INFLUENCE OF PRIMACY, MIDDLE AND RECENCY BIASES ON THE PSYCHOMETRIC CHARACTERISTICS OF THE NATIONAL EXAMINATIONS COUNCIL MULTIPLE-CHOICE MATHEMATICS ITEMS

[1]AYANLADE, O. A. & [2]AJEIGBE, T. O.
[1&2]Department of Educational Foundations & Counselling, Faculty of Education
Obafemi Awolowo University, Ile-Ife
ayanladeolajide2016@yahoo.com, taiaje@oauife.edu.ng, ajetaiolu@gmail.com

Abstract
The study determined the difficulty and discrimination of National Examination Councils (NECO) Multiple-choice Mathematics items from 2014 to 2016. The study also investigated influence of primacy, middle and recency biases on the psychometric characteristics (difficulty and discrimination) of NECO Multiple-choice Mathematics items from 2014 to 2016. These were with a view to providing statistical evidence on how the three biases consistently affect item difficulty and discrimination of NECO Multiple-choice Mathematics items across different population sets within the period of three consecutive years. The study adopted the ex-post facto research design. The population for the study comprised 53,986 Senior Secondary School Students (SS III) who registered and sat for the National Examinations Council Mathematics paper III from 2014 to 2016 in Osun state. A sample of 2,311 candidates was selected using multi-stage sampling procedure. Two Local Government Areas (LGAs) were selected from each of the three senatorial districts of the State, using simple random sampling techniques. From each of the selected LGAs, two schools were selected using simple random sampling technique. An intact class of senior secondary students (SS III) was used from each of the 12 selected schools. The instrument used for the study are the NECO Mathematics questions for the three years. Students' corresponding responses as contained in the Optical Mark Reader (OMR) were scored dichotomously using Keys collected from NECO and calibrated under R software to generate difficulty and discrimination indices. The results of the study indicated that the item difficulty and discrimination parameters of Multiple-choice Mathematics items across the three years. For year 2014: 52(86.7) and 43(71.7) fell under moderate difficulty and discrimination respectively; for year 2015, 57(95.0) and 34(56.7) fell under moderate difficulty and discrimination respectively; lastly for year 2016, 42(70.0) and 12(20.7) fell under moderate difficulty and discrimination respectively. Finally, the result also showed that the primacy, middle and recency biases has no significant influence on difficulty and discrimination levels of National Examination Council (NECO) Mathematics paper III in the year2014 (For the difficulty; $F_{(2,57)}$.=0. 412; p>0.05 and For the discrimination; $F_{(2,57)}$.= 0.739; p>0.05) and 2016(For the difficulty; $F_{(2,57)}$.= 0.023; p>0.05 and For the discrimination; $F_{(2,57)}$.= 0.275; p>0.05). And the result further

showed that the primacy, middle and recency biases also has no significant influence on difficulty level whilst there is a significant influence in terms of discrimination of NECO Multiple-choice Mathematics items in 2015(For the difficulty; $F_{(2,57)}$.= 0.539; p>0.05 and For the discrimination; $F_{(2,57)}$.= 6.409; p<0.05). It is concluded that difficulty and discrimination indices of the National Examination Council Multiple-choice Mathematics items are stable and not affected by primacy, middle and recency biases across the three years that were reviewed.

Keywords: Primacy, Middle, Recency, Psychometric Characteristics, Multiple-choice and Biases

Introduction

Assessment refers to a related series of measures used to determine a complex attribute of an individual or group of individuals. This entails gathering and interpreting data on a student's level of achievement of learning objectives. Assessments are also used to identify particular student problems and strengths, allowing educators to provide specialized academic support, educational programs, or social services to students. Also, assessments are created by a diverse range of organizations and individuals, including teachers, district administrators, colleges, private corporations, state education departments, and groups that contain a combination of these individuals and institutions. Assessment is crucial because it motivates pupils to study. Most students, whether they like it or not, tend to focus their concentration on the best or most expedient manner to pass their 'tests.' We may utilize our evaluation strategies to control the types of learning that take place based on this knowledge. Assessment procedures that focus primarily on knowledge recall, for example, are prone to promote superficial learning. On the other side, we are more likely to see higher levels of student performance or achievement if we use evaluation procedures that require critical thinking or creative problem solving. As previously said, one of the purposes of evaluation is to motivate and direct learning. Assessment procedures that are well-designed play an important role in educational decision-making and are an important part of continuing quality improvement activities at the lesson, course, and/or curriculum levels.

