

**Journal of Applied Research and  
Multidisciplinary Studies (JARMS)**

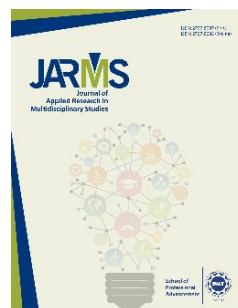
**Volume 3 Issue 2, Fall 2022**

ISSN(P): 2707-5087 ISSN(E): 2707-5095

Homepage: <https://journals.umt.edu.pk/index.php/jarms>



Article QR



**Title:** Psychometric Analysis of Senior Secondary School Certificate Examination (SSCE) 2017 NECO English Language Multiple Choice Test Item in KWARA State Using Item Response Theory

**Author (s):** Jimoh Kasali<sup>1</sup>, Oluwaseyi Aina Opesemowo<sup>2</sup>, Yinusa Akintomide Faremi<sup>3</sup>

**Affiliation (s):** <sup>1</sup>Department of Educational Foundations and Counselling, Obafemi Awolowo University, Ile-Ife

<sup>2</sup>Department of Educational Foundation, Faculty of Education, Ajayi Crowther University, Oyo

<sup>3</sup>Department of Educational Foundations and Management University of Eswatini Kwaluseni Campus, Eswatini

**DOI:** <https://doi.org/10.32350/jarms.32.01>

**History:** Received: June 30, 2022, Revised: July 6, 2022, Accepted: August 1, 2022

**Citation:** Kasali, J., Opesemowo, O. A., & Faremi, Y. A. (2022). Psychometric analysis of senior secondary school certificate examination (SSCE) 2017 NECO English language multiple choice test item in KWARA state using item response theory. *Journal of Applied Research and Multidisciplinary Studies*, 3(2), 00-00.  
<https://doi.org/10.32350/jarms.32.01>

**Copyright:** © The Authors

**Licensing:**



This article is open access and is distributed under the terms of [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

**Conflict of Interest:**

Author(s) declared no conflict of interest



A publication of  
School of Professional Advancements  
University of Management and Technology, Lahore, Pakistan

# **Psychometric Analysis of Senior Secondary School Certificate Examination (SSCE) 2017 Neco English Language Multiple Choice Test Items in Kwara State Using Item Response Theory**

Jimoh Kasali <sup>1\*</sup>, Oluwaseyi Aina Opesemowo<sup>2</sup>, Yinusa Akintoye Faremi<sup>3</sup>

<sup>1</sup>Department of Educational Foundations and Counselling, Obafemi Awolowo University, Ile-Ife

<sup>2</sup>Department of Educational Foundation, Faculty of Education, Ajayi Crowther University, Oyo

<sup>3</sup>Department of Educational Foundations and Management University of Eswatini, Kwaluseni Campus, Eswatini

## **Abstract**

The study determined the dimensionality of 2017 National Examinations Council (NECO) English Language multiple-choice test item and estimated the item parameter indices (discrimination, difficulty, guessing and carelessness) using four parameter logistic model. The ex-post facto design was employed for the study. The population for the study comprised all candidates/test-takers who enrolled and sat for June/July Senior School Certificate Examination (SSCE) 2017 NECO English Language Examination in Kwara State, Nigeria with 12,000 samples purposively selected from sixteen Local Government Area in the State. The research instruments used for the study were Optical Marks Record Sheets for the NECO June/July 2017 English Language objectives items. The responses of the testees were scored dichotomously. The data collected were calibrated using four parameters logistic model. The results showed that the 2017 English Language multiple-choice item among SSCE students in Kwara State does not violate the assumption of unidimensionality which made the items reliable for use in assessing knowledge of students in English language. Also, the results showed that only two items were able to suit the 4-PLM based on the rule of thumb. While the remaining items does not suit the 4-PLM. It was recommended among others that NECO and other examination bodies should intensify more efforts toward improving the standard of the English Language test items using 4-PLM, which is the new trend for estimating item parameter indices.

