ASSESSING UNIDIMENSIONALITY AND ITEM PARAMETER ESTIMATES OF FOUR DIFFERENT PAPER TYPES OF ENGLISH LANGUAGE MULTIPLE-CHOICE TESTS USING THREE-PARAMETER MODEL

Taiwo Oluwafemi AJEIGBE
taiaje@oauife.edu.ng, ajetaiolu@gmail.com
Department of Educational Foundations and Counselling
Faculty of Education, Obafemi Awolowo
University, Ile-Ife.

and

Olayinka Ibiyemi ODERINDE
layinkaoderinde@yahoo.com
Department of Guidance and Counselling
Faculty of Education, University of Ibadan
University, Ibadan

Abstract

The study examined unidimensionality, ascertained and compared stability of item difficulty, discrimination and guessing tendencies across four different paper types of English Language multiple-choice tests of Distance Learning Centre in Obafemi Awolowo University. These were with a view of establishing parameter stability of the English Language Multiple-Choice. The study adopted a causal comparative design because students' responses were obtained from the database. The 2449 students who sat for the first contact examination during the 2015/2016 session were used as sample size for the study. The instruments used for the study were four different paper types of English Language for 2015/2016 first contact, consisting of 60 Multiple-choice items each. The items were calibrated to generate item difficulty, discrimination and guessing tendency using X-Calibre 4.2 software package. One-way analysis of variance was used to estimate statistical difference in terms of item difficulty, discrimination and guessing tendency across the four different paper types. Results obtained showed that each test paper type is unidimensional in nature. Also, out of the 60 items 25(41.7), 39(65.0) and 38(63.3); 32(53.3), 36(60.0) and 42(70.0); 30(50.0), 37(61.7) and 38(63.3); and 27(45.0), 36(60.0) and 40(66.7) fell under moderate difficulty, discrimination and acceptable guessing value of 0.00 – 0.25 across the four different paper types respectively, The results finally showed that there was no statistical significant difference in item difficulty ($F_{(3, 239)}=0.028$; $p>0.05$); discrimination ($F_{(3, 239)}=0.212$; $p>0.05$); and guessing ($F_{(3, 239)}=0.425$; $p>0.05$) respectively. The study

therefore concluded that the assumption of unidimensional was not violated and stable parameter estimates were recorded across paper types.

Keywords: *Unidimensionality, Item, Parameter estimates, English Language, Multiple-Choice Tests, Different paper types, Three-parameter model*

Introduction

One of the major components in teaching and learning processes at any educational levels is students' assessment. Its importance cannot be overstretched because it is a mode of measuring the extent to which learning has taken place. Although there are many forms of assessing learning outcome, it seems multiple-choice has gained wider acceptability among stakeholders in educational settings these days (Baghaei & Amrahi, 2011). This may not be unconnected to its objectivity in nature compared to easy form. A well-developed multiple-choice test is capable of improving learning process, covering more content areas; providing prompt feedback to the students (Haladyna, 2004). Also, it requires skillful preparation of items of acceptable parameter estimates (difficulty, discrimination and guessing), and subsequently resulting to reduced measurement errors, ensured validity and accurate interpretation of test scores (Moss, 1995). A widely used measure of identifying measurement errors and evaluation of educational, psychological, vocational assessment is the item response theory (IRT). This is sometimes referred to as the latent trait theory which is used to establish the relationship between item behaviour and ability level of the test takers.

Prior to the use of item response theory, basic assumptions are bound to be observed, namely; unidimensionality, local independence and item characteristic curve. The concept of unidemensionality deals with existence of a single overriding factor, with other factors having minor or relatively small factor components that could affect or influence the items scores (Stout, 1990). The appropriateness of IRT model and evidence of construct validity of tests rely on the unidimensionality assumption (Sheng, 2005). The concept of unidimensionality occurs when each of the items in a test measures a single trait, which in principle assumed that local independence (Ajeigbe & Afolabi, 2014). The assumption is predicated on IRT model that measures a single latent trait of a given test. According to Reckase, 1979; Drasgow and Parsons, 1983 the primary factor should contribute or explain at least 20.97% of the variance as sufficient unidimensional for using item response theory analysis.

The second assumption of local independence emphasizes that the correctness or wrongness of any given item in a test is not a function of answering any other item correctly or wrongly. This means that each item is independent of one and another and one item did not give a clue to another (Ajeigbe & Afolabi, 2014). The third assumption of item characteristic curve (ICC) is monotonic in nature. It shows the

relationship between examinees' latent trait and probability of responding correctly to a given item.

