OPEN ACCESS
Journal of Advances in Education and Philosophy
Abbreviated Key Title: J Adv Educ Philos
ISSN 2523-2665 (Print) |ISSN 2523-2223 (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: https://saudijournals.com

Original Research Article

Guessing Parameters and Multiple-Choice Test Items of 2016 - 2017 National Examination Council (NECO) Mathematics Examination Instruments in Calabar Education Zone of Cross River State

Asuquo, Umo Bassey[1*], Ekon, Samuel Jacob[1] and Chigbu, Nelly Nobel[1]

[1]Measurement and Evaluation Unit, Department of Science Education, Faculty of Education, Ebonyi State University, Abakaliki, Nigeria

DOI: 10.36348/jaep.2022.v06i09.007 | **Received:** 27.08.2022 | **Accepted:** 22.09.2022 | **Published:** 25.09.2022

***Corresponding author:** Asuquo Umo Bassey
Measurement and Evaluation Unit, Department of Science Education, Faculty of Education, Ebonyi State University, Abakaliki, Nigeria

Abstract

This study this study therefore examines guessing parameters and 2016 and 2017 NECO Mathematics multiple choice test items in Calabar education zone. The research objectives specifically compared the guessing parameter of the 2016 and 2017 NECO Mathematics multiple-choice test items; and also, compared the measurement theories and year of examination on reliability coefficient of 2016 and 2017 NECO Mathematics multiple-choice test items in Calabar education zone, Cross River State. To guide the study two research questions and hypotheses were further formulated and tested at 0.05 level of significance. The instrumentation research design was adopted for the study. One thousand three hundred and fifty-one (1,351) students were selected using the stratified random sampling procedure. Two instruments, namely the 2016 and 2017 NECO mathematics multiple choice tests were used for data collection. Using ANOVA and correlated t-test to analyze data collected from the field, the result reveals that guessing parameter and reliability indices of the 2016 and 2017 NECO mathematics multiple choice tests were significantly different. It was concluded that guessing parameters differed significantly with the years of study and also, there was a significant influence of measurement theory and year of test on item reliability on NECO mathematics examination papers significant difference exists between psychometric properties using the different measurement theories. Based on these findings, it was recommended that NECO should use the item response theory (IRT) in assessment for its relative merit over the CTT.

Keywords: Guessing Parameters, Multiple-Choice Test Items, item response theory (IRT), Mathematics Examination Instruments, Calabar Education Zone.

INTRODUCTION

All over the world, education is recognized as an essential tool for development. One of the ways of ascertaining the level of development in any nation is to assess the outcome of her education policies either in terms of its product from the educational sector or other assessment parameters. The most reliable channel of assessing the product is through evaluation. According to Enyi (2002), evaluation is concerned with determining the extent to which educational objectives and desired behavioural changes in learners have been attained, as well as making value judgments on the worth of the attributes. Capper (1996) supported the definition by Enyi (2002) when he said that, better test means better teaching and better teaching means, better learning. Evaluation can be done in the form of tests.

Sidhu (2005) defined test as an examination to reveal the relative standing of individuals in the group with respect to intelligence, personality, aptitude or achievement. Examination is administered to the testee for determining the extent to which he/she has attained previously identified objectives in a learning situation. The objectives here may be based on cognitive achievement, attitude, interest, personality, social adjustment or psychomotor skills. Examination is judged worthwhile when it possesses acceptable psychometric properties. In simple terms psychometric properties refer to the reliability and validity of instrument. Kline (2000) defined psychometrics as

those aspects of psychology that are concerned with psychological testing. Psychometrics is a field of study concerned with the theory and technique of psychological measurement. One part of the field is concerned with the objective measurement of skills and knowledge, abilities, attitudes, personality traits and educational achievement, the other part of the field is concerned with statistical research bearing on measurement theory (example item response theory, interclass correlation). As a result of these focuses, psychometric research involved two major tasks: the construction of instruments and the development of procedures for measurement (Thurstone, 1959).