Multiple-choice assessment is unquestionably one of the most durable and effective forms of instructional technology still in use today. Fredrick J. Kelly is often cited as the developer of the multiple-choice item format (Rogers, 1995). In 1916, he published the *Kansas Silent Reading Test* in the *Journal of Educational Psychology* where students who wrote the test were required to circle the correct answer rather than writing their answer for each item. The multiple choice item formats were an important breakthrough in educational testing because it served as an objectively scored task that used a structured-response format where the student was presented with one correct option and two or more incorrect options or distractors. The task was to select the correct option.

Multiple-choice items are commonly employed in educational testing because they allow for the direct measurement of a wide range of knowledge, skills, and competences across a wide range of disciplines and curriculum areas, such as the ability to understand concepts and principles, make decisions, infer conclusions, and reason statements must be completed, data must be interpreted, and information must be applied. Multiple-choice items are efficient to administer, they are easy to score objectively, and they can be used to sample a wide range of content domains in a relatively short time using a single test administration (Haladyna & Rodriguez, 2013; Rodriguez, 2016). Compared with essays and other constructed-response tasks, which are prone to subjective scoring and require more time for recording answers, multiple-choice items can be scored more accurately, and they require students to spend less time on recording answers (Haladyna, 2004). Because of these significant advantages, multiple-choice testing is regarded as a cost-effective method of educational assessment.

Primacy, Middle and Recency biases refers to the biases that occurs in positioning the correct key in the first, middle and the last options by the examiners whilst making the multiple-choice items. In educational measures, there are two basic frameworks for achieving excellent test items. These are Classical Test Theory (CTT) and Item Response Theory (IRT). The Classical Test Theory comprises three concepts. These are: test (observed) score, true score and error score. According to Hambleton and Jones (1993), within these three concepts, several models have been formulated, of which the central model is the "Classical test model". This model links the observed test score (X) to the sum of the two unobserved (or often called latent) variables, true score (T) and error score (E). Mathematically, the Classical test model is represented by $X = T + E$ The equation has two unknowns (the true score (T) and error score (E)), thereby making it not solvable. However, according to Hambleton and Jones (1993), the use of the Classical test model represented by $X = T + E$ is made possible by three assumptions. These are: {a} True score (T) and error score (E) are uncorrelated {b} the average error score in the population of examinees is zero and {c} Error scores on parallel tests are uncorrelated. Thus, under the Classical Test Theory, the examinee's test score would be the sum of the scores received on all the items in the test. This, according to Tomkowickz and Wright (2007), is referred to as number-correct scoring.

This method of scoring produces maximum likelihood trait estimates based on raw scores (that is, total number of correctly answered items). In this method, examinees who answer correctly the same number of items, irrespective of the items' level of difficulties and discriminations, earn the same scale score. Thus, the nature of the items parameters (that is, difficulty and discrimination levels) are not considered in the scoring of examinees' performance. The importance of item parameters estimation in test development cannot be ignored. In fact, difficulty and discrimination indices are statistics that guide test development (Ayanwale, 2017). Hence, it is however imperative to empirically investigate whether the item response theory parameter estimates such as difficulty, discrimination estimated using various

IRT models and local dependence and differential items functioning indices will be sensitive to the positioning of the correct keys option in the set of the alternatives answers.

Biases in psychometrics characteristics can cause measurement errors in students' scores (limited feedback to correct errors in student understanding, results can be skewed by reading ability or test-wiseness, developing good items takes time, and measuring ability to organize and express ideas is impossible). Other possible biases that test developers may not have considered include primacy, middle, and recency biases. However, it is also unknown whether items with keys in the central positions have any impact or effect on item difficulty and item discrimination of Multiple-Choice test items compared to those with keys in the first or last positions in both the four and five option lengths of multiple-choice tests. It is also unknown to what extent these biases could cause variations in psychometrics characteristics of National Examinations Council (NECO) Mathematics multiple-choice items.

The main purpose of the study was to assess the influence of primacy, middle and recency biases on the psychometric characteristics of the national examinations council mathematics multiple-choice items. Specifically, an attempt was made to:

i. determine item difficulty and discrimination of NECO Mathematics Multiple-choice items from 2014 to 2016 in Osun state; and
ii. investigate influence of primacy, middle and recency biases on the psychometric characteristics (difficulty and discrimination) of NECO Mathematics Multiple-choice items from 2014 to 2016.