---

\* Corresponding Author: [jimoh.bukola@yahoo.com](mailto:jimoh.bukola@yahoo.com)

**Keywords:** Carelessness, dimensionality, discrimination, difficulty, guessing item response theory

## Introduction

English language as a school subject plays a strategic role in the school system because almost all the school subjects are taught in English in countries where the English Language is an official language. A child cannot learn most of the elementary facts or ideas unless he/she understands the language in which these ideas are expressed. Thus, at the Senior School Certificate Examination (SSCE) level, a credit pass or failure in the subject (English Language) determines the educational advancement of Senior Secondary School (SSS) students. The achievement of students in English Language most especially in external examination has been a source of concern to education stakeholders (including parents, teachers, educators and researchers). Students must have credit in Mathematics and English Language which are major requirements in most courses before admission into tertiary institutions. Given this requirement, there is need to determine the level of achievement in both subjects across the states in Nigeria based on the geopolitical zones, for a clear picture of the entire situation (Atanda, 2011). The alarming rate of failure and poor performance of students in English language in most external examinations has been a source of concern to the education stakeholders. The West Africa Examination Council (WAEC) in 2014 traced the persistent poor performance of students in the examination to lack of adequate preparation. Many scholars also blame the poor performance of students in English language on the fact that learners are not willing to read or lack of adequate reading material to engage with. Many students find it difficult to interact meaningfully with reading materials before them. Thus, the consistent poor performance of students in external examinations such as WAEC and National Examinations Council (NECO) particularly in English Language raises serious concern.

A sample of test taker's action is frequently observed using psychological and educational assessments. Majority of educational psychologists concerned with assessing capabilities and abilities of test takers, as a result, understanding how test-takers' ability influence the correctness of an answer on an item (Lord, 2012). An examinee having the required knowledge on the item is expected to produce a correct response on test item, whereas an examinee without the required information on the

item is expected to give an incorrect answer. However, in the case of multiple-choice assessments, this known-correct assumption may not always correctly reflect what happens in the real sense. An examinee's responses on a multiple-choice test can be divided into three categories: responses that reflect the examinee's genuine ability (e.g., accurate or incorrect); correct responses resulting from lucky guesses; and false responses arising from nervousness, carelessness, or distraction. Because the latter two sorts of aberrant responses do not reflect the examinee's actual knowledge, they may lead to an incorrect evaluation of the examinee's true aptitude. Test items were often weighted equally in traditional tests, the impact of abnormal responses was restricted. In the case of item response theory, on the other hand, lucky predictions and thoughtless errors can lead to an estimation bias (Liao, Ho, Yen, & Cheng, 2012).

### **Literature Review**

In Social Sciences, item response theory (IRT) models are commonly employed. Although IRT models were first used in education, they are now used in a variety of fields, including personality (Loken & Rulison, 2010). In a multiple-choice exam, IRT is critical for scale development and generating correct latent trait estimations. With the rise in IRT applications comes the need to carefully analyze various parametric forms for IRT models, as well as their interpretation and implications for conclusions. Modeling based on item response theory (IRT) has a lengthy history and a large body of work (Baker & Kim, 2004). Item response theory (IRT) is a method of measuring a hypothesized latent construct such as ability or attitude in current educational and psychological settings. These hidden characteristics can't be measured directly on people; instead, they have to be measured through replies to items or questions in a test or survey. In a test or survey, IRT methods are widely used to derive latent scores for individual respondents on traits, abilities, competency, or attitude. In a testing environment, IRT is arguably best understood in terms of latent trait ability. In fact, educational testing was one of the first areas where IRT was used. The IRT scoring method takes into account the respondent's latent variable as well as the item's difficulty and discriminatory characteristics. IRT is employed in a variety of domains, including psychometrics, educational sciences, sociology, health professions, and computer adaptive testing (CAT). In addition, because IRT models employ information about

an item's attributes to evaluate and refresh an instrument, it can be used in test or instrument development.

Abnormal responses might lead to ability estimation errors, incorrect items may be chosen. The three-parameter logistic (3PL) model is an appropriate answer for evaluating an examinee's ability in circumstances when guessing is likely to be a factor affecting the examinee's test responses (Amarnani, 2009). To model the effect of guessing, the 3PL model includes a guessing parameter (the lower asymptote).

According to Ojerinde (2013), latent traits are setup as a variable that is not directly detectable yet quantifiably affects discernible attributes. Through the perception of these attributes, it is feasible to make surmises about the presence or extent of these qualities by standard measurements which are the test items. The connection between these items and the latent trait is thought to be direct and the items are thought to be restrictively independent. The response to an item ought to be represented completely by this latent trait. Any covariance among the items is because of their regular reliance on the accepted latent trait and there ought to be no covariance among item responses along some other latent measurement for the basis of objectivity to be fulfilled. A later conceptualisation of the basic connection between response to an item and the ability controlled by an individual is the item response theory, which is probabilistic in its methodology. The general of the IRT Model as follow:

Rasch model defines the following:

$$P_i(\theta_j) = \frac{e^{(\theta_j - b_i)}}{1 + e^{(\theta_j - b_i)}}, \quad (i)$$

where  $e$  denotes a constant 2.718, where  $j$  is the power level and  $b_i$  is the individual item's trouble boundary. The position boundary is the item trouble boundary  $b_i$ , since it reflects the ability level at which a large percentage of examinees correctly address the item. Although the theoretical range of  $b_i$  estimation is from to, the average range is assumed to be from - 3 to +3. The two-boundary logistic model adds one item boundary, the discrimination boundary, to the Rasch model and is defined as

$$P_i(\theta_j) = \frac{e^{ai(\theta_j - b_i)}}{1 + e^{(\theta_j - b_i)}}, \quad (ii)$$

where  $a_i$  is the discrimination parameter, and 1.7 is the scaling constant that determines the ICC's slope or steepness. While the theoretical range of  $a_i$  is  $[-\infty, \infty]$ , the value of  $a_i$  for the correct answer to an item is generally positive, and the value seen in practice is normally less than 2.5. (Baker & Kim, 2004). A low  $a_i$  value and a relatively flat ICC indicate that the item is ineffective at distinguishing between different skill levels. The three-parameter model's equation is

$$P_i(\theta_j) = c_i + (1 - c_i) \frac{e^{a_i(\theta_j - b_i)}}{1 + e^{(\theta_j - b_i)}}, \quad (\text{iii})$$

where  $c_i$  is the pseudo-guessing parameter, which reflects the chance of correctly guessing the object using just guessing.

To simulate a parameter for the upper asymptote in the item characteristic curve, Barton and Lord (1981) suggested 4PL IRT. Due to nervousness and carelessness, this model accounts for unexpected wrong responses (missing) of examinees with a high ability level. The probability of accurate response given the ability level is written as equation in the general form of this model.

$$P_{4PL}(\theta) = c + (d - c) \frac{1}{1 + \exp[-1.702a(\theta - b)]}, \quad (\text{iv})$$

Where  $P_{4PL}(\theta)$  ranges from the lower asymptote  $c$  to 1, and  $P_{4PL}(\theta)$  ranges from  $c$  to the upper asymptote parameter  $d$ . (i.e., slipping parameter) of item. Although Barton and Lord (1981) advocated for a single upper asymptote for all test items, the basic form of the 4PL model allowed for a distinct upper asymptote to be estimated for each test item. In the last decade, one-, two-, and three-parameter logistic (1PL, 2PL, and 3PL) IRT models for dichotomous items have gotten a lot of attention (Magis, 2013). Recently, the 4PL IRT model was not a widely used IRT model among practitioners and researchers due to a lack of evidence for its benefits, difficulties in estimating the upper asymptote, and the lack of computer software programs that practitioners and researchers could use to implement the 4PL IRT model (Loken & Rulison, 2010). With the emergence of increasingly powerful computer software packages such as the "mirt" package in the R program, the 4PL IRT model has grown more prominent in recent years, notably in the literature on IRT and computerized adaptive testing (CAT) (Meng et al., 2019).

Item Response Theory is based on three fundamental assumptions, the first of which is unidimensionality. Unidimensional item response models are those that assume a single latent ability. "What is necessary for the unidimensionality assumption to be met satisfactorily is the presence of one dominant factor that effects test performance," Adedoyin and Adedoyin (2013) said. The second assumption is local independence, which states that an examinee's chances of correctly answering a question are unaffected by the answers given to other questions on the test. The aim of researching a test's internal structure is to demonstrate that all of the items operate together, hence the art of assessing dimensionality is to determine the least number of latent ability domains defined in a test. According to Stevina (2011), the number of abilities or constructs measured by a test or a set of items is referred to as dimensionality in assessment. In light of the foregoing, Stevina (2011) defined dimensional structure as the relationship between the test items and the latent proficiencies that the test is supposed to measure.

According to McDonald (2000), the issue of dimensionality entails more than (successfully) identifying a set of proficiencies that explain for item responses. He pointed out that, in addition to determining the number of dimensions that underpin the item responses, the relationship between the items and dimensions is critical in dimensionality evaluation.

In assessment settings, a set of items is considered to be unidimensional if the data is based on a single characteristic, but multidimensional if the data is based on numerous traits. The Multidimensional IRT (MIRT) is a mathematical model that describes the relationship between two or more unobservable variables defined as dimensions and the chance of an examinee successfully answering a specific test item (Ackerman, Gierl & Walker, 2003). Items on an exam may evaluate multiple domains of abilities; nevertheless, this is not an issue as long as the evaluation is assessing the same composite for all students.

