or too difficult. Item discrimination is a measure of whether an item discriminated between candidates who knew the test well and candidates who did not. Distractor Index measures the effectiveness of alternatives, that is, to determine whether distractors (incorrect but plausible options) tend to be chosen by the less able examinees and not by the more able examinees. Therefore, since test item affects the vital psychometric properties of measuring quality of examinations, examination bodies are expected to construct test items in such a manner that items are free from writing errors to ensure that the items measures what it exactly want to measure without any ambiguity. In view of this there is the need to embark on this study in order to examine the comparability of the items constructed by WAEC and NECO since they award certificates use for similar purposes.

Statement of the problem

Knowledge of mathematics is fundamental to the success of several subjects in school curriculum and it is, also indispensable for national technological growth. In a developing country such as Nigeria, performance of student in Mathematics which is bedrock of nation's technological advancement should be given complete attention. The performance of students in WAEC and NECO are not comparable for instance, according to examiners reports, in year 2018 NECO Mathematics examination 82.32% pass at credit level while 49.98% of the students who sat for WAEC Mathematics examinations in the year passed at credit level. Since the examinations were attempted by same set of students taught by the same teacher, and WAEC and NECO have similar syllabi, and award certificates for same purposes their items are expected to be of comparable standard with comparable performance. It is probable that some technical weaknesses exist in the tests developed and administered by the WAEC and NECO. Therefore, it is necessary to examine and compare the quality of the items constructed by the examination bodies.

Purpose of the study

The purpose of the study was to examine the comparability of the psychometric properties (difficulty levels, discriminating powers and guessing indices) of the Mathematics examination items constructed by WAEC and NECO in Nigeria using item Response Theory (IRT) approach.

Theoretical framework

The theoretical framework for the study focuses on 3-parameter Logistic model Item Response theory (IRT) developed by Lord, (1980). The model has basic assumptions of un-dimensionality and Local Independence. In the 3-parameter Logistic model, the probability of a correct response to a dichotomous item. Usually, a multiple –choice item is presented mathematically as follows:

$$P_i(\theta) = C_i + \frac{1 - C_i}{1 + e^{-D a_i(\theta - b_i)}}$$

Where:

θ : is the test taker's ability

a_i : is the item discrimination index

b_i : is the difficulty parameter

C_i : is the guessing index

e : is the base of natural logarithm and is approximately equal to 2.714

D : is the arbitrary constant (normally $D = 1.7$)

IRT which is also known as latent response theory is the probability of answering an item correctly or of attaining a particular response level in relation to individual ability and characteristic of the item. The goal of IRT is to predict the probability at which a testee of a given ability level responds to an item correctly. In IRT ability level is measured on a transformable scale having a mid-point of zero, a unit measure of one with the theoretical range of ability from negative infinity to positive infinity, however, practical consideration usually limit the range of values from -3 to +3 (Hambleton et al., 1991).

According to Zaman et al. (2008). The ability range in IRT estimates is between $-\infty$ to $+\infty$ theoretical but typically they range from + 3.0 for examples with high abilities on the test to -3.0 for examples with low abilities. The difficulty estimates in IRT for items range from +3 to -3 the item with difficulty level +3 and -3 are labelled as "very difficult" and "Very easy" respectively. There are three IRT model for test items that are dichotomously scored known as three, two and one - parameter IRT models to describe the psychometric properties of an item. The distinction among the model is the number of parameter used to describe the items the parameter are a-parameter (discriminating power), b-parameter (difficulty level) and c-parameter (guessing factor). The value of item difficulty denoted by b-parameter is a location parameter that indicate the position of the item characteristic curve in relation to the ability that is required for a testee to have 50% chance of getting the item right the item discrimination denoted by a-parameter provides information on how well an item separate testee with high and low ability level while guessing factor denoted by c-parameter indicates the ability level at which testee guess answer correctly that is the effect of guessing on the probability of a correct response. The values of these parameter indicates the ability level at which they occur which practically ranges from -3 to +3. IRT provides a framework for evaluating how well individual item in a test or examination function. IRT enables the psychometricians to develop and design examination items, maintain item banks and equate the difficulties of items for successive version of examination which allow comparison between result overtime.

According to Yu (2008), IRT address the weakness of Classical Test Theory (CTT). CTT does not provide information about how examinees at different ability level perform on the item. IRT is a necessary tool which has to be at any testing centers for a valid instrument (Tshering, 2006). According to Adedoyin (2010), for more objective educational measurement, IRT theoretical frame work should be incorporated by examination bodies in Africa for the construction of examination items.