Assessing students using multiple-choice tests requires analysing individual test item's behaviour through a modern item statistics theory such as Item response theory, that engages different models capable of screening the items for possible measurement errors, as well as ascertaining item parameter estimates (difficulty, discrimination and guessing). Well-constructed multiple-choice test items will be void of measurement errors, capable of distorting the test score interpretation, as well as the end use of such test. Item response theory accommodates one-parameter logistic model (1PLM), two-parameter model (2PLM) and three-parameter model (3PLM) in estimating items' behaviour relative to individual examinee's ability. The one-parameter logistic emphasizes difficulty parameter (b-parameter), two-parameter emphasizes difficulty and discrimination parameters (a and b parameters), while the three-parameter emphasizes difficulty, discrimination and guessing parameters (Ibrahim, 2013).

$$P_i(\Theta) = \frac{1}{1 + \exp -1.7(\Theta - b)} \qquad \ldots\ldots (1)$$

According to Lord and Novick (1968) (cited in Ojerinde, 2012), 1PLM is written in mathematical expression as:

From the equation 1, theta (Θ) is constant to each item and the central parameter is 'b', which is difficulty parameter which measured by person's ability (Bond & Fox, 2001). It assumes that discrimination and guessing are constraint to zero and this is called rasch model (van der Linden &Hambleton, 1997). The mathematical expression can be presented graphically in item characteristic curve (ICC) as given in Figure 1.

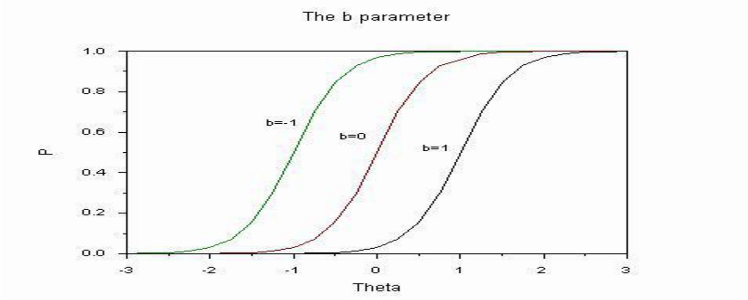

Figure 1:b-parameter (ICC)

The complexity of item response theory added another parameter called a-parameter (discrimination) that is 2PLM. The parameter "a" is estimated by taking the slope of the line tangent to the ICC at "b", as shown in Figure 2. In this case, "a" and "b" parameters vary with respect to individual item in the test, holding guessing parameter constant (David, 2011). The mathematical expression is written as:

$$P_i(\Theta) = \frac{1}{1 + \exp [-1.7a(\Theta - b)]} \qquad \ldots\ldots (2)$$

Where: Pi(Θ) = the probability that an examinee with ability level *Θ* answers item correctly;
b = the item difficulty parameter;
a = the item discrimination parameter; and
1.7= scaling factor (D).

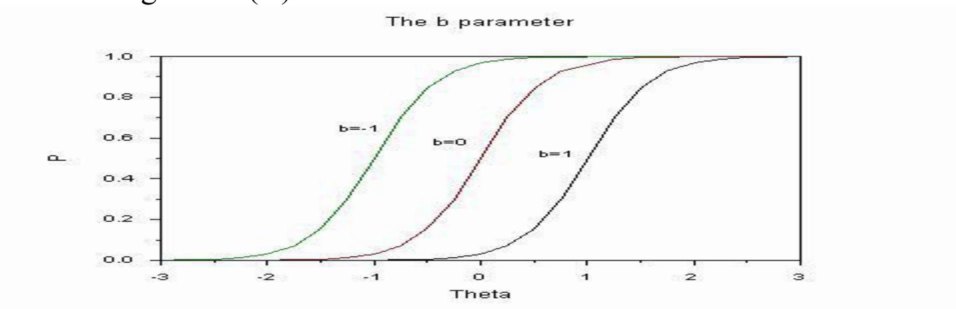

Figure 2:b-parameter (ICC)

The three-parameter model, around which this study was centred, emphasized three-parameter model (3PLM). The mode included guessing parameter, which is known as lower asymptote of the ICC (Ojerinde, Popoola, & Ajeigbe, 2020). The c-parameter is estimated at the low point of the curve that moves toward negative infinity on the horizontal axis, as shown in Figure 3. It is the probability that sometimes, students with low ability level get a particular item correctly through guessing. Theoretically, *c* ranges from 0.0 to 1.0, but is typically < 0.3 (Ojerinde et al, 2012). The mathematical

$$P_i(\Theta) = c + \frac{(1-c)}{1+\exp[-1.7a(\Theta-b)]} \quad \ldots\ldots (3)$$

expression is written as:
Where: Pi(Θ) = the probability that an examinee with ability level *Θ* answers item correctly;
b = the item difficulty parameter;
a = the item discrimination parameter;
1.7= scaling factor (D); and
c= the lower asymptote parameter.