Testee-item interaction may cause different composites of ability to be measured for testees with varied backgrounds on particular exams. Multidimensional models, like unidimensional models, are based on two assumptions. Monotonicity and Local Independence are two of them. According to the monotonicity assumption, as an examinee's skill level rises, so does the likelihood of the examinee properly answer any given test item (Smith, 2009). One of the primary goals of education in Nigeria, as

stated in the National Policy on Education (FGN, 2004), is to prepare young people for future difficulties and to develop them to satisfy the country's manpower demands. As a result, conducting examinations both inside and outside of schools as a basis for assessment becomes extremely important.

In this context, the study was conducted to determine the dimensionality of 2017 NECO English Language multiple-choice test item and estimate the item parameter indices (discrimination, difficulty, guessing and carelessness) using four parameter logistic model in order to ascertain how the items suitable for the 4-PLM which is a new trend in estimating item parameter indices.

Based on the above objectives, the following research questions were raised. (a) What is the dimensionality of 2017 NECO English Language multiple choice test items among SSCE students in Kwara State? (b) What are the item parameter indices (discrimination, difficulty, guessing and carelessness) using four parameter logistic model?

### **Methodology**

The ex-post facto design was employed for the study. The population for the study comprised all candidates/test-takers who enrolled and sat for June/July SSCE 2017 National Examinations Council (NECO) English Language Examination in Kwara State, Nigeria. The sample comprised 12,000 candidates who sat for the examination in three senatorial districts in the state (i.e Kwara South, Kwara North and Kwara Central). The sample purposively selected from sixteen Local Government Area in the state. This technique was used to ensure a fairly equal representation of sample. The research instruments used for the study were Optical Marks Record Sheets for the National Examination Council (NECO) June/July 2016 SSCE English Language objectives items. The responses of the testees were scored dichotomously. Data collected were analysed using Dimensionality Test (DIMTEST) package.

### **Results**

#### **Research Question One**

What is the dimensionality of 2017 NECO English Language multiple-choice test items among SSCE students in Kwara State?

To answer the research question one, the dimensionality of the 2017 NECO English Language multiple-choice test items among SSCE students



in Kwara State, examinees' responses were subjected to a test of essential unidimensionality using the Dimensionality Test (DIMTEST) in DIMPACK 1.0 package

**Table 1**

*Dimensionality of 2017 NECO English Language Test Items*

TL	TGbar	T	<i>p</i> -value
10.2316	9.4853	0.7426	0.2289

The result of the test of essential unidimensionality was used to examine the assumption of unidimensionality of the 2017 NECO English Language test items form a secondary dimension. This was done by differentiating the test into two subtests, namely the Assessment Subtest (AT) and the Partitioning Subtest (PT). The AT is the item chosen as those that measures best in the direction most opposite to that of the PT items. The AT was empirically selected using the HCA/CCPROX cluster procedure and DETECT statistic in DIMTEST. This item cluster was tested to determine if it was dimensionally distinct from the remainder of the test. A random sample of 30% of the students' responses was used to select the AT, and the remaining 70% of the examinees' responses were used as PT.  $p > 0.05$  is interpreted as statistically insignificant indicating essential unidimensionality. Table one showed that the AT was not dimensionally distinct from the remaining item of the test ( $T = 0.7426$ ,  $p$ -value = 0.2289, one-tailed); therefore, the assumption of unidimensionality was upheld. This implies that the 2017 English Language multiple-choice item among SSCE students in Kwara State does not violate the assumption of unidimensionality.

## Research Question Two

What are the item parameter indices (discrimination, difficulty, guessing, and carelessness) using the 4 parameters logistic model?

To answer this research question, the calibration of students' responses in the 2017 English Language multiple-choice items among SSCE students in Kwara State using a 4-parameter logistic IRT model in JMetrik (Baker & Kim, 2004; Adeyemow & Opesemowo, 2020). The result is presented in Table 2 showing the discrimination, difficulty, guessing, and carelessness parameters of individual items.