Some authors who have worked on psychometric properties of examination items constructed by WAEC and NECO have identified difference in psychometric properties of the items constructed by the examination bodies. Olutola (2015) in his study discovered that WAEC

multiple – choice Biology examination have more difficult items than NECO multiple choice Biology examination and that the higher mean difficulty index discovered in WAEC may be cause by the number of options in WAEC multiple choice Biology examination. As the four options formats in WAEC have a higher difficulty than five options format in NECO. Olatunji (2007) reported that four options for WAEC multiple choice tests have better discriminating indices than NECO multiple choice test in Economics.

Research question

How comparable are the psychometric properties of Mathematics items constructed by WAEC and NECO?

Research hypotheses

The following research hypotheses were postulated for the study at 0.05 level of significance:

- (1) There is no significant difference between the difficulty indices of the mathematics examination items constructed by WAEC and NECO
- (2) There is no significant difference between the discriminating indices of the Mathematics items constructed by WAEC and NECO
- (3) There is no significant difference between the guessing indices of the Mathematics items constructed by WAEC and NECO.

METHODOLOGY

The study employed a descriptive research design of survey type. The population of the study comprised all public senior Secondary School Students in the final year who enrolled for 2019 WAEC and NECO mathematics examinations. A sample of 1,200 students were selected for the study using multistage random sampling procedure. The instruments used for the study consisted of mathematics objective test items constructed by WAEC (50 items) and NECO (60 items) for year 2017 examinations. The instruments were administered on the 1,200 sampled students in 24 public secondary schools randomly selected for the study under similar conditions as given by the examination bodies. The data collected were analysed using Bilog MG software statistical analysis to generate item difficulty levels, discriminating powers and guessing indices of the examination items. Hypotheses formulated were analysed using student's independent t-test.

Three parameter IRT was adopted in this study because the study examined all the three parameters. That is, difficulty levels discriminating factor and guessing index of each item. The three IRT model are based on the logistics (cumulative) distribution function. The logistics equations when graphed produced plot that are called Items Characteristics Curves (ICC) ability levels the X - axis while the probability of an examinee correctly answering the equation is denoted $p(\theta)$ on the Y - axis which produced values for each item. The linchpin of IRT is the item characteristics curve produced for each item which generated difficulty levels,

Table 1. Descriptive statistics of difficulty parameter for WAEC and NECO.

Type of examination	N	Mean (X)	Standard deviation	Maximum value	Minimum value
WAEC difficulty	50	5.11	3.47	10.42	0.88
NECO difficulty	60	6.07	3.51	15.61	-0.01

Table 2. Descriptive statistics of Discriminating parameter for WAEC and NECO.

Type of examination	N	Mean (X)	Standard deviation	Maximum value	Minimum value
WAEC discrimination	50	1.11	1.76	6.62	0.06
NECO discrimination	60	0.74	1.05	3.54	0.08

Table 3. Descriptive statistic of guessing parameter for WAEC and NECO.

	N	Mean (X)	Standard deviation	maximum Value	Minimum value
WAEC guessing	50	0.17	0.12	0.30	0.31
NECO guessing	60	0.16	0.09	0.05	0.00

discriminating power and guessing index for each item.

The two basic assumptions of IRT are unidimensionality of latent ability and local independence. Unidimensionality implies that items in a test must be developed to measure one and only one area of ability knowledge. This assumption is sometimes empirically assessed by investigating whether or not a dominant factor exists among all items of the test (Oloda, 2017). Local independence states that the probability of an examinee answering a test item correctly is not affected for better or for worse by his/her performance in any other item will not give a clue to the knowledge of another. Unidimensionality is the most important assumption common for all IRT models. Consequently, IRT model is adopted in this study to ensure the unidimensionality of each item constructed by the examination bodies. Validity of a test is strengthened by ascertaining the unidimensionality of its items.

Test for unidimensionality and local independence

The method used to ascertain unidimensionality for the items used for the study was confirmatory factor analysis to determine whether or not a dormant factor existed among all the items used in the study. The local independence of the items were assessed using tetra – chronic correlation between items with exactly the same ability. The items were found to meet the conditions for unidimensionality and Local Independence.

RESULTS

Research question

How comparable are the difficulty levels, discriminating powers and distractor indices in the mathematics

examination items constructed by WAEC and NECO?

To assess comparability of item parameters of WAEC and NECO mathematics test items, descriptive statistics of their indices were established using Bilog MG software statistical analysis for NECO and WAEC Examination items. Tables 1 to 3 present the descriptive statistics of item parameters of WAEC and NECO test items. It can be observed from Table 1 that overall difficulty mean and standard deviation for WAEC test items was 5.11 (SD = 3.47), and 6.07 (SD = 3.51) for NECO test items. It can be observed from Table 2 that overall discrimination indices mean and standard deviation for WAEC test items was 1.11 (SD = 1.76), and = 0.74 (SD = 1.05) for NECO test items. It can be observed from Table 3 that overall guessing mean and standard deviation for WAEC test items was 0.17 (SD = 0.12), and 0.16 (SD = 0.09) for NECO test items.

