

## Assessment of Psychometric Properties of Government Tests in Osun State Unified Promotion Examinations

<sup>1</sup>Femi T. Adekunle ; <sup>2</sup>E. R. I Afolabi

<sup>12</sup>Department of Educational Foundations and Counselling, Obafemi Awolowo University, Ile-Ife

### Abstract

*The quality of government tests in the Osun State Senior Secondary Schools' Unified Promotion Examinations was examined by investigating its dimensionality, item difficulty, discrimination indices, local item independence and model fit of the items. For this study, Descriptive survey design was used. The population of the study comprised 27,470 Senior Secondary 2 (SS2) students from Osun State who took the 2019 Unified Promotion Examinations (UPE). 4,423 SS2 students who sat for Government test in the examinations were purposefully selected. Profoma was used for data collection where the OMR sheets of students were retrieved from the Osun State Ministry of Education. The data was analyzed with the sirt models package because it's include the multidimensionality compensatory and non-compensatory IRT Models. The Government test yielded DETECT = -0.288, ASSI = -0.390, and RATIO = -0.482. The results revealed that the Government examination was mainly one-dimensional. The difficulty index ranged between 50% and 51%, and the discrimination index returned discrimination parameter estimates that were more than 1.7. Out of the 40 items tested, item 7 was found to violate item local independence. The 4PL returned the best fit ( $P < 0.05$ ,  $RMSEA > 0.05$ ). The study concluded that government test used in the unified promotion examination in Osun Sate was unidimensional; almost all the items of the Government tests were moderately difficult and highly discriminated examinees that are proficient in government from those who were not.*

**Keywords:** Dimensionality, Multidimensional Item Response Theory, Psychometric Property

### Introduction

The assessment of individual's traits, such as achievement, ability, proficiency, attitude, belief, or other constructs, is frequently accomplished through the administration of different instruments based on the need and administered according to standardized procedures. According to studies (Adekunle et al, 2021; Adekunle & Faleye, 2019; Osterlinds, 2006; Ajeigbe & Afolabi, 2014; Kinanee, Bosede, & Orluwene, 2017), academic performance in the test of government, whether good or poor, is not steady but varies yearly, with low performance being more pronounced. The drop in students' performances, on the other hand, has been related to a variety of factors, which are generally classified as either test attributes or personal factors. Rather than inabilities of students, declines in students' academic achievement are frequently attributed to flaws in

the psychometric properties of administered tests (Adekunle et al, 2021; Ajeigbe & Afolabi, 2014). Individual teachers in each school set questions and marked the students' answer scripts for the terms' examinations internally at all levels prior to the implementation of the unified promotion examinations programme in the Osun State public secondary schools. Teachers worked at their own pace, and there was a high rate of non-completion of subject schemes of work (Osun Education Factsheet, 2019). This could open the door to favoritism, examination malpractices, and leakage of examinations questions. According to Osadebe, 2014; Akamigbo, Eneja, & Ikeh, 2013, one of the most serious issues in Nigerian senior secondary schools is that teachers, after assuming that the curriculum for social science subjects (Economics, Government, and Geography) of the Senior Secondary Certificate Examination (SSCE) has

been covered, will assess senior secondary three (SSS III) students with unreliable summative achievement tests in the form of mock (last test for SS III students before sitting for the Senior Secondary Certificate Examinations). To create summative achievement assessments, most teachers would rapidly duplicate questions from any previous question paper, failing to establish validity and reliability (Osadebe, 2014; Akamigbo, Eneja & Ikeh, 2013).

### **Examinations and Students' Academic Achievements in Government**

According to Kwasu (2013), government, as a subject, is taught at the senior secondary school level and is one of the examination subjects for the National Examination Council (NECO) or the West African Examination Council's Senior Secondary School Certificate (WAEC). According to Ogunjawa and Udoh (2015), one of the Nigerian state's cardinal aspirations is to consolidate the country's fledgling democratic experience; hence the state should focus on efficient teaching and learning of Government as a subject in the Senior Secondary Schools. This will go a long way toward instilling democratic values and civic mindsets in learners, which will lead to peace, stability, progress, and national integration in the long run. However, many students have lost interest in the subject over the years since it is regarded as unimportant in comparison to the sciences. Several studies have found that students' capacity and motivation to learn have a significant impact on secondary school achievement.