**Table2***2017 NECO English Language Items Using Four Parameter Logistic Model*

Item	model	a-param	Remark	b-param	Remark	c-param	Remark	u-param	Remark	Over Remark
1	L4	1.50	*	-1.15	*	0.39	**	0.98	**	**
2	L4	138.60	*	2.33	*	0.21	*	0.85	**	**
3	L4	1.42	*	-2.28	*	0.19	*	0.96	**	**
4	L4	2.63	*	-0.63	*	0.40	**	0.96	**	**
5	L4	2.01	*	-1.05	*	0.37	**	0.98	**	**
6	L4	2.84	*	-0.97	*	0.28	*	0.93	**	**
7	L4	1.28	*	-2.11	*	0.09	*	0.96	**	**
8	L4	1.60	*	-1.90	*	0.12	*	0.95	**	**
9	L4	1.53	*	-1.96	*	0.13	*	0.98	**	**
10	L4	2.86	*	-0.53	*	0.35	**	0.96	**	**
11	L4	6.84	*	-0.38	*	0.43	**	0.60	*	*
12	L4	1.67	*	-0.98	*	0.30	*	0.99	**	**
13	L4	2.85	*	-0.83	*	0.36	**	0.97	**	**
14	L4	1.35	*	-2.26	*	0.13	*	0.96	**	**
15	L4	1.35	*	-2.38	*	0.07	*	0.93	**	**
16	L4	7.51	*	-0.43	*	0.53	**	0.60	*	*
17	L4	1.87	*	-1.97	*	0.25	*	0.96	**	**
18	L4	1.54	*	-1.96	*	0.11	*	0.93	**	**
19	L4	1.80	*	-1.79	*	0.16	*	0.95	**	**
20	L4	2.70	*	-0.93	*	0.21	*	0.94	**	**

Item	model	a-param	Remark	b-param	Remark	c-param	Remark	u-param	Remark	Over Remark
21	L4	1.91	*	-1.57	*	0.21	*	0.97	**	**
22	L4	2.95	*	-0.95	*	0.38	**	0.89	**	**
23	L4	2.90	*	-1.39	*	0.27	*	0.91	**	**
24	L4	1.64	*	-1.16	*	0.50	**	0.95	**	**
25	L4	2.27	*	-1.10	*	0.42	**	0.97	**	**
26	L4	91.55	*	2.38	*	0.12	*	0.88	**	**
27	L4	2.98	*	-0.59	*	0.13	*	0.86	**	**
28	L4	2.88	*	-1.04	*	0.34	*	0.98	**	**
29	L4	1.24	*	-2.05	*	0.15	*	0.98	**	**
30	L4	1.43	*	-2.10	*	0.14	*	0.98	**	**
31	L4	2.69	*	-0.75	*	0.25	*	0.97	**	**
32	L4	1.41	*	-1.67	*	0.07	*	0.99	**	**
33	L4	1.80	*	-1.41	*	0.17	*	0.99	**	**
34	L4	2.82	*	-1.06	*	0.28	*	0.98	**	**
35	L4	2.92	*	-0.77	*	0.09	*	0.98	**	**
36	L4	1.42	*	-2.11	*	0.11	*	0.97	**	**
37	L4	2.04	*	-1.14	*	0.21	*	0.99	**	**
38	L4	1.40	*	-2.08	*	0.06	*	0.97	**	**
39	L4	1.34	*	-1.60	*	0.04	*	0.98	**	**
40	L4	1.53	*	-1.37	*	0.08	*	0.96	**	**
41	L4	91.40	*	2.38	*	0.18	*	0.97	**	**
42	L4	1.65	*	-2.01	*	0.05	*	0.98	**	**

Item	model	a-param	Remark	b-param	Remark	c-param	Remark	u-param	Remark	Over Remark
43	L4	97.92	*	2.37	*	0.16	*	0.98	**	**
44	L4	1.61	*	-1.94	*	0.10	*	0.97	**	**
45	L4	1.29	*	-2.05	*	0.06	*	0.99	**	**
46	L4	131.26	*	2.34	*	0.21	*	0.85	**	**
47	L4	1.25	*	-2.25	*	0.06	*	0.96	**	**
48	L4	1.38	*	-1.91	*	0.07	*	0.98	**	**
49	L4	1.61	*	-1.55	*	0.23	*	0.99	**	**
50	L4	7.93	*	-0.41	*	0.48	**	0.60	*	**
51	L4	2.27	*	-0.95	*	0.24	*	0.94	**	**
52	L4	1.72	*	-1.71	*	0.15	*	0.95	**	**
53	L4	1.96	*	-2.20	*	0.05	*	0.98	**	**
54	L4	7.49	*	-0.48	*	0.45	**	0.60	*	**
55	L4	2.29	*	-2.24	*	0.08	*	0.97	**	**
56	L4	1.21	*	-1.62	*	0.26	*	0.98	**	**
57	L4	1.81	*	-2.25	*	0.11	*	0.95	**	**
58	L4	2.92	*	-1.14	*	0.14	*	0.95	**	**
59	L4	121.15	*	2.35	*	0.18	*	0.91	**	**
60	L4	1.34	*	-2.11	*	0.05	*	0.93	**	**
61	L4	1.86	*	3.96	**	0.28	*	0.60	*	**
62	L4	1.49	*	-1.94	*	0.06	*	0.98	**	**
63	L4	2.96	*	-0.43	*	0.49	**	0.60	*	**
64	L4	1.86	*	-1.85	*	0.07	*	0.96	**	**