Testing for proficiency dates back to 2200 B. C., when the Chinese emperor used grueling tests to asses fitness for office. Modern psychometrics dates to Sir Francis Galton (1822 – 1911). Charles Darwin's cousin interested in individual differences and their distribution (1884 – 1890) tested 17,000 individuals and demonstrated that objective tests could provide meaningful scores. Clark Wissler (Cattell's student) did the first basic validational research, examining the relation between the old 'mental test' scores and academic achievement. Alfred Binet (1905) introduced the first modern intelligence test, which directly tested higher psychological processes. The origin of psychometrics also has connections to the related field of psychophysics (Michell, 1999).

The psychometric properties that every measuring instrument should possess are validity and reliability. They are the characteristics of a test and other measures that identify and describe the attributes of an instrument. Psychometric properties are not statistics per se, but they are generally represented by quantitative values. These values are often calculated using statistical procedures. Some common psychometric properties of a test are item difficulty, item discrimination, the option distraction power all from the CTT perspective and the item location, the "a" parameter and the guessing parameter from the IRT perspective. Okoye (1996) defined validity as the extent to which a test measures what it intends to measure. Enyi (2002) reminded that if a test possesses other qualities but lacks validity then, it will be considered not being useful. Reliability of a test refers to the degree to which a test measures accurately and consistently yielding comparable results when administered a number of time (Akuezeulo & Aju, 2003). The test of validity and reliability can be ascertained through item analysis.

The quality of test items in any public examination is always examined through the item analysis of examinees responses. Nwoabia (1990) defines item analysis as a process which examines student's responses to individual test items inorder to assess the quality of those items. It is concerned with ascertaining the worth of test items; it helps to improve both items and the test by revising and discarding ineffective items. Item analysis usually calls for the computation of some indices such as the difficulty index, discrimination index and item distractors under the Classical Test Theory (CTT) and the a, b, and c parameter using the item characteristic curve and the test information curve in item response theory (IRT). A test can be studied from different angles and items in the test can be evaluated using the classical test theory and the item response theory. CTT was originally the leading framework for analyzing and developing standardize tests.

In order to achieve the conduct of valid and credible examinations, independent examination bodies were established, these include: The West African Examinations Council (WAEC), The National Examinations Council (NECO), The Joint Admissions and Matriculation Board (JAMB) and others. The duties of The National Examinations Council (NECO) are drafting of questions, time tabling, and administration of examinations, marking, scoring, grading and perhaps certification. One of the last acts of Abdulsalami Abubakar's military administration in Nigeria was the promulgation of a decree in 1999 that created the National Examinations Council (NECO). NECO was to take over the responsibilities of the National Board for Educational Measurement (NBEM) which was created in 1992. NECO was to take exclusive charge of the conduct of the Senior Secondary Certificate Examination (SSCE) for school-based candidate. NECO conducted its maiden Senior Secondary Certificate Examination (SSCE) in the mid-2000.

Mathematics as a subject affects all aspects of human life at different degrees. This is because of man's social, economic, political, geographical and technological exploits and use of numbers. In education, Mathematics is the bedrock of all sciences and technologically based subject. Hence, the relevance of Mathematics cannot be overemphasized. As a core subject, Mathematics is offered by all students in the secondary schools, whether Science or Arts inclined. Therefore, any examination instrument on Mathematics must be valid and reliable. This study therefore examines guessing parameters of 2016 and 2017 NECO Mathematics multiple choice test items in Calabar education zone.

Research Questions

For the study to have a focus, the following research questions were formed.

i. To what extent do the guessing parameters of the 2016 and 2017 NECO Mathematics multiple-choice tests differ?
ii. To what extent do the measurement theories and year of test influence reliability coefficient of 2016 and 2017 NECO Mathematics tests differ?

Research Hypotheses

The following null hypotheses were formulated to guide the study:

i. There is no significant difference between the guessing parameter of the 2016 and 2017 NECO Mathematics multiple choice test items.
ii. There is no significant influence of measurement theories and year of tests on the reliability of 2016 and 2017 NECO Mathematics tests.

METHODOLOGY

The research area for this study was Calabar Education Zone of Cross River State, Nigeria. The Calabar Education Zone is located in the Southern part of Cross River State, Nigeria and lies between latitude $4^0 28'$ and $6^0 35'$ North of the equator and longitude $7^0 50'$ and $9^0 28'$ East of the Greenwich meridian, with an area of 20,156km^2 (National Board of Technical Education, 2010).