Also, Kwasu, (2013) and Rono & Rono, (2016) found that students performances are poor in government compared to the sciences in their study on student's interest and academic performance. The study emphasized the importance of government to secondary school students in areas such as civic education, government roles and obligations, good governance circumstances, and political stability indices, among other things. The efficiency of

teaching and learning government as a subject in senior secondary schools was investigated by Ogunjawa and Udoh (2015). According to the findings, teaching the subject will assist students develop democratic virtues and civic attitudes, resulting in peace, stability, progress, and national integration. The study found that having access to professionally-qualified teachers is critical. However, there is a scarcity of data on students' success in internal examinations and the quality of evaluation in public secondary schools.

High failure rates in internal examinations have been attributed to different factors among which are lateness and irregular school attendance by teachers and students, non-completion of subject schemes by teachers, inability of teachers to account for students' performance, inadequate and ineffective supervision of academic activities at the school and district levels, and low confidence in public schools by parents (Osun Education Factsheet, 2019). However, the quality test items to determine proficiency and how well students have acquired the expected mastery of the subject area to perform creditably well in both internal and external examinations or when called upon to demonstrate such skills has not been explored, which is why this study was conducted.

In the field of education, Item Response Theory (IRT) is also employed as a standardized tool used for examining the psychometric properties of new and established instruments. In contrast, a review of measurement-based articles (Adebule, 2005; Andala, Digoloji, & Kamandii, 2014 & Adeyemo, 2018) published in journals specific to the field of education between 2004 and 2018 explored test characteristics in IRT analysis to evaluate the psychometric properties of external examinations in science based subjects and English language, but there is a paucity of literature in education using IRT to explore the quality of test items used for government tests in the Osun State unified promotion examinations.

### **Multidimensional Item Response Theory Model**

Kose and Demirtasli (2012) opined that multidimensional IRT (MIRT) are models which explain the relationship between two or more unobservable variables, conceptualized as constructs or dimensions, and the probability of the examinee who is correctly answering any test item by the mathematical model (Ackerman, Gierl & Walker, 2003 as cited by Kose & Demirtasli, 2012). Multidimensional models, like unidimensional models, are based on specific assumptions. Monotony and local independence are two of them. Monotonicity suggests that when an examinee's ability level rises, so does the likelihood of the examinee properly answering any given test item (Kose and Demirtasli 2012). Local independence is defined as the probability of solving any item without regard to the outcome of any other item, while person and item factors are controlled for (Embretson & Reise, 2000; Kose & Demirtasli, 2012). Multidimensional Item Response Theory (MIRT), according to Ha (2017), is an extension of unidimensional IRT that allows for the simultaneous study of many constructs.

### **Purpose of the Study**

The purpose of this study is to evaluate the psychometric qualities as well as the multi-traits dimensions that underpin student performance in government in secondary schools in Osun State.

### **Research Questions**

1. How many dimensions underlie Government tests in the Osun State Senior Secondary Schools' Unified Promotion Examination?
2. To what extent do the items in government test in the Osun State senior secondary school Unified Promotion Examination satisfy the assumption of local independence of items?
3. What are the levels of difficulty of the items on the Government test?
4. What are the discriminating indices for the Government test items?

### **Methodology**

The research design that was adopted for this study was descriptive survey design. The study was carried out in Osun State, Nigeria. All senior secondary school II (SSS2) students in Osun State public secondary schools make up the study's population. The Osun State Ministry of Education (OSMOE, 2019) provided secondary data on the total number of SS II pupils which revealed that there were 27, 470 SSS II students. The sample for this study consisted of 4,423 students out of the 27,470 that sat for the 2019 Unified Promotion examinations in Osun State using a multistage sampling procedure. From the existing three senatorial districts in the state, 9 local government areas (LGAs) were selected using simple random sampling technique. All the high schools in the LGAs selected were purposively selected to participate in the study based on the existence of SSII students in the schools. This sample was based on the requirements for estimation of item parameter for obtaining adequate model calibration under the multidimensional graded response model cited in Jian, Wang & Weiss (2016).

The study extracted students' responses from the optical mark recognition (OMR) answer sheets of the sample from the study population (Government study). The examinations were multiple choice types. The responses of the sample population were scored in binary format by assigning 1 (one) to item that was correctly answered and 0 (zero) to an item that was not correctly answered. Data collected were analyzed using both confirmatory Dimensionality evaluation to Enumerate Traits (confirmatory DETECT) to determine whether the test was unidimensional or not. Also, the test items were calibrated with Multidimensional 4 parameter logistics item response theory (4-PLIRT) using item analysis statistics built in Supplementary Item response Theory (sirt package in R).

**Results.**

*Research Question 1:* How many dimensions underlie Government test used in the Osun state senior secondary schools' unified promotion examination?