Item	model	a-param	Remark	b-param	Remark	c-param	Remark	u-param	Remark	Over Remark
65	L4	1.38	*	-1.98	*	0.09	*	0.96	**	**
66	L4	2.54	*	-0.98	*	0.32	*	0.95	**	**
67	L4	1.33	*	-2.10	*	0.08	*	0.94	**	**
68	L4	6.18	*	-0.54	*	0.54	**	0.60	*	**
69	L4	2.89	*	-0.77	*	0.33	*	0.98	**	**
70	L4	2.96	*	-0.65	*	0.16	*	0.85	**	**
71	L4	9.22	*	-0.41	*	0.41	**	0.60	*	**
72	L4	0.82	*	28.43	**	0.02	*	0.91	**	**
73	L4	1.73	*	-1.47	*	0.19	*	0.97	**	**
74	L4	1.45	*	-2.01	*	0.11	*	0.97	**	**
75	L4	1.77	*	-1.24	*	0.37	**	0.97	**	**
76	L4	147.00	*	2.34	*	0.10	*	0.93	**	**
77	L4	1.07	*	-2.10	*	0.10	*	0.99	**	**
78	L4	1.19	*	-1.84	*	0.10	*	0.99	**	**
79	L4	1.35	*	-1.70	*	0.09	*	0.99	**	**
80	L4	8.63	*	-0.41	*	0.44	**	0.60	*	**
81	L4	0.82	*	22.01	**	0.08	*	0.91	**	**
82	L4	165.23	*	2.33	*	0.17	*	0.97	**	**
83	L4	142.28	*	2.34	*	0.21	*	0.91	**	**
84	L4	2.94	*	-0.55	*	0.21	*	0.98	**	**
85	L4	1.45	*	-1.01	*	0.29	*	0.96	**	**
86	L4	1.39	*	-1.98	*	0.06	*	0.95	**	**

Item	model	a-param	Remark	b-param	Remark	c-param	Remark	u-param	Remark	Over Remark
87	L4	1.21	*	-2.12	*	0.13	*	0.97	**	**
88	L4	166.55	*	2.33	*	0.08	*	0.84	**	**
89	L4	0.82	*	23.97	**	0.13	*	0.91	**	**
90	L4	2.44	*	-0.77	*	0.38	**	0.96	**	**
91	L4	0.82	*	27.06	**	0.07	*	0.91	**	**
92	L4	2.30	*	-0.71	*	0.41	**	0.97	**	**
93	L4	0.82	*	25.60	**	0.05	*	0.91	**	**
94	L4	0.82	*	24.46	**	0.08	*	0.91	**	**
95	L4	0.82	*	22.98	**	0.08	*	0.91	**	**
96	L4	0.82	*	30.12	**	0.05	*	0.91	**	**
97	L4	0.82	*	26.39	**	0.02	*	0.91	**	**
98	L4	0.82	*	21.56	**	0.11	*	0.91	**	**
99	L4	13.03	*	-0.45	*	0.39	**	0.84	**	**
100	L4	1.62	*	-0.61	*	0.54	**	0.96	**	**

Note. \* Suitable, \*\* Not Suitable

Table 2 showed the 2017 NECO English Language items using the four-parameter logistic model. It can be deduced that all the 100 items were calibrated using IRT 4-PLM in JMetrik statistical software. However, the rule of thumb for any item to be considered as being suitable must fulfill the certain condition of the parameter indices. The a-parameter is a discrimination index that must be greater than 0.2, the b-parameter is the difficulty index that ranges from -3 to +3, and the c-parameter is the guessing parameter which must be less than 0.35, while the u-parameter is the carelessness and should not be greater than 0.75. Subsequently, each item must fulfill the criteria about the rule of thumb for the overall remark (Baker & Kim, 2004). Based on the result in Table 2, it could be seen that only two items (i.e., items 11 and 16) were able to satisfy the rule of thumb while the remaining 98 items does not suit the 4-PLM because they violated the rule of thumb in one way or the other.