Testing of hypotheses

Ho₁: There is no significant difference between the difficulty index of the examination items constructed by WAEC and NECO. Table 4 show that the mean difference was -0.96 but NECO test items was a bit difficulty compare to WAEC test. Independent samples t-test statistics further showed that the mean difference was not statistically significant (t = 1.44, df = 108, p = 0.26). This result further reveals that on the overall, the difficulty index of WAEC and the NECO mathematics test

Table 4. Independent sample t-test of difficult parameters for WACE and NECO.

		Levene's test for equality for variances			t-Test for equality of means		
		F	Sig	t	Df	Sig (2- tailed)	Mean differences
Difficulty	Equal variance assumed	1.283	0.260	-1.435	108	0.154	-0.959
	Equal variances no assumed			-1.436	104.814	0.154	-0.959

Table 5. Independent sample t-test of discriminating parameters for WACE and NECO.

		Levene's Test for equality for variances				Mean difference
		F	Sig	t	Df	
Discrimination	Equal variance assumed	9.864	0.002	1.351	108	0.366
	Equal variances no assumed			-1.436	76.826	0.363

Table 6. Independent sample t-test of guessing parameters for WACE and NECO.

		Levene's test for equality for variances				Mean difference
		F	Sig	t	Df	
Guessing	Equal variance assumed	5.474	0.021	0.224	108.000	0.005
	Equal variances no assumed			0.220	95.724	0.005

items are comparable.

Ho₂: There is no significant difference between the discriminating power of the examination items constructed by WAEC and NECO. Table 5 shows that the mean difference was 0.36, which shows that WAEC tests items distinguish very well between testees with low and high ability compare to NECO items. Independent sample t-test of discriminating indices show that the mean difference was statistically significant ($t=1.351$ $df = 108$, $P = 0.002$). The implication of these results was that on the whole there was the difference in the discriminating power of WAEC and that of NECO mathematics test items.

Ho₃: There is no significant difference between guessing index of the examination items constructed by WAEC and NECO. Table 6 shows that the mean difference was 0.005 which shows that the guessing indices of WAEC and NECO items are comparable. Independent sample t-test statistics showed that the mean difference was not statistically significant ($t = 0.22$, $df = 108$, $p = 0.82$). This result further reveals that on the overall, the guessing index of WAEC and that of NECO mathematics test items are comparable.

DISCUSSION

The result of the study shows that there were no

significant difference between WAEC and NECO item parameters except for discrimination index that proved otherwise. Despite this, the two examining bodies are not inferior to one another since their items are of equal quality, and certificate issued by them can be on the same metric scale. More importantly, results from the hypotheses suggests that there was no significant difference between the difficulty indices of WAEC and NECO test items. It implies that their test consist of sets of items with comparable difficulty which could be used to examine the testees. This is against the public outcry that NECO items are more difficult than WAEC. Thus, the two public examining bodies produced test items with comparable difficulty and one is not superior to other. This findings lay credence to the findings of studies carried out by Kolawole (2007); Alfred (2013); Metibemu (2016) and Ogbebor (2017) which says that there was no significant difference between the difficulty level of WAEC and NECO multiple-choice items in mathematics. However, the study disagrees with the submission of Adewale (2008); Bandele and Adewale (2013) and Thomas et al. (2016) that NECO examinations are more difficult than WAEC test items. More so, the study disagree with the findings of Abiri (2006) and Olutola (2015) which say 10 difficulty indices of multiple choice test with a fewer number of options say four differ significantly to anyone with a larger number of options.

Nevertheless, their item discrimination shows a

significant difference. Items of WAEC distinguish better between examinees with high and low ability. The higher discrimination of WAEC items might be due to the fact that WAEC adopt fewer options (4 options) as against NECO items with 5 options. Olatunji (2007) reported that test item with fewer options had the best discriminating index. Findings from this study were consistent with the earlier work by Olatunji (2007); Olutola (2015); Thomas et al. (2016) that items constructed by WAEC have more discriminating items than NECO test items. However, researchers including Alfred (2013); Metibemu (2016) and Ogbebor (2017) concluded in their different studies that there was no statistically significant difference in the slope of test items constructed by public examining bodies in Nigeria such as WAEC, NECO etc. that is, their test items discriminate equally while their asymptotic parameter showed insignificant difference. It can be said that the reliability of the estimate requires assumption that the tendency for an examinee to guess on any item is entirely a quality of the item which has exactly the same fixed effect on all examinees. However, examinees vary in their tendency to guess, some examinees guess a lot, some a little, and some hardly ever. The findings of this study agree with the work by Alfred (2013); Metibemu (2016) and Ogbebor (2017) that there was no statistically significant difference in the guessing parameter of test items constructed by public examining bodies in Nigeria.