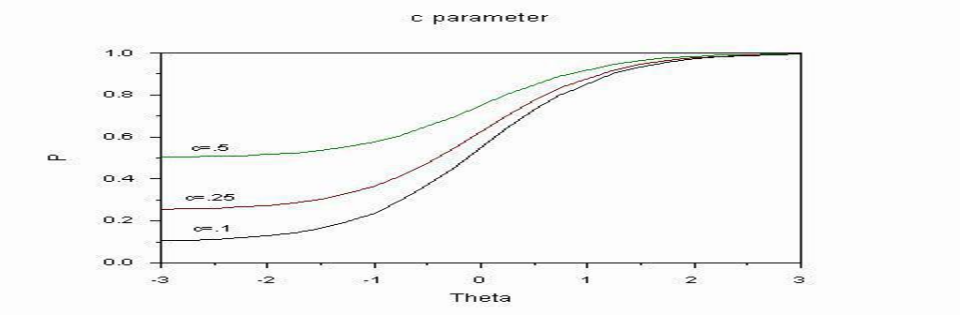

Figure 3: c-parameter

This study was premised on the three-parameter logistic model, as recommended by Thissen, & Wainer (2011), rather than using the normal ogive, which only emphasized discrimination and difficulty parameters which are related to item characteristics curve (ICC). The parameter estimates; difficulty, discrimination and guessing are essential in item response theory. Difficulty is a function of individual items, with respective latent trait (ability) of the testees to answer an item correctly or wrongly. Discrimination, on the other hand, has to do with the ability of an item to differentiate between higher ability and lower ability students. The guessing parameter is the extent to which an item is prone to answering it correctly with necessary having and in-depth knowledge of the subject matter.

To adequately address the subject matter in this study, the historical information about distance learning program in Obafemi Awolowo University needed to be briefly elucidated. The advent of the course unit system in Nigerian Universities in the 1977/1978 session, the concessional examination was abolished in 1978 and admission to universities became centralized through the Joint Admissions and Matriculation Board (JAMB) Akinyoola (2019). Well over one million applicants seek admission every year into universities, which now has a capacity of about 250,000. The competition is now at its peak and good preparation will result in better performance by the students.

This background knowledge establishes the need for a Pre-degree programme in the Arts, Commercial and Sciences, that would serve as a landmark to equip prospective students for University admission, give them a good head-start and promote superior studentship in the university. Due to the large number of students who are willing to gain admissions into universities of their choices, with a view to studying certain courses passionately, there is the need to assess them, as far as assessment is a strategy to discriminate students in terms of performance, using multiple choice question item tests. In the present educational system, multiple choice questions (MCQs) are mostly used and very common for assessing the cognitive domain capabilities of students who come for the Center for Distance Learning (CDL) Programme of the Obafemi Awolowo University, in both entrance examinations and contacts examinations. Multiple-choice questions have the inherent merits of providing a large number of

examination items that encompass many content areas; administering in a relatively short period of time; and grading of results by the computer system. Critically speaking, designing good multiple-choice questions is a cumbersome, tasking and time-consuming process. Item analysis as an approach to improving test quality using statistics and expert judgment to evaluate tests based on the quality of individual items, item sets, the entire sets of items, and the relationship of each item to other items.

Statement of Research Problem

Multiple-choice tests have gained patronage in the Nigeria education system as a means of assessing students' learning outcome. A recent dimension in the development of multiple-choice items is the use of different paper types, introduced to reduce cheating among students. As good as this development seems to be, little or no research evidence has been documented on the psychometric properties of different paper types of multiple-choice tests. In this regard, the place of IRT as a means of judging and ascertaining the adequacy of item parameter estimates prior to the final administration cannot be overstressed. There have been controversies on the underlying applications of the unidimensionality assumption of IRT model especially on language measurement. To this end, the study, among other things, aimed to ascertain if the unidimensionality assumption holds when different paper types are used in English Language multiple-choice tests. In addition, there are more literature evidences on multiple-choice option length, sample size, models and software packages than different paper types. For different paper types of the tests to be regarded as fair, valid and able to generate reliable results that can be used for making valid decisions, its stability in terms of difficulty, discrimination and guessing must be ensured. Since there has not been statistical and empirical evidence of different paper types of multiple-choice test items used by the distance learning of the Obafemi Awolowo University, the study therefore aimed at unveiling the item parameter estimates of the four different English Language multiple-choice paper types used to assess the students learning. To this end, it is germane to have documented empirical evidence of analytical parameter estimates of items' difficulty, discrimination and guessing of the different paper types of English Language multiple-choice test. Hence, the questions on unidimensionality and stability of the four different paper types of the English Language multiple-choice used by the distance learning of the Obafami Awolowo University to assess students' performance were addressed in this study.

Research Objectives

The objectives of the study were to

(a) examine unidimensionality of the four different paper types of English Language multiple-choice test items;

(b) ascertain the item difficulty, discrimination, and guessing parameters of the four different paper types of English Language multiple-choice test items; and

(c) compare stability of item difficulty, discrimination and guessing tendencies across the four different paper types.

Research Questions

(1) Do the English Language Multiple-choice tests comply with the assumption of unidimensionalty across four different paper types?

(2) What are the item difficulty, discrimination, and guessing parameters of the four different paper types of English Language multiple-choice test items?

Research Hypotheses

The following research hypotheses are generated based on the stated objectives of the study.