The study adopted the instrumentation design for the study. This is because the research seeks to investigate measuring instruments to ascertain some level of certainty concerning their properties. Instrumentation design according to Mehrens and Lehman (1996) in Okeme (2009) is a type of design which aims at developing and certifying the efficacy of an instrument for the measurement of a given behavior or construct. The population of the study consisted of 7,591 senior secondary three (SSIII) students in Calabar Education Zone of Cross River State Nigeria in the 2018/2019 academic year (Department of Planning, Research and Statistics, Secondary Education Board Calabar, 2018). This population spread across eighty-nine (89) public secondary schools. This population comprises of both male and female students. The stratified random sampling technique was used to select 1,351 senior secondary three (SSIII) students drawn from 20 out of the 89 secondary schools in the education zone. The student sample size was 17.78% of the total population of 7,591. The instruments for this study were Mathematics multiple-choice test items of both 2016 and 2017NECO Examinations. There were sixty items each of the 2016 and 2017 NECO examination being multiple choice items of five-point response options. ANOVA and correlated t-test were used to test the hypotheses formulated for this study.

Data Presentation and Findings

Hypothesis one: There is no significant difference between the guessing parameters of 2016 and 2017 NECO mathematics multiple choice test.

The independent variable in this hypothesis is the year of examination which is categorical (2016 and 2017) while the dependent variable is a guessing parameters which is a continuous variable. To test the null hypothesis of significant difference between the guessing parameters of 2016 and 2017 NECO mathematics multiple choice item the correlated t-test was applied with results as showed in Table 1.

Table 1: Correlated t-test of the guessing parameter of 2016 and 2017 multiple-choice test (N = 1351)

Year	No. of items	$\overline{x}$	SD	Df	t-cal	Sig.
2016	60	0.30	0.91	1350	2.63*	.001
2017	60	0.13	0.18			

* p<.05; t-cri = 1.96

In Table 1 since the calculated t-value of 2.63 is greater than the critical t-value at 0.05 level of significance with 1350 degree of freedom, the null hypothesis that there is no significance difference between the guessing parameter of 2016 and 2017 multiple choice test was rejected and the alternative hypothesis upheld.

Hypothesis two: There is no significant influence of measurement theory and year of test on the reliability of the NECO mathematics test.

The independent variables are year of examination which is a categorical variable and the measurement theory which is also categorical (CTT and IRT). Dependent variable is a reliability which is continuous variable. To test the null hypothesis of the influence of model theory and year of examination on the reliability, the 2-way repeated measures ANOVA was used, with results as presented in Table 2.

Table 2: 2-way repeated measure ANOVA of influence of measurement theory and year of test on item the reliability (KR_{20}) coefficients of NECO mathematics test.

Source factor		SS	df	MS	F	Prob.
Year	Sphericity assumed	.029	1	.029	16.00	.000
	Greenhouse-Geisser	.029	1.000	.029	16.00	.000
	Huynn feldt	.029				
	Lower bound	.029	1.000	.029	16.00	.000
Error (year)	Sphericity assumed	.002	1	.002		
	Greenhouse-Geisser	.002	1.000	.002		

Source factor		SS	df	MS	F	Prob.
	Huynn feldt	.002				
	Lower bound	.002	1.000	.002		
Model (theory)	Sphericity assumed	.001	1	.001	0.12	0.001
	Greenhouse-Geisser	.001				
	Huynn feldt	.001				
	Lower bound	.001	1.000	.001	.012	0.001
Error (theory)	Sphericity assumed	0.000	1	.001		
	Greenhouse-Geisser	0.000		.001		
	Huynn feldt	0.000				
	Lower bound	0.000	1.000			
Yr. *Model (theory)	Sphericity assumed	.001	1	.001	.091	.040
	Greenhouse-Geisser	.001	1.000	.001	.091	.040
	Huynn feldt	.001				
	Lower bound	.001	1.000	.001	.091	.040
Error (theory)	Sphericity assumed	.001	1	.001		
	Greenhouse-Geisser	.001	1.000	.001		
	Huynn feldt	.001				
	Lower bound	.001	1.000	.001		

Table 2 shows the summary of repeated measures ANOVA with corrected F-values. The table is split into sections for each of the effects for year of test measurement theory, interaction and their associated error term. From the table, it was observed that there was significance in each of the main effects and the interaction. That is there is significant influence of year of examination, measurement theory and their interaction on the reliability. Therefore, the null hypothesis of no significant influence is rejected and the alternative hypothesis upheld.