Research Question

Question One: What are the difficulty and discrimination index of NECO Multiple-choice Mathematics items from 2014 to 2016 in Osun state?

Research Hypothesis

Hypothesis One: There is no significant influence of primacy, middle and recency biases on the psychometric characteristics (difficulty and discrimination) of NECO Multiple-choice Mathematics items from 2014 to 2016.

Methodology

This study used the ex-post factor research design. It is an investigation, according to Nwogu (2006), in which objects or individuals are gathered, compiled, and analyzed with the intention of informing decision-making based on a perception that they are typical of the entire group. This design is used since it satisfies the requirements of this study for an efficient analysis. It is used to gather information about and describe in a methodical way the traits, traits, and details of the study's population. This design's strength is that it allows for thorough explanation of pertinent factors in connection to the provided population. The description of events as they actually occur is important to this survey study design. It is only used to gather

and evaluate data. It does not involve manipulation of information.
The population for the study comprised 53,986 Senior Secondary School Students (SS III) who registered and sat for the National Examinations Council Mathematics paper III from 2014 to 2016 in Osun state. A sample of 2,311 candidates was selected using multi-stage sampling procedure. Two Local Government Areas (LGAs) were selected from each of the three senatorial districts of the State, using simple random sampling techniques. From each of the selected LGAs, two schools were selected using simple random sampling technique. An intact class of senior secondary students (SS III) was used from each of the 12 selected schools.
The research instrument for the study titled Pro-forma of Optical Mark Reader (OMR) Sheet that contain Students' responses from the year 2014, 2015 and 2016 respectively of the National Examination Council (NECO) Mathematics paper III was used for this study. The instrument was a standardized test developed by experts in NECO as such there was no need for validation and reliability because the test is already valid and reliable.
Data was extracted from the data-base of the National Examination Council (NECO) Mathematics paper III Optical Mark Reader (OMR) sheet for the year 2014, 2015 and 2016. The extraction contains all details of candidates including their center number, registration number, sex, responses and scores.
Data collected was analyzed based on the research questions raised in the study. Data collected was subjected to analysis using SPSS to transform the extracted Optical Mark Responses (OMR) through Z-score and T-score.

Results
Research question
What are the difficulty and discrimination of NECO Multiple-choice Mathematics items from 2014 to 2016 in Osun state?

Table 1: Contingency Table showing the difficulty and discrimination of NECO Multiple-choice Mathematics items

	2016				2015				2014			
Model	Diff	R	Disc	R	Diff	R	Disc	R	Diff	R	Disc	R
Item1	50.00328	D	0.099777	G	-0.22481	M	1.193402	P	-0.22617	M	1.341041	P
Item2	1.459551	D	0.535189	G	-0.71417	M	0.714217	G	-0.23932	M	0.765968	G
Item3	-0.54167	M	0.706836	G	-0.46653	M	0.882725	G	-0.72843	M	1.25376	P
Item4	-0.5112	M	0.558035	G	-0.1091	M	0.658197	G	-0.80971	M	1.105436	P
Item5	-0.26422	M	1.371275	P	-0.44679	M	1.697053	P	0.24015	M	0.891548	P
Item6	0.223274	M	1.382429	P	0.005101	M	0.762173	G	0.274646	M	0.937061	P
Item7	-0.95523	M	1.948702	P	-0.86675	M	1.430311	P	-0.55983	M	1.284273	P
Item8	-3.08844	E	0.26205	G	-0.96006	M	0.360828	G	0.315026	M	0.608155	G
Item9	-0.41011	M	1.668093	P	-0.57046	M	1.283177	P	-0.38912	M	1.00126	P
Item10	0.816911	M	1.119387	P	0.047376	M	1.447324	P	0.18976	M	0.905669	P
Item11	-0.73095	M	3.032252	P	-0.67724	M	1.293366	P	1.124969	D	0.733977	G
Item12	-0.61635	M	2.660975	P	0.292209	M	0.859008	G	0.202101	M	0.774055	G
Item13	-59.1398	E	-0.01463	G	-0.41611	M	1.270622	P	0.606315	M	0.851904	P
Item14	-0.72856	M	2.900345	P	-0.42674	M	1.961632	P	-0.23529	M	1.068732	P
Item15	4.995394	D	0.258736	G	0.155338	M	1.124167	P	-0.26803	M	1.277931	P
Item16	-0.77173	M	2.927988	P	-0.54956	M	1.519576	P	0.420099	M	0.780029	G
Item17	-0.8691	M	2.584257	P	-0.47707	M	1.517382	P	-0.11488	M	0.618653	G
Item18	0.181176	M	0.703397	G	-0.35331	M	1.496393	P	0.944373	M	0.613824	G
Item19	-1.42182	E	1.431331	P	-0.26757	M	1.31037	P	1.074681	E	0.866962	P
Item20	-0.27485	M	1.882705	P	-0.39255	M	1.503916	P	0.704748	M	0.989717	P
Item21	-0.79772	M	1.954762	P	-0.66321	M	1.957416	P	0.280123	M	1.14272	P
Item22	-1.30095	E	1.31091	P	-0.98656	M	1.334468	P	0.338905	M	0.933832	P
Item23	-2.39505	E	-0.47603	G	-0.43512	M	0.427195	G	-0.05838	M	0.865681	P
Item24	0.482624	M	1.695894	P	1.376838	D	0.423527	G	-0.15116	M	2.022475	P
Item25	-1.04735	E	1.793308	P	-0.7058	M	1.460126	P	-0.04731	M	1.13327	P
Item26	-0.86685	M	1.803855	P	-0.27549	M	1.128354	P	0.393538	M	1.05219	P
Item27	-0.21987	M	1.206588	P	0.384454	M	1.02495	P	0.183218	M	0.884425	P
Item28	-0.22047	M	1.40758	P	0.085833	M	1.15396	P	0.314088	M	0.685256	G
Item29	-0.32849	M	2.292141	P	14.49381	D	0.034357	G	-0.29103	M	1.341443	P
Item30	-0.44999	M	1.904877	P	-0.32854	M	1.807873	P	0.296828	M	0.802477	G