(1) There will be no significant difference in the item difficulty of the English Language Multiple-choice tests across the four different paper types.
(2) There will be no significant difference in the item discrimination of the English Language Multiple-choice tests across the four different paper types.
(3) There will be no significant difference in the item guessing of the English Language Multiple-choice tests across the four different paper types.

Methodology

The study adopted causal comparative research design. Since the variables of interest could not be manipulated, existed pre-groups (students were classified based on paper types), homogeneous groups and the students' responses were obtained from the existing data base of the distance centre of the Obafemi Awolowo University, Ile-Ife. The population for study comprised all the distance learning students during 2015/2016 academic session. An intact class of students (2,449) who sat for the first contact examination during the 2015/2016 session constituted the sample size for the study. The instruments used for the study were four different paper types of English Language for 2015/2016 first contact, consisting of 60 Multiple-choice items, each with a four-option length. The items were scored dichotomously ("1" for the right option and "0" for the wrong option). For the calibration, two separate data files (control and data files) were created and prompted into X-calibre 4.2 for calibration to generate coefficients for item difficulty, discrimination and guessing parameters. Before conducting Factor analysis, Kaiser-Meyer-Olkin and Bartlett's Test of Sphericity were established for sample adequacy for the data. The hypotheses were tests for statistical significant at $p<0.05$ for all the four different paper types of English Language multiple-choice tests responded to by the students. In addition, one-way analysis of variance was used to estimate statistical difference in terms of item difficulty, discrimination and guessing tendency across the four different paper types.

Results

Research Question One: Do the English Language Multiple-choice tests comply with the assumption of unidimensionalty across four different paper types?

One of the preconditions for item analysis is to examine the dimensionality of the tests. To this end, the study examined the dimension of the four different paper types. The students' responses for each of the four different paper types were marked as "1" for the right option and "0" for the wrong option. Thereafter, the data was subjected to factor analysis as propounded by Stout (1984), and the resulting outputs are presented in Tables 1, 2, 3, and 4 and the corresponding Figures 1, 2, 3, and 4 respectively.

Table 1: Total Variance Explained for English Language Multiple-Choice Items Paper One

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	5.311	8.851	8.851
2	2.328	3.880	12.731
3	1.629	2.714	15.446
4	1.570	2.617	18.063
5	1.504	2.506	20.569
6	1.476	2.461	23.029
7	1.400	2.333	25.363
8	1.355	2.258	27.620
9	1.339	2.232	29.853
10	1.299	2.164	32.017
11	1.259	2.098	34.115
12	1.256	2.093	36.208
13	1.224	2.041	38.249
14	1.185	1.974	40.223
15	1.182	1.969	42.193
16	1.136	1.894	44.086
17	1.117	1.861	45.947
18	1.110	1.850	47.797
19	1.072	1.786	49.584
20	1.065	1.775	51.358
21	1.046	1.743	53.102
22	1.032	1.720	54.821

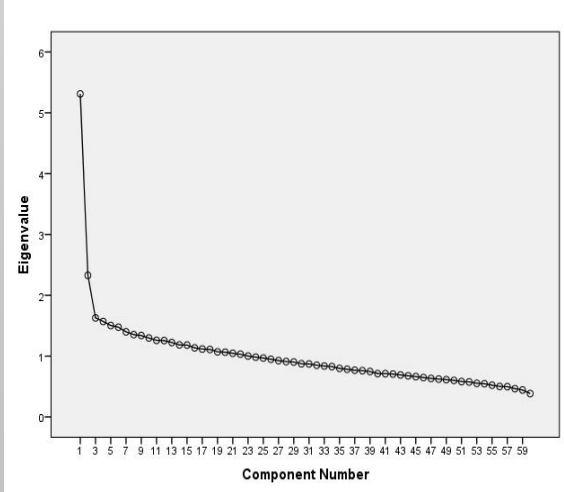

Figure 1: Scree Plot for English Language for Paper One

Table 1 showed 22 underlying factors with eigenvalues higher than one. The total variance of the first factor (5.311) was distinctly greater than twice of the second factor (2.328). The variance of the first factor explained 8.85% and the second accounted for 3.88% of the residual variance, while the remaining variance was accounted for by the other 28 factors. Also, the results yielded 22 factors with eigenvalues greater than one, which accounted for 54.82% of the total variance. In addition, the ratio of the first and

second factor produced 2.282, higher than required ratio 2.00. Based on these results, it was implied that there is only one underlying factor in the factor structure of the 60 English Language multiple-choice test items of paper type one. This is an evident that the English Language multiple-choice test item of paper type one is unidimensional in nature. The unidimensionality could also be inferred from the scree plot in Figure 1, showing a distinct factor.