The responses of the students were assessed in order to ascertain its dimensionality. The number of dimension(s) underlying students' performance was determined because dimensionality entails isolating the number of dimension(s) accounting for the variation observed in students' performance in each test. To do this, the students' responses were subjected to two levels of analysis. The appropriate number of dimensions underlying each examination was found after determining whether the test was unidimensional or not. Confirmatory Dimensionality Evaluation to Enumerate Contributing Traits (confirmatory DETECT) was used to assess the tests' unidimensionality, and then exploratory Dimensionality Evaluation to Enumerate Contributing Traits was used to determine the optimal number of dimensions underlying each examination (exploratory DETECT).

When the Dimensionality Evaluation to Enumerate Contributing Traits (DETECT) index is less than 0.2, the Approximate Simple Structure Index (ASSI) is less than 0.25, and the approximate simple structure index ratio index (RATIO) is less than 0.36, a test is considered unidimensional. When DETECT is greater than 1.00, it is strongly multidimensional, moderately multidimensional when DETECT is greater than 0.40 but less than 1.00, and barely multidimensional when DE. When a test has multidimensionality, the values of the Approximate Simple Structure Index (ASSI) and the approximate simple structure index ratio index are used to measure the test's dimensionality. When ASSI is close to 1 and RATIO is close to 1, a simple structure is assumed (Baker, 2001; Hasmy, 2014). Supplementary Item Response Theory Models was used to conduct the analysis. The result is as follows:

**Table 1: Dimensionality Assessment of Government Examination**

Index	Estimate
DETECT	-0.288
ASSI	-0.390
RATIO	-0.482

**DETECT < 0.20, ASSI < 0.25 and RATIO < 0.36**

The test of dimensionality of Osun State senior secondary schools' 2019 UPE Government examination was seen in Table 4.3. The Government test yielded DETECT = -0.288, ASSI = -0.390, and RATIO = -0.482, as shown in the table. The results revealed that the Government examination was mainly one-dimensional (DETECT < 0.20, ASSI < 0.25, and RATIO < 0.36). The result implies that the Government test was unidimensional; the diversity in student responses to the Government test was accounted for by one dominant dimension.

*Research Question 2:* To what extent do the items in government test in the Osun State senior secondary school Unified Promotion Examination satisfy the assumption of local independence of items?

To provide an answer to this research question, the responses of the examinees t/in government test was subjected to Yen Q3 statistic. Yen Q3 statistics is the linear correlation existing among the residuals of pairs of items on a test. According to Yen (1984), Q3 value for pair of items above 0.2 indicates dependence (that is, one of the items in the examination violates the assumption of local independence). To calculate the LID values of test items, two stages of analysis are conducted: selection of the item response model that is most suitable for estimating the probability of correct response that is used for calculating the residual

used for the item local independence assessment. Thereafter, the LID index for the test items was calculated based on the IRT model that fitted the data best.

### Model-data fit assessment of Government test

Table 2 (a) presents the result of the model-data fit assessment of the Government test.

**Table 2(a): Model-data fit of Government test**

Model	AIC	SABIC	BIC	-2logLikelihood
1PL	238009.400	238158.800	238289.100	237927.400
2PL	235772.000	236063.500	236317.700	235612.000
3PL	235853.100	236290.300	236671.600	235613.200
4PL	230489.900	231072.900	231581.300	230170.000

Table 2 (a) showed the model-fit assessment of the Government fit to one-parameter logistic model (1PL), two-parameter logistic model (2PL), three-parameter logistic model (3PL) and four-parameter logistic model (4PL) respectively. The table showed that when the test was calibrated with 1PL, AIC = 238009.400, SABIC = 238158.800, BIC = 238289.100, -2logLikelihood = 237927.400. When the test was calibrated with 2PL, AIC = 235772.000, SABIC = 236063.500, BIC = 236317.700, -2logLikelihood = 235612.000. When the test was calibrated with 3PL, AIC = 235853.100, SABIC = 236290.300, BIC = 236671.600, -2logLikelihood = 235613.200 and when the test was calibrated with 4PL, AIC = 230489.900, SABIC = 231072.900, BIC =

231581.300, -2logLikelihood = 230170.000. The result showed that 4PL returned the smallest AIC, SABIC, BIC and -2loglikelihood values among all the IRT models that were used in the calibration of the Government test. The implication of the result is that 4PL was the most appropriate IRT model for the calibration of Osun state senior secondary schools' unified promotion examination government test. Thus, the Yen Q3 statistics for the test was calculated based on 4PL. The resulting Q3 statistics for the Government test is presented in Table 2 (b).