Source: Author's Analysis, 2022

NB: Diff-Difficulty, Disc-Discrimination, R-Remarks, P-Poor, G-Good, E-Easy Item, M-Moderately Difficult Item, D-Difficult Item

Table 2: Summary of Item Difficulty and Discrimination for Mathematics Multiple-Choice Items across Three years.

Years	(b)	No of Item/%	(a)	No of Item/%
2014	Easy	4 (6.7)	Excellent (a=1.70)	2 (3.3)
	Moderate	52 (86.7)	Good (1.35=a=1.69)	5 (8.3)
	Difficulty	4 (6.7)	Moderate (0.65=a=1.34)	43 (71.7)
			Marginal 0.35=a=0.64)	9 (15.0)
			Poor (a=0.34)	1 (1.7)
2015	Easy	1 (1.7)	Excellent (a=1.70)	4 (6.7)
	Moderate	57 (95.0)	Good (1.35=a=1.69)	12 (20.0)
	Difficulty	2 (3.3)	Moderate (0.65=a=1.34)	34 (56.7)
			Marginal 0.35=a=0.64)	8 (13.3)
			Poor (a=0.34)	2 (3.3)
2016	Easy	10 (16.7)	Excellent (a=1.70)	22 (36.7)
	Moderate	42 (70.0)	Good (1.35=a=1.69)	9 (15.0)
	Difficulty	8 (13.3)	Moderate (0.65=a=1.34)	12 (20.7)
			Marginal 0.35=a=0.64)	7 (11.7)
			Poor (a=0.34)	10 (16.7)

**difficulty (b) and discrimination (a)*

Table 2 showed item difficulty and discrimination parameters of Multiple-choice Mathematics items across different Years. For year 2014: 52(86.7%) and 43(71.7%) fell under moderate difficulty and discrimination respectively; for year 2015, 57(95.0%) and 34(56.7%) fell under moderate difficulty and discrimination respectively; lastly for year2016, 42(70.0%) and 12(20.7%) fell under moderate difficulty and discrimination respectively.

Hypothesis One: There is no significant influence of primacy, middle and recency biases on the psychometric characteristics (difficulty and discrimination) of NECO Multiple-choice Mathematics items from 2014 to 2016.