Table 2: Total Variance Explained for English Language Multiple-Choice Items Paper Two

Component	Initial Eigenvalues		
	Total	% of Variance	Cum %
1	5.570	9.284	9.284
2	2.167	3.611	12.895
3	1.668	2.780	15.675
4	1.514	2.523	18.198
5	1.480	2.467	20.666
6	1.421	2.369	23.035
7	1.407	2.345	25.380
8	1.360	2.266	27.647
9	1.317	2.195	29.841
10	1.286	2.144	31.985
11	1.267	2.111	34.096
12	1.244	2.073	36.169
13	1.221	2.035	38.204
14	1.181	1.968	40.172
15	1.153	1.922	42.095
16	1.122	1.870	43.965
17	1.119	1.864	45.829
18	1.104	1.840	47.669
19	1.064	1.773	49.442
20	1.047	1.744	51.186
21	1.034	1.724	52.910
22	1.017	1.695	54.605
23	1.001	1.668	56.273

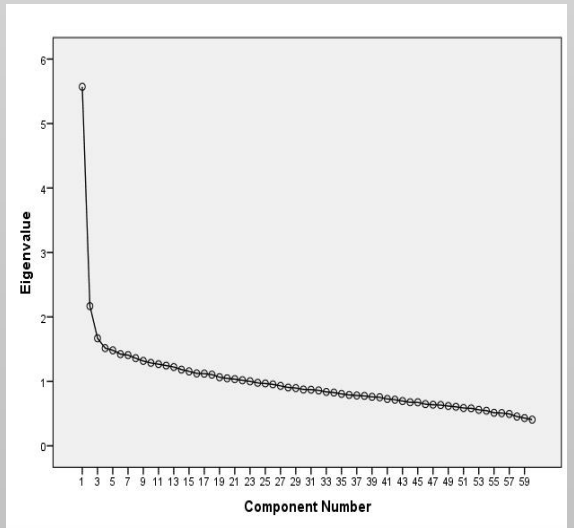

Figure 2: Scree Plot for English Language for Paper Two

Table 2 showed 23 underlying factors with eigenvalues higher than one. The total variance of the first factor (5.570) was distinctly greater than twice of the second factor (2.167). The variance of the first factor explained 9.28% and the second accounted for 3.61% of the residual variance, while the remaining variance was accounted for by the

other 27 factors. Also, the results yielded 23 factors with eigenvalues greater than one, which accounted for 56.27% of the total variance. In addition, the ratio of the first and second factor produced 2.570, higher than the required ratio 2.00. Based on these results, it was implied that there is only one underlying factor in the factor structure of the 60 English Language multiple-choice test items of paper type one. Similarly, this is an evidence that the English Language multiple-choice test items of paper type two are unidimensional in nature. The unidimensionality could also be inferred from the scree plot in Figure 2, showing a distinct factor.

Table 3: Total Variance Explained for English Language Multiple-Choice Items Paper Three

Component	Initial Eigenvalues		
	Total	% of Variance	Cum %
1	5.306	8.843	8.843
2	2.213	3.688	12.531
3	1.723	2.872	15.403
4	1.654	2.757	18.160
5	1.510	2.516	20.676
6	1.431	2.384	23.060
7	1.376	2.293	25.352
8	1.360	2.266	27.618
9	1.296	2.160	29.778
10	1.291	2.152	31.930
11	1.269	2.115	34.045
12	1.258	2.096	36.141
13	1.234	2.057	38.198
14	1.200	2.001	40.199
15	1.193	1.989	42.188
16	1.176	1.959	44.147
17	1.148	1.914	46.061
18	1.099	1.832	47.893
19	1.061	1.769	49.662
20	1.038	1.730	51.393
21	1.031	1.718	53.111
22	1.024	1.707	54.818
23	1.013	1.688	56.506

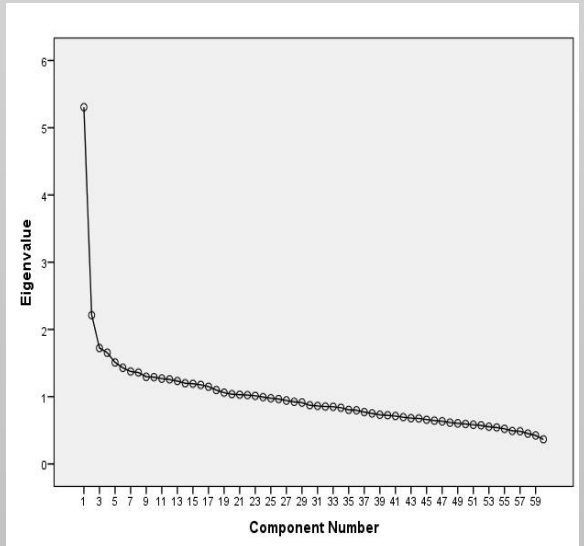

Figure 3: Scree Plot for English Language for Paper Three

Table three showed 23 underlying factors with eigenvalues higher than one. The total variance of the first factor (5.306) was distinctly greater than twice of the second factor (2.213). The variance of the first factor explained 8.84% and the second accounted for 3.69% of the residual variance, while the remaining variance was accounted for by the

other 27 factors. Also, the results yielded 23 factors with eigenvalues greater than one, which accounted for 56.51% of the total variance. In addition, the ratio of the first and second factor produced 2.398, higher than the required ratio 2.00. Based on these results, it was implied that there is only one underlying factor in the factor structure of the 60 English Language multiple-choice test items of paper type three. This is an evidence that the English Language multiple-choice test items of paper type three are unidimensional in nature. The unidimensionality could also be inferred from the scree plot in Figure 3 showing a distinct factor.