DISCUSSION OF MAJOR FINDINGS

The finding of this study in the first hypothesis shows that there is a significant difference between the guessing parameter of NECO 2016 and 2017 mathematics multiple choice test. This means that the degree of guessing is not the same for the two years. This may be so owing to distinction in item difficulty in the two tests as guessing occurs when the ability level required in an item is higher than that of the testee. The study also revealed that the guessing parameters of 2016 and 2017 NECO mathematics test ranges from 0 to 0.30 and 0 to 0.08 respectively. The findings on this parameter are consonant with literature which posits that a parameter ranges in practice from 0.0 to 0.30 (Nailo 2004, Natarajan, 2009).

The finding also corroborates Uyani and Culer (2014) who carried out a study on classical test theory and item response theory of items using item parameters, 25 item Turkish test was used as instrument for study. The study revealed that there was a significant difference between the guessing parameters.

Furthermore, Praisipong, 2016; Gad, 2011, who avered that guessing model produced higher correlations between the parameter values and estimated values, the finding is also supported by the study of Prohoda, Parkard, McMahan and Jones (2006) who compared uncorrected and corrected scores on examinations using choice format and found out that there was a significant difference between the corrected and uncorrected scores.

One goal of a valid and reliable classroom test is to decrease the chance that random guessing could result in the right response. The greater the number of plausible distractors, the more accurate, valid and reliable the test typically becomes. Teachers should try as much as possible to construct items with plausible distractors, and items with such quality that the probability of the students getting the response right by guessing is low.

The results of the second hypothesis showed that there was a significant influence of the measurement theory and year of test on the item reliability coefficients of NECO mathematics test. In fact, each main effect: theory and year was significant. The two way interaction of measurement theory and year of test (theory *year) was also significant at .05 level. This means that the degree of consistency of the instrument and of the items differ over the years, as well as measurement theory. This may be so because of the uniqueness of the instruments as well as the year of test. The finding of this study agrees with Anastasi (1992) who avered that the test reliability as consistency of scores obtained when retested with an identical test does vary with test. It gives answers as to how good the test is? (Anaglwogu, 2005; Obukohoro, 2005; Idaka, 2007; Undorbuoye, 2005, Ashibi, 2005; Udom, 2004 and Ebot, 2007) are also concurrent of reliability as a function of theory of measurement and typical instrument used.

Also, the results of the finding agrees with the Obinne (2011) who carried out a study to compare the standard error of measurement (SEM) of Biology

examination conducted by NECO and WAEC from 2000 – 2002 using one parameter model of item response theory (IRT). The result showed significant differences in the SEM of biology examinations conducted by NECO and WAEC in 2000, 2001, 2002. However, Adegoke (2014) carried out a study on comparison of item statistics of physics achievement test using a classical test and item response theory frameworks. The results of the test revealed that there is significant difference between reliability coefficients using the CTT and the IRT frameworks.

Reliability is the quality of a test which produces scores that are not affected much by chance. Students sometimes randomly miss question they really knew the answer to or sometimes get an answer correct just by guessing, teachers can sometimes make an error of scoring inconsistently with subjectively scored test. These cause low reliability. Classroom teachers can solve the problem of low reliability by setting tests with many items, many item test is more reliable than shorter ones and the more objective a test is, the fewer random errors there will be in scoring. So teachers concerned about reliability are often drawn to subjective format. Teachers should often use a detailed scoring rubic to make scoring as objective and therefore as reliable as possible.

CONCLUSION AND RECOMMENDATIONS

The finding of this study shows that there is a significant difference between the guessing parameter of NECO 2016 and 2017 mathematics multiple choice test. And that that there was a significant influence of the measurement theory and year of test on the item reliability coefficients of NECO mathematics test. In fact, each main effect: theory and year was significant. Based on the findings, the following recommendations were made;

i. Guessing as a form of insincerity and corruption should be discouraged by using correction for guessing and moral education. Testees should be advised to leave blank any item they do not know.
ii. No effort should be spared to increase the reliability of instrument before administration as this speaks of the credibility of the instrument. Item quality, test length and sample size may be correlates of reliability depending on the measurement theory used.
iii. Based on these findings, it was recommended that NECO should use the item response theory (IRT) in assessment for its relative merit over the CTT.