Conclusion

It can be concluded that mathematics examination items constructed by WAEC and NECO have comparable difficulty and guessing indices but differ in discriminating index. The disparity in the performance of the examinees might be as a result of difference in method of computation of results by the examination bodies and other variables but not the difference in the quality of their items.

Recommendations

Based on the findings of the study, the following recommendations are made:

- (i) Certificates issued by WAEC and NECO could be used freely for same purposes without any discrimination since the items of the examinations are of comparable standard.
- (ii) The examination bodies should ensure that their items are constructed in a way to ensure high discrimination between brilliant and dull examinees
- (iii) Government should constitute a joint examination evaluation team to ensure that examination items constructed by different examination bodies but for same

purposes are of comparable standard.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

REFERENCES

- Abiri JO (2006). Elements of Evaluation Measurement and Statistical Techniques in Education. Ilorin: Library and Publication Committee, University of Ilorin, Nigeria.
- Adedoyin OO (2010). Using IRT approach to detect gender biased items in public examination. *Educational Research and Reviews Academic Journals* 5(7):385-399.
- Adewale AE (2008). Comparative Analysis of the Psychometric properties of WAEC, NECO and NABTEB Mathematics Achievement Examinations. Unpublished Ph.D Thesis, University of Ado-Ekiti, Nigeria.
- Alfred AO (2013). Scoring Tests with dichotomous and Polychromous items. Maureen Grass Dean of Graduate School University of Georgia.
- Anigbo LC (2018). Comparison of 2017 west Africa Examination Council (WAEC) and National Certificate Examination Council (NECO) Senior Secondary School Certificate Examination Mathematics Question. *International Journal of Education and Evaluation* 4(3):21-28.
- Awogbemi CA, Oloda FSS, Alagbe SM (2015). A Correctional analysis of students' achievement in WASSCE and NECO Mathematics. *IOSR Journal of Research and Method in Education* 5(4):14-21.
- Bande SO, Adewale AE (2013). Comparative analysis of the reliability and validity coefficient of WAEC, NECO and NABTEB constructed mathematics examination. *Journal of Educational and social research* 3(2):761-764.
- Kolawole EB (2007). A Comparative Analysis of Psychometric properties of two Nigerian examining bodies for Senior Secondary Schools Mathematics. *International Research Pakistan Journal of Applied Sciences* 2(8):913-915.
- Lord FM (1980). Application of item Response theory to practical testing Problems. New Jersey; Lawrence Erlbaum Association, Inc.
- Metibemu MA (2016). Comparison of Classical test theory and item Response theory in the development and Scoring of Senior Secondary School Physics tests in Ondo State. Unpublished Ph.D thesis. Institute of Education, University of Ibadan.
- Hambleton RK, Swaminathan H, Rogers HJ (1991). Fundamental of item responses Theory. Social Sciences Research. SAGE.
- Ogbebor UC (2017). Construction of Mock Economics Test for Senior Secondary School Students in Delta State, Nigeria Using Classical Test and item Response theories. Unpublished Ph.D Thesis. Institute of Education, University of Ibadan.
- Olatunji DS (2007). Effects of number of options on psychometric properties of multiples choice tests in Economics. Unpublished M.Ed thesis, University of Ilorin, Ilorin.
- Oloda FS (2017). Differential items function of test items of joint senior secondary II mathematics promotion Examination in Ondo state, Nigeria. An unpolished Ph.d thesis of Ekiti State University, Ado-Ekiti, Nigeria.
- Olutola AT (2015). Item difficulty and discrimination Indices of multiple choice Biology tests; *Liceo Journal of Higher Education Research*. 11(1):16-30.
- Oshkosh UW (2002). Item Analysis, Testing Services Support; <http://www.uwosh.edu/testing/facultyinfo/ikemanalysis.php>.
- Thomas ES, Uchegbue H, Ugbe EB (2016). Comparative item analysis of students' WAEC and NECO Mathematics objectives Test item scores. *Educational Assessment and Research Network in Africa*, (3).
- Tshering G (2006). IRT in item Banking, study of DIF items and Test construction Item response Theory in Item Banking, Study of Differentially functioning items and Test construction. M.Sc thesis,

University of Twente, Netherlands.

Yu CH (2008). True score Model and Item Response theory Retrieved from <http://www.creativewistom.com/teaching/WBI/measurement>.

Zaman A, Kashmir A, Mubarak M, Ali A (2008). Student' Ranking, Based on their Abilities on objective type: Comparison of CTT and IRT. Edu – COM 2008 international conferences on sustainability in higher Education: Direction for change Edith Cowan University, Australia, held on 19th-21st November, 2018 Research online 592-599 <http://ro.ecu.edu.au/ceducom/52>