Table 4: Total Variance Explained for English Language Multiple-Choice Items Paper Four

Component	Initial Eigenvalues		
	Total	% of Variance	Cum %
1	5.113	8.521	8.521
2	2.210	3.683	12.204
3	1.657	2.761	14.965
4	1.575	2.624	17.589
5	1.481	2.468	20.057
6	1.469	2.448	22.505
7	1.400	2.333	24.839
8	1.377	2.296	27.134
9	1.341	2.235	29.370
10	1.302	2.171	31.540
11	1.272	2.119	33.660
12	1.242	2.069	35.729
13	1.227	2.045	37.774
14	1.167	1.945	39.719
15	1.165	1.941	41.660
16	1.155	1.924	43.584
17	1.140	1.900	45.485
18	1.103	1.838	47.323
19	1.073	1.789	49.112
20	1.062	1.771	50.883
21	1.059	1.765	52.648
22	1.038	1.730	54.378
23	1.015	1.692	56.070
24	1.004	1.673	57.743

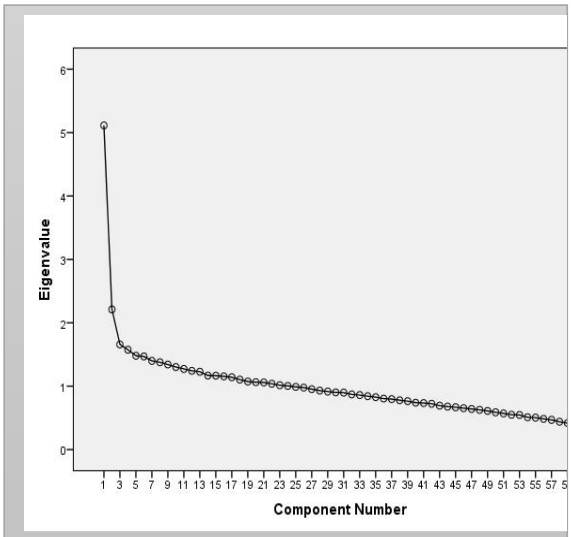

Figure 4: Scree Plot for English Language for Paper Four

Table four showed 24 underlying factors with eigenvalues higher than one. The total variance of the first factor (5.113) was distinctly greater than twice of the second factor (2.210). The variance of the first factor explained 8.52% and the second accounted for 3.68% of the residual variance, while the remaining variance was accounted for by the other 26 factors. Also, the results yielded 24 factors with eigenvalues greater than one, which accounted for 57.74% of the total variance. In addition, the ratio of the first and second factor produced 2.314, higher than the required ratio 2.00. Based on these results, it was implied that there is only one underlying factor in the factor structure of the 60 English Language multiple-choice test items of paper type one. Similarly, this is an evidence that the English Language multiple-choice test item of paper type four is unidimensional in nature. The unidimensionality could also be inferred from the scree plot in Figure 4 showing a distinct factor.

Research Question Two: What are the item difficulty, discrimination, and guessing parameters of the four different paper types of English Language multiple-choice test items?

To answer this question, the items were calibrated for parameter estimates (difficulty, discrimination, and guessing), using X-cliabre 4.2. The categorizations were done as proposed by Meyer (2014). The results are presented in Table 5.

Table 5: Summary of Item Difficulty, Discrimination and Guessing for English Language Multiple-Choice Items across Four Different Paper Types

Paper Types	(b)	No of Item/%	(a)	No of Item/%	(c)	No of Item/%
1	Easy	14 (23.3)	Excellent (a≥1.70)	6 (10.0)	0.00 – 0.25	38 (63.3)
	Moderate	25 (41.7)	Good (1.35≤a≤1.69)	14 (23.3)	0.26 – 0.40	22 (36.7)
	Difficulty	21 (35.0)	Moderate (0.65≤a≤1.34)	39 (65.0)		
			Marginal 0.35≤a≤0.64)	1 (1.7)		
			Poor (a≤0.34)	0 (0.0)		
2	Easy	10 (16.7)	Excellent (a≥1.70)	1 (1.7)	0.00 – 0.25	42 (70.0)
	Moderate	32 (53.3)	Good (1.35≤a≤1.69)	22 (36.7)	0.26 – 0.40	18 (30.0)