Table 3: One-way Analysis of Variance (ANOVA) of the Relative Influence of Primacy, Middle and Recency Biases on the Psychometric Characteristics (Difficulty and Discrimination) of 2014 NECO Multiple-Choice Mathematics Items

		SS	df	MS	F	Sig.
DIFF	Between Groups	0.598	2	0.299	0.412	0.066
	Within Groups	41.315	57	0.725		
	Total	41.912	59			
DISC	Between Groups	0.194	2	0.970	0.739	0.482
	Within Groups	7.472	57	0.131		
	Total	7.666	59			

Note: SS = Sum of Square, df = Degree of Freedom, MS = Mean Square

Table 3 showed the results of one-way analysis of variance obtained based on the influence of primacy, middle and recency biases on the difficulty and discrimination of the 2014 NECO Multiple-choice Mathematics test. For the difficulty; $F_{(2,57)}$.=0. 412; $p>0.05$ was obtained. This implies that there was no statistical significant difference in the item difficulty of the Multiple-choice Mathematics test items with reference to primacy, middle and recency biases. For the discrimination; $F_{(2,57)}$.= 0.739; $p>0.05$ was obtained. This implies that there was no statistical significant difference in the item discrimination of the Multiple-choice Mathematics test items with reference to primacy, middle and recency biases. Since the p-value of 0.066 and 0.482 were obtained for the difficulty and discrimination respectively, which is greater than 0.05 level of significant, hence, the hypothesis that there is no significant influence of primacy, middle and recency biases on the psychometric characteristics (difficulty and discrimination) of NECO Multiple-choice Mathematics items 2014 is hereby accepted.

Table 4: One-way Analysis of Variance (ANOVA) of the Relative Influence of Primacy, Middle and Recency Biases on the Psychometric Characteristics (Difficulty and Discrimination) of 2015 NECO Multiple-Choice Mathematics Items

		SS	df	MS	F	Sig.
DIFF	Between Groups	22.637	2	11.318	0.539	0.587
	Within Groups	1198.0	57	21.018		
	Total	1220.6	59			
DISC	Between Groups	2.109	2	1.054	6.409	0.003
	Within Groups	9.378	57	0.165		
	Total	11.487	59			

Note: SS = Sum of Square, df = Degree of Freedom, MS = Mean Square

Table 4 showed the results of one-way analysis of variance obtained based on the influence of primacy, middle and recency biases on the difficulty and discrimination of the 2015 NECO Multiple-choice Mathematics test. For the difficulty; $F_{(2,57)}$.= 0.539; $p>0.05$ was obtained. This implies that there was no statistical significant difference in the item difficulty of the Multiple-choice Mathematics test items with reference to primacy, middle and recency biases. For the discrimination; $F_{(2,57)}$.= 6.409; $p<0.05$ was obtained. This implies that there was statistical significant difference in the item discrimination of the 2015 Multiple-choice Mathematics test items with reference to primacy, middle and recency biases.

Table 5: One-way Analysis of Variance (ANOVA) of the Relative Influence of Primacy, Middle and Recency Biases on the Psychometric Characteristics (Difficulty and Discrimination) of 2016 NECO Multiple-Choice Mathematics Items

		SS	df	MS	F	Sig.
DIFF	Between Groups	4.908	2	2.454	0.023	0.977
	Within Groups	6117.60	57	107.33		
	Total	6122.51	59			
DISC	Between Groups	0.483	2	0.242	0.275	0.760
	Within Groups	49.999	57	0.877		
	Total	50.482	59			

Note: SS = Sum of Square, df = Degree of Freedom, MS = Mean Square

Table 5 showed the results of one-way analysis of variance obtained based on the influence of primacy, middle and recency biases on the difficulty and discrimination of the 2016 NECO Multiple-choice Mathematics test. For the difficulty; $F_{(2,57)}$.= 0.023; $p>0.05$ was obtained. This implies that there was no statistical significant difference in the item difficulty of the Multiple-choice Mathematics test items with reference to primacy, middle and recency biases. For the discrimination; $F_{(2,57)}$.=

0.275; p>0.05 was obtained. This implies that there was no statistical significant difference in the item discrimination of the Multiple-choice Mathematics test items with reference to primacy, middle and recency biases. Since the p-value of 0.977 and 0.760 were obtained for the difficulty and discrimination respectively, which is greater than 0.05 level of significant, hence, the hypothesis that there is no significant influence of primacy, middle and recency biases on the psychometric characteristics (difficulty and discrimination) of NECO Mathematics paper III Multiple-choice items 2016 is hereby accepted.