**Table 2 (b): Inter-correlational Matrix of the residual of Osun State senior secondary schools' unified promotion examination Government test items**

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.10	1.00								
3	0.00	0.05	1.00							
4	0.04	0.06	-0.01	1.00						
5	-0.01	0.12	0.04	0.16	1.00					
6	-0.01	0.04	-0.06	0.13	0.16	1.00				
7	0.01	0.02	<b>0.29</b>	0.02	0.01	-0.04	1.00			
8	0.00	-0.03	-0.05	-0.03	-0.01	-0.01	-0.01	1.00		
9	0.04	0.02	0.13	0.00	0.00	0.01	0.14	0.02	1.00	
10	-0.05	-0.01	-0.05	0.02	0.04	0.02	0.03	0.05	-0.05	1.00
11	-0.08	-0.03	-0.03	-0.02	0.00	-0.03	-0.06	0.02	-0.03	-0.03
12	-0.01	0.03	-0.01	0.06	0.05	0.03	0.01	-0.01	0.06	-0.04
13	-0.05	-0.07	-0.01	-0.06	-0.05	-0.01	0.01	0.03	0.03	0.02
14	-0.01	-0.01	-0.04	-0.02	-0.05	0.01	-0.03	0.02	-0.06	0.03
15	0.01	-0.03	0.03	-0.02	-0.04	-0.05	0.08	0.02	0.05	0.00
16	-0.03	-0.09	-0.01	-0.05	-0.02	-0.01	-0.03	-0.04	0.00	-0.02
17	0.02	0.01	0.07	0.00	-0.03	-0.05	0.09	-0.07	0.05	-0.05
18	-0.07	-0.11	-0.05	-0.06	-0.03	0.03	-0.09	0.00	-0.01	-0.07
19	-0.05	-0.09	0.10	-0.01	-0.05	-0.10	0.06	-0.08	-0.01	-0.09
20	-0.03	-0.07	-0.08	-0.07	-0.07	-0.05	-0.06	-0.01	0.03	-0.02