	Difficulty	18 (30.0)	Moderate (0.65≤a≤1.34)	36 (60.0)		
			Marginal 0.35≤a≤0.64)	0 (0.0)		
			Poor (a≤0.34)	1 (01.7)		
3	Easy	11. (18.5)	Excellent (a≥1.70)	0 (0.0)	0.00 – 0.25	41 (68.3)
	Moderate	30 (50.0)	Good (1.35≤a≤1.69)	22 (36.7)	0.26 – 0.40	19 (31.7)
	Difficulty	19 (31.7)	Moderate (0.65≤a≤1.34)	37 (61.7)		
			Marginal 0.35≤a≤0.64)	1 (1.7)		
			Poor (a≤0.34)	0 (0.0)		
4	Easy	13 (21.7)	Excellent (a≥1.70)	1 (1.7)	0.00 – 0.25	40 (66.7)
	Moderate	27 (45.0)	Good (1.35≤a≤1.69)	22 (36.7)	0.26 – 0.40	20 (33.3)
	Difficulty	20 (33.3)	Moderate (0.65≤a≤1.34)	36 (60.0)		
			Marginal 0.35≤a≤0.64)	1 (1.7)		
			Poor (a≤0.34)	0 (0.0)		

**difficulty (b), discrimination (a). and guessing (c)*

Table 5 showed item difficulty, discrimination and guessing parameters of English Language multiple-choice items across different paper types. For the 60 items in paper one: 25(41.7) and 39(65.0) fell under moderate difficulty and discrimination respectively, while 38(63.3) items were within the acceptable range (0.00 – 0.25) of guessing; for paper two, 32(53.3) and 36(60.0) fell under moderate difficulty and discrimination respectively, while 42(70.0) items were within the acceptable range (0.00 – 0.25) of guessing; for paper three, 30(50.0) and 37(61.7) fell under moderate difficulty and discrimination respectively, while 38(63.3) items were within the acceptable range (0.00 – 0.25) of guessing; lastly for paper four, 27(45.0) and 36(60.0) fell under moderate difficulty and discrimination respectively, while 40(66.7) items were within the acceptable range (0.00 – 0.25) of guessing.

Hypotheses Testing

There will be no significant difference in the item difficulty, discrimination and guesting parameter estimates of the English Language Multiple-choice tests across the four different paper types.

To test this hypothesis, the calibrated item by item difficulties, discrimination and guessing tendencies across the four different paper types were subjected to a one-way analysis of variance in order to determine if there exist significant differences. The results are presented in Table 6.

Table 6: Difference in Item Difficulty, Discrimination and Guessing Parameter Estimates of English Language Multiple-choice Test across Four Different Paper Types

Parameters	Sources	Sum of Squares	df	Mean Square	F	Sig.
a	Between Groups	.049	3	.016	.212	.888
	Within Groups	18.265	236	.077		
	Total	18.314	239			
b	Between Groups	.188	3	.063	.028	.994
	Within Groups	526.809	236	2.232		
	Total	526.996	239			
c	Between Groups	.001	3	.000	.425	.735
	Within Groups	.256	236	.001		
	Total	.257	239			

Table 6 showed the results of the one-way analysis of variance with respect item difficulty of English Language Multiple-choice test in the four different paper types. The difference is not statistical significant ($F_{(3, 239)}$=0.028; p>0.05), since the p=0.994 and higher than the significant level of 0.05. This implied that the item difficulty parameter estimates across the four different paper types are comparable. Table 5 also showed the results of the one-way analysis of variance with respect to item discrimination of English Language Multiple-choice test in the four different paper types. The difference is not statistical significant ($F_{(3, 239)}$=0.212; p>0.05), since the p=0.888 and higher than the significant level of 0.05.. This implied that the item discrimination parameter estimates across the four different paper types are comparable. Finally, Table 5 showed the results of the one-way analysis of variance with respect guessing parameter of English Language Multiple-choice test in the four different paper types. The difference is not statistically significant ($F_{(3, 239)}$=0.425; p>0.05), since the

p=0.734 and higher than the significant level of 0.05. This implied that the item guessing parameter estimates across the four different paper types are comparable.

Discussion

The study ascertained unidimensionality and item parameters of four different paper types of English Language multiple-choice test of distance learning of the Obafemi Awolowo University. The results from research question one that investigated unidimensionality of the English Language multiple-choice items across the four different paper types showed that all the four paper types were unidimensional. In each of the paper types, the eigenvalue of the first factor was distinctly higher than the eigenvalue of the second factor, as contained in the principal component analyses for the four different paper types. The findings corroborated that of Reckase (1999); Orlando, Sherbouve, and Thissen (2001); and Ajeigbe and Afolabi (2014), where the eigenvalue of the first factor was substantively greater than the next, the factor structure as evident of complying to the assumption of unidimensionality of item response theory. The results relating to the item difficulty, discrimination and guessing parameter of the English Language multiple-choice items across the four different paper types had greater numbers of the items under moderate difficulty and discrimination values across different paper types. This means that there was no significant variation in the difficulty and discrimination across the four paper types. The results are tangential with the findings of Hambleton (2000); and Kamata and Vaughn (2014) which showed higher significant relationship between item difficulty and discrimination. In addition, well above average of the items were within the acceptable range of guessing value across the four different paper types. Finally, the results of the hypotheses found that: firstly, the difficulty parameter is stable across the four types of the English Language Multiple-choice. Secondly, the discrimination parameter is stable across the four types of the English Language Multiple-choice. Lastly, the guessing parameter is stable across the four types of the English Language Multiple-choice. The results obtained are corroborated with findings of Courville (2004), where similar parameter estimates were recorded in mathematics achievement test under different measurement framework. Furthermore, the use of three parameter model (difficulty, discrimination and guessing) showed the parameter indices were stable, as this result was supported by DeAyala (2012), who reported comparable and stable item parameter indices across different item subsets, using item response theory approach.