Discussion of Findings

The results of the research question one of this study which estimated the difficulty and discrimination parameters for the Optical Mark Record (OMR) sheet extracted from the National Examination Council (NECO) Mathematics paper III in the year 2016, 2015 and 2014 respectively; item difficulty and discrimination parameters of Multiple-choice Mathematics items across different Years. For year 2014: 52(86.7%) and 43(71.7%) fell under moderate difficulty and discrimination respectively; for year 2015, 57(95.0%) and 34(56.7%) fell under moderate difficulty and discrimination respectively; lastly for year 2016, 42(70.0%) and 12(20.7%) fell under moderate difficulty and discrimination respectively.

According to the past study, Brown (2004) proposes that difficulty level refers to the ease with which a multiple-choice item can be completed. It is critical to ensure that a test item has an appropriate difficulty level because this will have a substantial impact on the test's capacity to gauge the genuine ability of the test takers. The study concluded that the majority of the test items were moderately challenging, while some were easy, and only a few were considered extremely difficult. According to the study, the difficulty level distribution on test items was appropriate because it did not place undue burden on low-achieving students or demotivate high-achieving students. In terms of discriminating power, the study discovered that the test items lacked the necessary discriminating power.

The results of the research question two shows that there is no significant influence of primacy, middle and recency biases on the psychometric characteristics (difficulty and discrimination) of NECO Multiple-choice Mathematics items in 2014. In 2015, there is no significant influence of primacy, middle and recency biases on the psychometric characteristics in terms of difficulty, whilst there is significant influence of primacy, middle and recency biases on the psychometric characteristics in terms of discrimination of NECO Multiple-choice Mathematics items. And in 2016, there is no significant influence of primacy, middle and recency biases on the psychometric characteristics (difficulty and discrimination) of NECO Multiple-choice Mathematics items. However, the finding of this study revealed that the primacy, middle and recency biases has no statistical significant influence on difficulty and discrimination levels of National Examination Council (NECO) Mathematics paper III in the year 2014(For the difficulty; $F_{(2,57)}$.=0. 412; p>0.05 and For the discrimination; $F_{(2,57)}$.= 0.739; p>0.05) and 2016(For the difficulty; $F_{(2,57)}$.=

0.023; p>0.05 and For the discrimination; $F_{(2,57)}$.= 0.275; p>0.05). And the result further showed that the primacy, middle and recency biases also has no statistical significant influence on difficulty level whilst there is statistical significant influence in terms of discrimination of NECO Multiple-choice Mathematics items in 2015(For the difficulty; $F_{(2,57)}$.= 0.539; p>0.05 and For the discrimination; $F_{(2,57)}$.= 6.409; p<0.05), (Ayanlade 2022).

Conclusion

The study therefore concluded that the psychometric characteristics of National Examination Council Multiple-choice Mathematics items were half-accurate and can adequately measure the student's ability in Mathematics. The National Examination Council Multiple-choice Mathematics items were reliable and valid to a certain extent but needs to improve in construction of high standard items in terms of discrimination index. Based on these, it was recommended that teachers should consistently test the ability of their students to ensure they know the materials for each subject very well, and the test experts and developers should consider the sensitivity of key location while generating an item. Besides, further research can also be carried out to investigate the relative influence of primacy, middle, and recency biases on the psychometric properties of other examination bodies' subjects.

References

Ayanwale, M.A. (2017). Efficacy of Item Response Theory in score ranking and concurrent validity of dichotomous and polytomous response mathematics achievement test in Osun State, Nigeria. Unpublished ph.D thesis. Institute of Education. University of Ibadan.

Brown, H. D. (2004). Language Assessment: Principles and Classroom Practices. Pearson Education

Haladyna, T. M. (2004). Developing and validating multiple-choice test items. New York: Lawrence Erlbaum Associates.

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. Mahwah, NJ: Lawrence Erlbaum.

Haladyna, T. M., & *Rodriguez*, M. C. (*2013*). Developing and validating test items. New York Taylor & Francis. www.taylorfrancis.com5

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Education Measurement: Issues and Practice, 12*(3), 38-47.

Nworgu, B. G. & Agah J. J. (2006). Application of three parameters logistic model in the calibration of a Mathematics achievement test. *Journal of Educational Assessment in Africa, 7: 162-172.*

Rogers, T. B. (1995). *The psychological testing enterprise: An introduction*. Belmont,CA: Wadsworth.

Tomkowickz, J.T. & Wright, K.R. (2007*). Investigation of the effect of test equating*

and scoring methods on item parameter estimates and student ability scores. A Paper presented at the annual conference of American Educational Research Association, Chicago, April 10.