21	-0.02	-0.02	0.05	0.01	-0.01	-0.01	0.04	-0.07	0.02	0.01
22	0.04	0.01	-0.07	-0.04	0.02	0.02	-0.04	0.01	-0.04	-0.02
23	-0.05	-0.02	0.10	-0.03	-0.03	-0.02	0.05	-0.06	0.00	-0.02
24	-0.03	-0.03	-0.07	-0.02	-0.03	-0.04	-0.08	-0.03	-0.07	-0.02
25	-0.03	0.01	0.04	-0.01	0.03	0.04	0.08	-0.08	0.02	0.01
26	-0.02	-0.02	0.15	0.01	-0.03	-0.03	0.09	-0.03	0.05	-0.04
27	0.01	-0.02	-0.10	-0.05	0.00	-0.01	-0.04	-0.01	-0.06	-0.03
28	-0.02	-0.08	-0.07	-0.02	-0.02	-0.04	-0.04	-0.03	-0.06	-0.04
29	0.02	-0.01	-0.03	0.01	0.02	0.03	0.00	-0.03	0.06	0.02
30	-0.03	-0.07	-0.06	-0.01	-0.04	-0.06	-0.07	-0.01	-0.08	-0.03
31	-0.08	-0.02	-0.11	-0.03	-0.04	-0.04	-0.14	-0.02	-0.06	-0.05
32	0.03	0.00	-0.14	0.05	0.04	0.05	-0.13	0.00	0.02	-0.03
33	-0.08	-0.06	-0.05	-0.03	-0.03	-0.08	-0.04	-0.07	-0.01	-0.05
34	-0.04	-0.02	0.08	-0.04	-0.02	0.02	0.10	0.02	0.04	-0.01
35	-0.09	-0.09	-0.03	-0.09	-0.06	-0.04	-0.01	0.01	-0.06	-0.05
36	0.00	0.00	-0.09	0.04	-0.07	-0.02	-0.02	0.02	-0.01	-0.06
37	-0.06	-0.01	0.06	-0.03	0.02	0.01	0.09	-0.01	0.02	-0.03
38	-0.07	-0.06	0.06	-0.06	-0.08	-0.07	0.12	-0.04	0.01	-0.02
39	-0.04	-0.01	-0.03	-0.03	-0.06	-0.07	-0.03	-0.11	-0.09	0.00
40	-0.06	-0.05	0.10	-0.04	-0.09	-0.09	0.10	-0.08	-0.01	0.00
	11	12	13	14	15	16	17	18	19	20
11	1.00									
12	0.03	1.00								
13	0.05	-0.06	1.00							
14	0.03	0.00	0.03	1.00						
15	-0.03	0.10	0.00	0.09	1.00					
16	0.02	-0.06	0.11	0.08	-0.05	1.00				
17	-0.05	0.04	0.00	0.01	0.00	-0.05	1.00			
18	0.09	-0.04	0.02	-0.01	-0.01	0.00	-0.08	1.00		
19	0.07	0.04	-0.03	-0.01	-0.03	0.03	0.04	0.05	1.00	
20	0.01	-0.07	0.05	-0.01	-0.07	0.05	0.06	0.04	0.00	1.00
21	-0.05	-0.04	-0.06	-0.08	-0.09	-0.05	-0.02	-0.08	0.03	-0.03
22	-0.03	-0.01	-0.01	-0.11	-0.03	-0.05	-0.09	-0.03	-0.07	-0.06
23	-0.03	0.03	-0.05	0.04	0.06	-0.02	0.02	-0.02	0.00	-0.10
24	0.00	-0.02	-0.06	-0.01	-0.04	-0.01	-0.04	-0.02	0.00	-0.03
25	-0.04	0.04	-0.08	0.00	-0.02	-0.05	0.08	-0.04	0.02	-0.06
26	-0.04	0.01	-0.05	-0.02	0.01	-0.03	-0.01	-0.02	0.04	-0.07
27	-0.02	-0.05	-0.06	-0.03	-0.03	-0.05	-0.03	-0.04	-0.09	-0.01
28	-0.02	-0.06	-0.06	-0.01	0.00	-0.01	-0.03	-0.01	-0.04	-0.05
29	-0.04	0.04	-0.04	0.04	0.01	0.01	0.14	-0.06	-0.01	0.02
30	0.05	-0.05	-0.01	-0.01	-0.05	-0.01	0.01	0.03	0.03	0.03
31	0.00	-0.04	-0.03	0.00	-0.09	-0.02	-0.05	-0.02	-0.07	0.03
32	-0.02	0.04	-0.02	0.04	-0.03	-0.04	0.01	-0.02	-0.05	-0.03
33	-0.02	-0.05	-0.04	-0.01	-0.09	-0.05	-0.06	0.03	0.05	-0.02
34	-0.02	0.03	0.02	-0.01	0.00	0.01	0.10	-0.02	0.02	0.02
35	0.00	-0.10	0.00	0.02	-0.06	0.00	-0.09	0.06	0.04	0.01
36	-0.07	0.02	-0.06	-0.06	0.03	-0.10	0.04	-0.07	-0.03	0.00
37	-0.04	-0.02	-0.04	-0.02	-0.03	-0.07	0.04	-0.04	0.01	-0.04
38	0.01	-0.05	-0.01	-0.05	-0.11	-0.04	0.07	-0.06	0.06	0.03
39	0.04	-0.02	0.00	-0.05	-0.11	-0.01	0.00	0.04	0.12	0.05
40	0.02	-0.06	-0.04	-0.07	0.00	-0.03	0.04	-0.03	0.13	0.02
	21	22	23	24	25	26	27	28	29	30
21	1.00									
22	0.05	1.00								
23	0.01	-0.01	1.00							
24	0.02	0.07	-0.03	1.00						
25	0.04	0.05	0.09	0.00	1.00					

26	0.03	0.02	0.13	-0.01	0.08	1.00					
27	-0.01	0.00	0.04	0.08	0.05	-0.04	1.00				
28	-0.06	-0.01	-0.02	0.08	0.00	-0.03	0.07	1.00			
29	0.03	-0.02	0.07	-0.04	0.11	0.08	0.00	-0.01	1.00		
30	-0.06	-0.03	-0.01	0.04	-0.04	0.00	0.04	0.01	-0.02	1.00	
31	-0.08	-0.03	-0.05	0.01	-0.06	-0.03	0.04	0.07	-0.01	0.01	
32	-0.08	0.01	0.02	0.01	0.00	-0.01	0.06	0.05	0.13	0.01	
33	-0.08	-0.04	-0.02	-0.02	-0.04	-0.04	-0.01	-0.02	-0.01	0.07	
34	-0.01	0.01	0.05	-0.04	0.02	0.01	-0.06	0.00	0.05	-0.01	
35	-0.06	-0.05	-0.04	0.03	-0.07	0.00	-0.03	0.02	-0.10	0.06	
36	-0.08	0.00	-0.05	-0.01	0.01	-0.04	0.01	-0.02	-0.03	0.03	
37	0.02	-0.08	0.05	0.00	0.07	0.02	-0.02	0.00	-0.01	-0.02	
38	0.06	-0.03	-0.01	-0.01	0.06	0.03	-0.07	-0.01	0.03	0.02	
39	0.09	0.00	-0.02	0.00	0.01	-0.05	-0.01	-0.01	0.01	0.07	
40	0.08	0.00	-0.02	-0.03	0.05	0.07	-0.08	-0.09	-0.03	0.03	
	31	32	33	34	35	36	37	38	39	40	
31	1.00										
32	0.13	1.00									
33	0.03	-0.03	1.00								
34	0.01	0.05	-0.08	1.00							
35	0.06	-0.02	0.11	-0.01	1.00						
36	0.05	0.08	0.02	-0.04	-0.01	1.00					
37	0.00	0.00	-0.04	0.14	0.00	-0.09	1.00				
38	0.03	-0.05	0.03	0.04	0.04	0.03	0.03	1.00			
39	0.05	-0.05	0.06	0.01	0.03	-0.02	0.05	0.18	1.00		
40	-0.06	-0.14	0.00	-0.01	-0.07	0.01	0.01	0.18	0.13	1.00	