Conclusion

The study concluded that each of the four different paper types of English Language multiple-choice tests did not violate the unidimensionality of item response theory. It also concluded that the item calibrations, in terms of difficulty, discrimination and guessing parameter, are stable and comparable across paper types.

Recommendations

Based on the results of the study, the following recommendations were put forward;

- Test developers should ensure that the assumption of unidimensionality for using item response theory should be ascertained.
- Test developers should consistently compare the stability of the parameter estimates in terms of its difficulty, discrimination and guessing parameter, so that some groups of students will not be at advantage over others because of the question type they answered.
- The management should consider establishing a testing and psychometric unit that will be in charge of item calibration for proper documentations.
- Item banks can be encouraged after establishing or ascertaining item parameter estimates of different paper types for future use.
- The assumption of unidimensionality of the IRT and parameter estimates could be replicated in other subjects or disciplines across different paper types.

References

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), 30-37.

Akinyoola, O. A. (2019). Historical Evaluation of JAMB and its Efficiency in Educational Development in Nigeria, 1978-2019. *Science and Education Development Institute.* 11 (1): 11 – 23.

Baghaei, P., & Amrahi, N. (2011). The Effect of the Number of Options on the Psychometric Characteristics of Multiple-choice Items. *Psychological test and assessment modeling*, 53(2), 192-211. https://doi.org/10.4304/jltr.2.5.1052-1060.

Bond, T.G. and C. M. Fox (2001).Applying the Rasch model. NJ: Lawrence Erlbaum. Crocker, L., and Algina, J. (1986). *Introduction to Classical and Modern Test Theory.* NewYork: CBS College Publishing.

Courville, T. G. (2004). An empirical comparison of item response theory and classical test theory item/person statistics. Unpublished doctoral dissertation, Texas A&M University. http://txspace.tamu.edu/bitstream/handle/1969.1/1064/etd-tamu-2004B-EPSY-Courville-2.pdf?sequence=1

DeAyala, R. J., (2012). *The theory and practice of item response theory*. New York: The Guilford Press.

Drasgow, F., & Parsons, C. K. (1983). Application of unidimensional item response theory models to multi-dimensional data. *Applied Psychological Measurement*,7, 189-199.

Haladyna, T. M. (2004). Developing and Validating Multiple-choice Test Items. Lawrence Erlbaum Associates Publishers, Mahwah, New Jersey, USA. 25. https://doi.org/10.4324/9780203825945.

Hambleton, R. K., (2000). *Item response theory principles and applications*. Boston, MA: Kluwer-Nijhoff Publishing.

Ibrahim, A., (2017). Empirical comparison of three methods for detecting differential item functioning in dichotomous test items. *Journal of Teaching and Teacher Education*, 5 (1), 1-18.

Kamata, M., & Vaughn, J. (2014). *Cognitive diagnostic assessment in education: Theory and applications.* Cambridge University Press.

Lord, F. M. and Novick, M. R. (1968). *Statistical Theories of Mental Test Scores.* Reading, Massachusetts: Addison-Wesley Publishing Company.

Meyer, J.P. (2014) Applied Measurement with jMetrik. Published by Routledge 55 B/W Illustrations ISBN 9780415531979.

Moss, P. A. (1995). Themes and variation in validity theory. Educational measurement: Issues and Practice, 14(4), 5-13. https://doi.org/10.1111/j.1745-3992.1995.tb00854.x.

Ojerinde, D, Popoola, K., Ojo, F., Onyeneho, P. (2012): *Introduction to Item Response Theory: Parameter Models, Estimation and Application*, Lagos, Nigeria, Goshen Print Media Limited.

Ojerinde, A., Popoola, O., and Ajeigbe, T. O. (2020). *Application of Item Response Theory in Educational Assessment in Africa*. Lambert Academic Publishing.

Orlando, M., Sherbourve, C. D. & Thissen, D (2001). Summed-score linking using Item Response Theory: Application to depression measured. *Psychological Assessment*, 12(3), 354 – 359.

Reckase, M. D. (1999). Uni-factor latent trait models applied to multi-factor tests: Results and implications. *Journal of Educational Statistics*, 4, 207-230.

Sheng, Y. (2005). Bayesian Analysis of Hierarchical IRT models: Comparing and Combing the Unidimensional and Multidimensional IRT modes. *Unpublished Doctoral Dissertation. University of Missouri-Columbia.*

Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, 55, 293-325.

Van der Linden, W. J. and Hambleton, R. K. editors. Handbook of modern item response theory. Springer, New York, 1997.