### **Discussion of the Findings**

The results showed that the 2017 English Language multiple-choice item among SSCE students in Kwara State does not violate the assumption of unidimensionality. The results of this finding were in agreement with Jimoh, (2021) who reported that 2016 NECO Mathematics test was essentially unidimensional. The finding of this study also in line with studies of Jiao (2004); Tomblim and Zhang (2006) and Deng, Wells and Hambleton (2008). However, the finding against the studies of Jang and Roussons (2007); Li, Jiao, and Lissitz (2012). The findings of their studies showed a clear violation of unidimensionality assumption in the tests assessed. The findings also corroborate with the finding of studies of David et al., (2017) the findings of their studies showed that NECO Senior School Certificate June/July Multiple-choice Objective Tests in Government for the years 2013 and 2014 conformed to the assumption of unidimensionality. Another result of the study showed that only two items (i.e., items 11 and 16) were able to satisfy the rule of thumb while the remaining 98 items does not suit the 4-PLM because they violated the rule of thumb in one way or the other. However, this does not affect the validity and reliability of the 2017 NECO English multiple-choice questions because the item does not violate the assumption of unidimensionality. This study is in line with the observation of Jimoh (2022) that 2016 NECO Mathematics test items were suitable items as their difficulty parameters were within the range (-2 to 2) for which items difficulty parameter estimates are considered suitable,

majority of the items were not suitable with discriminating power and almost test items were not vulnerable to guessing.

### **Conclusion and Recommendations**

Based on the findings of this study, it was concluded that the 2017 English Language multiple-choice item among SSCE students in Kwara State does not violate the assumption of unidimensionality. It was recommended that the National Examinations Council (NECO) should intensify more efforts toward improving the standard of the English Language test items using 4-PLM because is the new trend for estimating item parameter indices (discrimination, difficulty, guessing, and carelessness). It is also recommended that all examining bodies (such as West African Examination Council, NABTEB etc) using multiple-choice test instruments should be encouraged to use 4-PLM of the item response theory approach when developing test items.

### **References**

- Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice: MIRT Instructional Module*.
- Adedoyin, O. O., & Adedoyin, J. A. (2013). Assessing the comparability between classical test theory (CTT) and Item response theory (IRT) models in estimating test item parameters. *Herald Journal of Education and General Studies*, 2(3), 107-114.
- Adegbile, J. A. (1999). *The relative effectiveness of three models of expository advance organiser on secondary students' learning outcomes in reading comprehension*. Unpublished Ph.D Thesis, Univesity of Ibadan, Ibadan.
- Adeyemo, E. O. & Opesemowo, O. A. (2020). Differential Test let Functioning (DTLF) in Senior School Certificate English Language Examination Using Multilevel Measurement Modelling. *Sumerianz Journal of Education, Linguistics and Literature*, 3(11), 249-253. <https://doi.org/10.47752/sjell.311.249.253>
- Amarnani, R. (2009). Two theories, one theta: A gentle introduction to item response theory as an alternative to classical test theory. *The*



*International Journal of Educational and Psychological Assessment*, 3, 104-109.

- Atanda A.I. (2011). A Survey of Secondary Students Achievement in English Language and Mathematics in Nigeria: Lessons for Secondary School Administrators in Nigeria. *Journal of Sociology and Education in Africa*, 10 (2), 126-147
- Baker, F. B. & Kim, S. (2004). *Item response theory: Parameter estimation techniques*. 2nd ed. New York Marcel Dekker.
- Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). United States of America: ERIC Clearing House on Assessment and Evaluation.
- Baker, F. B., & Kim, S. H. (2004). *Item response theory: parameter estimation techniques* (2nd ed.). New York: Marcel Dekker.
- Barton, M. A., & Lord, F. M. (1981). An upper asymptote for the three-parameter logistic item-response model (Research Report 18-21). Princeton, NJ: Educational Testing Service. doi: 10.1002/j.2333-8504.1981.tb01255.x
- David, A. A., Henry, O. O., Mayowa, O. O., Joseph, T. A., & Samuel, T. B. (2017). Analysis of the dimensionality of Nigerian Senior School Certificate Examination June/July 2013/2014 objective test in Government. *Journal of Science and Technology, Mathematics and Education*, 13(4), 83-93
- Deng, N., Wells, C., & Hambleton, R. (2008). A confirmatory factor analytical study examining the dimensionality of educational achievement tests. *NERA Conference proceedings 2008 paper 31*. [http://digitalcommons.ucom.edu/nera\\_2008/31](http://digitalcommons.ucom.edu/nera_2008/31)
- Jang, E. E., & Roussos, L. (2007). An investigation into the dimensionality of TOEFL using conditional covariance-based nonparametric approach. *Journal of Educational Measurement*, 44, 1-22.
- Jiao, H. (2004). Evaluating the dimensionality of the Michigan English Language assessment battery. *Spain Fellow Working papers in Second or Foreign Language Assessment*, 2, 27-51.