**Table 2** (b) showed the correlation coefficient existing among the residuals of pairs items of the Government test data (Yen Q3 statistics). The table showed that out of the several pairs of items, only one pair, items 3 and 7 showed correlation coefficient (0.29) greater than the benchmark, 0.2. The result showed that one of the items 3 and 7 violated the item local independence. Because item 3 preceded item 7, item 7 was taken to be the item that violated item's local independence. The implication of the result is that one item, item 7 violated item's local independence.

Research Question 3: *What are the difficulty levels of the items in Government used in the Osun State senior secondary schools' unified promotion examination?*

The responses of students who took the Government examinations were calibrated with a

unidimensional 4-PL IRT model to answer this research question. The following is a summary of the findings: Table 3 showed the difficulty metrics retrieved from the calibrated Government test for the 39 items on the Government test. Because one of the original 40 items (item 7) of the examination violated the assumption of item local independence, the calibrated test has 39 items. The difficulty parameters were judged based on the criterion established by Baker, (2001) and Hasmy (2014). According to the authors, an item is: Very difficult when  $b$  or MDIFF  $\geq 2$ , difficult when  $0.5 \leq b$  or MDIFF  $< 2$ , moderately difficult when  $-0.5 \leq b$  or MDIFF  $< 0.5$ , easy when  $-2 \leq b$  or MDIFF  $< -0.5$  and very easy when  $b$  or MDIFF  $< -2$ . The results are presented as follow:

**Table 3: Difficulty parameter estimates of the Government test items.**

Item	b	Remark	Item	b	Remark
1	0.05	Moderately Difficult	22	0.19	Moderately Difficult
2	0.15	Moderately Difficult	23	-0.11	Moderately Difficult
3	-0.08	Moderately Difficult	24	0.02	Moderately Difficult
4	0.13	Moderately Difficult	25	0.13	Moderately Difficult
5	0.08	Moderately Difficult	26	0.13	Moderately Difficult
6	0.07	Moderately Difficult	27	0.26	Moderately Difficult
8	0.12	Moderately Difficult	28	0.22	Moderately Difficult
9	-0.08	Moderately Difficult	29	-0.10	Moderately Difficult
10	0.31	Moderately Difficult	30	0.03	Moderately Difficult
11	0.06	Moderately Difficult	31	0.19	Moderately Difficult
12	0.29	Moderately Difficult	32	0.02	Moderately Difficult
13	0.14	Moderately Difficult	33	0.22	Moderately Difficult
14	0.17	Moderately Difficult	34	-0.03	Moderately Difficult
15	0.24	Moderately Difficult	35	0.13	Moderately Difficult
16	0.23	Moderately Difficult	36	0.26	Moderately Difficult
17	0.51	Difficult	37	-0.04	Moderately Difficult
18	0.17	Moderately Difficult	38	0.11	Moderately Difficult
19	0.10	Moderately Difficult	39	0.01	Moderately Difficult
20	0.13	Moderately Difficult	40	0.48	Moderately Difficult
21	0.48	Moderately Difficult			

Table 3 showed the difficulty estimates of the government test items. The table showed that 38 of the 39 calibrated items of the government test returned difficulty estimate falling within -0.5 and 0.51 and only one item, item 17 returned difficulty estimate that was less than 2 but greater than 0.5, 0.51. The result showed that almost all the items of the government test were moderately difficult. The implication is that the items on the government test can be scored correctly by at least 50.0% of the examinees that are averagely proficient in government.