REFERENCES

- Adegoke, B. A. (2014). Comparison of item statistics of physics achievement test using classical test and item response theory frameworks. *Educational Journal*, 22 (4), 23 – 43.
- Akuezuilo, E. O. & Agu, N. (2003). *Research and statistics in education and social sciences, methods and applications*: Awka: Nuelcenti publishers & Academic Press Ltd.
- Anagbogu, G. E. (2005). Analysis of psychometric properties of WAEC and NECO examinations and students ability parameter in Cross River State, Nigeria Unpublished Doctoral Dissertation. University of Calabar, Calabar, Nigeria.
- Anatasi, A. (1992). *Psychological testing* (7th ed.) New York: Macmillan.
- Ashibi, L. (2005). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, 5, 11-19.
- Capper, J. (1996). *Testing to learn……learning to test: Academy for educational development.* Washington DC.
- Ebot, A. (2007). Analysis of psychometric properties and students' academic performance in 2009 and 2010 secondary mathematics. An Unpublished Master's Thesis. University of Calabar.
- Enyi, G. S. (2002). *Introduction to measurement and evaluation for universities, polytechnics and colleges of. Education. A Professional approach*. Enugu: Harrison Pal.
- Gao, S. (2011). *The exploration of the relationship between guessing and latent ability in models*. An unpublished thesis Southern Illinois University Carbondale.
- Idaka, I. E. (2004). Attitude of academic staff to student's evaluation of instruction in tertiary institution in Cross River State, Nigeria. Unpublished M.Ed Thesis, Faculty of Education, University of Calabar, Nigeria.
- Kline, R. B. (2000). *A psychometric pimer*, London, UK: Free Association Books.
- Mehrens, C. & Lehman, P. (1996). Measurement and evaluation in education and psychology. *Journal of Educational Measurement*. 22(4), 316-318.
- Michell, J. (1999). *Measurement in psychology.* Cambridge: Cambridge University press. Do1: 10.1017\CB09780511490040.
- Nararajan, V. (2009). *Basic principles of IRT and application to practical testing & assessment Merit Tracers*, India.
- National Board for Technical Education (2010). Institutions. Retrieved from http://en.m.wikipedia on 20/3/2010.
- Nwaobia, E. M. (1990). *Monograph on introduction to advanced measurement and evaluation*. University of Nigeria, Nsukka.
- Obinne, A. D. E. (2011). A psychometric analysis of two major examinations in Nigeria: Standard error of measurement. *International Journal of Educational Science*, 3(2), 137-144.
- Obukohwo, E. N. (2005). *Current issues in Nigeria educational system*. Abraka Delsu Publishers.

- Okeme, I. (2009). Development and validation of psycho-productive skill items for measuring performance of students in agricultural science in secondary schools in Kogi State. Unpublished Ph.D Thesis of University of Nigeria, Nsuka.
- Pasipong (2016). Analysis test of understanding of vectors with the three parameter logistic model of item response theory and item response curves techniques. *Physical Review Physics Education Research*, 4(12), 61-70.
- Prihoda, J., Pinkard, R., Macmahan, A. & Jones, A. (2006). Correcting for guessing increases validity in choice examination in an oral and maxlilo facial pathology course. *Educational Methodologies*, 24(3), 161-170.
- Sidhu, S. K. (2005). *New approach to measurement and evaluation.* New Delhi: Sterling Publishers Ltd.
- Thurstone, L. L. (1959). *The measurement of values*. Chicago: The University of Chicago press.
- Udom, S. U. (2004). *Item response dimensionality and some factors related to students' performance in mathematics*. Unpublished M.Ed Thesis, Faculty of Education, University of Calabar, Nigeria.
- Undorbuoye, M. (2005). Content validity of teacher made test in secondary schools in Enugu State. (Unpublished Master's Thesis). University of Nigeria, Nsukka.
- Uyani, G. K. & Guler, N. R. (2014). Comparison of classical test theory and item response theory in terms of item parameters. *European Journal of Research on Education*, 2(1), 1-6.