- Jiao, H. (2004). Evaluating the dimensionality of the Michigan English Language assessment battery. *Spaan Fellow Working papers in Second or Foreign Language Assessment*, 2, 27–51.
- Jimoh, K. (2021). *Gender and culture-related differential item functioning in 2016 National Examinations Council Mathematics multiple choice questions in Nigeria*. Unpublished Ph.D. Thesis, Obafemi Awolowo University, Ile-Ife.
- Jimoh, K., & Adediwura. A. A. (2022). Estimation of Item Parameter Indices of NECO Mathematics Multiple Choice Test Items Among Nigerian Students. *Journal of Integrated Elementary Education*, 2(1), 43-54. <https://doi.org/10.21580/jieed.v2i1.10187>.
- Kyung, T. H. (2013). *Windows software that generates IRT parameters and item responses: research and evaluation program methods (REMP)*. Amherst: University of Massachusetts.
- Li, Y., Jiao, H., & Lissitz, R. W. (2012). Applying multidimensional item response theory models in validity test dimensionality: An example K-12 large-scale science assessment. *Journal of Applied Testing Technology*, 13,(2), 1-27.
- Liao, W. W., Ho, R. G., Yen, Y. C., & Cheng, H. C. (2012). The four-parameter logistic item response theory model as a robust method of estimating ability despite aberrant responses. *Social Behavior and Personality: An International Journal*, 40(10), 1679-1694. doi: 10.2224/sbp.2012.40.10.1679
- Loken, E. & Rulison, K. L. (2010). Estimation of a 4-parameter Item Response Theory model. *The British Journal of Mathematical and Statistical Psychology*, 63(3), 509-525. doi:10.1348/000711009X474502,
- Lord, F. M. (2012). *Applications of item response theory to practical testing problems*. New Jersey, NJ: Lawrence Erlbaum Associates
- Magis, D. (2013). A note on the item information function of the four-parameter logistic model. *Applied Psychological Measurement*, 37(4), 304-315. doi: 10.1177/0146621613475471
- McDonald, R. P. (2000). A basis for multidimensional item response theory. *Applied Psychological Measurement*, 24, 99 - 114.

- Meng, X., Xu, G., Zhang, J., & Tao, J. (2019). Marginalized maximum a posteriori estimation for the four-parameter logistic model under a mixture modelling framework. *British Journal of Mathematical and Statistical Psychology*, *Advanced online publication*. doi: 10.1111/bmsp.12185
- Ojerinde, D. (2012). *Introduction to item response theory, parameter models, estimation and application*. Abuja, Nigeria: Marvellous Press.
- Reeve, B. B. (2000). *Item and scale-level analysis of clinical and non-clinical sample responses to the MMPI-2 depression scales employing item response theory* (Unpublished doctoral dissertation). University of North Carolina, Chapel Hill.
- Smith, J. (2009). Some issues in item response theory: Dimensionality assessment and models for guessing. *Unpublished Doctoral Dissertation*. University of South California.
- Stivena, D. (2011). *Assessing dimensionality in complex data structures: A performance comparison of DETECT and NOHARM procedures*. Unpublished Ph.D. Dissertation, Arizona State University.
- Tomblin, J. B., & Zhang, X. (2006). The dimensionality of language ability in school-age children. *Journal Language and Hearing of Speech, Research Vol. 49*. 1193-1
- Wilson, K. M. (2000). An explanatory dimensionality assessment of the TOEFL test. Educational Testing Service; Research Report, RR-00-14. Princeton, New Jersey.
- Wright, B. D. (1999). Fundamental measurement for psychology. In: S. E. Embretson & S. L. Hershberger (Eds.). *The new rules of measurement*. (pp. 65-104). Mahwah, N J: Lawrence Erlbaum Associates.