*Research Question 4:* What are the discrimination indices of items in the tests of government in Osun

State senior secondary schools' unified promotion examination?

The discrimination of government tests was judged based on the criterion established by Baker, (2001) and Hasmy (2014). According to the authors, an item is described as Very highly discriminating when  $a$  or MDISC  $\geq 1.7$ , highly discriminating when  $1.35 \leq a$  or MDISC  $< 1.7$ , moderately discriminating when  $0.65 \leq a$  or MDISC  $< 1.35$ , lowly discriminating when  $0.35 \leq a$  or MDISC  $< 0.65$  and very lowly discriminating when  $a < -0.35$ . The results are presented as follow:



**Table 4:** Discrimination parameter estimate of the Osun State 2019 UPE Government examination

Item	a	Remark	Item	a	Remark
1	5.737	Highly discriminating	25	5.293	Highly discriminating
2	5.397	Highly discriminating	26	5.162	Highly discriminating
3	12.96	Highly discriminating	27	6.637	Highly discriminating
4	5.394	Highly discriminating	28	6.079	Highly discriminating
5	7.031	Highly discriminating	29	10.513	Highly discriminating
6	6.486	Highly discriminating	30	7.64	Highly discriminating
7	9.741	Highly discriminating	31	7.344	Highly discriminating
8	8.635	Highly discriminating	32	7.616	Highly discriminating
9	13.829	Highly discriminating	33	5.307	Highly discriminating
10	4.621	Highly discriminating	34	15.882	Highly discriminating
11	7.856	Highly discriminating	35	7.231	Highly discriminating
12	4.495	Highly discriminating	36	4.986	Highly discriminating
13	7.382	Highly discriminating	37	25.508	Highly discriminating
14	5.822	Highly discriminating	38	6.525	Highly discriminating
15	5.052	Highly discriminating	39	9.513	Highly discriminating
16	5.14	Highly discriminating	40	3.053	Highly discriminating
17	3.414	Highly discriminating			
18	5.173	Highly discriminating			
19	5.988	Highly discriminating			
20	7.048	Highly discriminating			
21	3.448	Highly discriminating			
22	4.639	Highly discriminating			
23	24.297	Highly discriminating			
24	7.803	Highly discriminating			

### **a: discriminating parameter**

Table 4 showed the discrimination parameter of Government test. The table showed that all the items of the calibrated 39-item government returned discrimination parameter estimates that were greater than 1.7. The result revealed that the government test's items highly discriminated examinees that are proficient in government from those who were not proficient in government.

### **Discussion**

The uni-dimensionality of the government supported the chief examiners' reports of the WAEC on candidates' performance in government which pointed out misinterpretation of the questions, inadequate preparation was among factors responsible for poor performance of students in the examination (WAEC, 2018). All these factors centered on verbal skills. This also corroborated the report of the findings of this study that only one dimension accounted for the variation observed in students' responses to the government test. From an expert's opinion (a lecturer from the department of political science) who gave the examination's questions face validity. The ability to read is necessary for success in the examination.

Based on the result of the dimensionality, model fits of the items used in Government revealed that 4PL was the most appropriate IRT model for the calibration of Osun State senior secondary schools' unified promotion examination for Government Test. The result of the study is in line with the findings of Loken and Rulison (2010) on "Estimation of a 4-parameter Item Response Theory model" which showed that 4PLM has a better model fit when compared with other models. Thus, the Yen Q3 statistics for the test was calculated based on 4PL. The resulting Q3 statistic for the Government test showed that out of the several pairs of item, only one pair, items 3 and 7 (one of items 3 and 7) violated the item's local independence. Also, item 3 preceded item 7, item 7 was taken to be the item that violated item's local

independence. The implication of the result is that one item, item 7, violated item's local independence.

The difficulty parameters were judged based on the criterion established by Baker, (2001); Hasmy (2014). The result showed that almost all the items of the Government test were moderately difficult. The implication is that the items on the Government test can be scored correctly by at least 50.0% of the examinees that are averagely proficient in Government. The discrimination parameter of Government test of Osun State Unified Promotion Examinations revealed the items highly discriminate examinees that are proficient in government from those who were not proficient in Government.

### **Recommendations**

Based on the findings, the following recommendations were made:

- i. The Department of Curriculum Development and Evaluation (CD&E) of the Osun State Ministry of Education should adopt and use M4PL IRT models to investigate the psychometric quality of test items used for the state unified promotion examinations.
- ii. The Department of CD&E of the Osun State Ministry of Education should also extend test dimensionality to other subjects using M4PL IRT models to determine the dimensions underlying the tests in order to help teachers identify skills and abilities needed to improve students' performance in both internal and external examinations;
- iii. The Department of CD&E of the Osun State Ministry of Education should use standardized tests for termly school internal examinations across the state not only as promotion examinations but to monitor the teachers and their mastery of the subject as well as their coverage of syllabus.

- iv. School heads should ensure that standardized tests are developed and used by teachers as an evaluation tool for preparing students for internal examinations.
- v. Teachers should be trained to use IRT models in developing and assessing tests characteristics of test items used for internal assessment of students.
- vi. Teachers should be encouraged to use standardized tests to investigate the least learned content areas of their students and effect remediation promptly.
- vii. Students should be encouraged to use standardized tests to reinforce and improve their learning.

### Conclusion

The study concluded that 4-parameter IRT model (4PL) is dependable and most appropriate for the evaluation of the psychometric quality of government test items used in the Osun State Unified Promotion Examinations. Also, the Paper I Multiple Choice Tests utilized for government in the state's unified promotion examinations are of good quality based on the analysis of data obtained. As a result, students should be encouraged to take standardized tests as a means of reinforcing and improving their knowledge.

### 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*, 22, 37-51.
- Adebule, S. O. (2005). Functional assessment for quality education in Nigeria. *International Journal of Research in Education*, 2(1&2), 173-178.
- Adekunle, F. T. & Faleye B. A. (2019). Path-analytic study of the factors influencing students' performance in senior secondary school in Osun State, Nigeria. *FUDMA Journal of Educational Foundations*, 2(2), 1-14.
- Adeyemo, E. O. (2018). Equating the state unified and West African School Certificate mathematics examination Items. *American International Journal of Contemporary Research*, 8(4). <http://doi:10.30845/ajicr.v8n4p10>.
- Ajeigbe, T. O. & Afolabi, E.R.I (2014). Assessing unidimensionality and differential item functioning in qualifying examination for senior secondary school students, Osun State, Nigeria. *World Journal of Education*, 4 (4).
- Akamigbo I. S., Eneja R. U.& Ikeh E. F (2013). Assessment of gender disparity in achievement test item format among student of economics in senior secondary schools. *Asian Journal of Educational Research*, 1(1), 2013
- Andalai H. O., Digolooii O. and Kamandeeiii M. (2014). Reliability of mock examinations for prediction of the Kenya certificate of secondary examination (kcse) results. *Journal of Research & Method in Education*, 4(1), 28-36
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates
- Ha, D. T. (2017). Applying multidimensional item response theory in validating an english final <http://doi:10.1007/978-0-387-89976-3>
- Kinanee, J., Bosede J. N, & Orluwene, J.W. (2017). A Comparative study on item characteristics of 2014 – 2016 mathematics objective tests in junior secondary school certificate examination questions in Rivers State. *International Journal of Mathematics Trends and Technology*. 52(8).
- Kose I. A. and Demirtasli N. C. (2012). Comparison of unidimensional and multidimensional models based on item

- response theory in terms of both variables of test length and sample size *Procedia - Social and Behavioral Sciences*, 46, 135 – 140.
- Kwasau, M. A. (2013). Political science education in Nigeria: an imperative tool for sustainable development. *European Scientific Journal*; 9(8)
- Ogujawa, A. I. and Udoh A. P. (2015). Strategies for improving the teaching and learning of government as a core-subject in senior secondary schools. *International Journal of Educational Planning & Administration*; 5(1), 51-61.
- Osadebe, P.U., 2014. Construction of economics achievement test for assessment of students. *World Journal of Education*, 4(2): 58-64.
- Osterlind, S. J. (2006). *Modern measurement: Theory, principles, and applications of mental appraisal*. Upper Saddle River, NJ: Prentice Hall.
- Osun Education Factsheet, (2018). Impact of an 8-year focus on education reform, assessing the performance of students in the west african senior school certificate examination (wassce). Retrieve online
- Reckase M. D. (2009). *Multidimensional Item Response Theory*. Springer Dordrecht Heidelberg: [http://doi: 10.1007/978-0-387-89976-3](http://doi:10.1007/978-0-387-89976-3)
- Rono D. and Rono O. C. (2016). An assessment of the attitudes of students towards history and government in selected secondary schools in Bomet County in Kenya. *Journal of Education and Practice*, 7, (19).
- WAEC, (2018). Chief examiners' report. West African Examination Council (WAEC), 2010. Chief examiner's report