



## Estimation of the Sources of Error in WAEC Mathematics Test Score. Using Generalizability Theory

Ohagwu Sylvanus C., Ene Catherine U. & Okeke, Agnes O.

Department of Science Education, University of Nigeria Nsukka

Corresponding author: [ohagw8usylvanus@yahoo.com](mailto:ohagw8usylvanus@yahoo.com)

### Abstract

Mathematics is one of the important subjects studied in Nigeria. The objective of mathematics lends credence to the fact that Nigeria wants to harness these benefits and estimate any sources of measurement errors. The study used generalizability theory in estimating errors in WAEC Mathematics test scores. A fully crossed  $s \times q \times r$  design was used. Mathematics Essay Test was used to collect data and Kendal coefficient of concordance in reliability and 0.719 was obtained. EduG version 6.1 was used in analysis and hypothesis tested at 95% C.I using standard error. The study showed that there are other sources of errors in any test items and the work should be replicated using wider geographical scope.

**Keywords:** Mathematics, Generalizability theory, Classical test theory, Measurement error, Questions and Raters.

### Introduction

Mathematics is one of the important subjects that is been studied in primary and post primary schools in Nigeria. Mathematics helps the learners to build the sense of reasoning. Moreover, mathematics figures the foundation of ultimate development in scientific and industrial research. To buttress this, Gimba and Agwagah (2012), mathematics is the bedrock of all sciences, technology and modern developing nations depend on mathematics to survive. Also, some complicated networks and constructions in this present world like computer networking can be understood using mathematical knowledge. Therefore, any facet that constitutes error in measurement needed to be estimated. Justice, Osei and Daniel (2015) opines that mathematics is seen by society as the knowledge that is necessary for the development of scientific and technological nation. Also, mathematics is one of the oldest disciplines that do not lend itself to a generally acceptable definition, even among mathematicians. Elaine (2013) sees mathematics as a science that deals with logic, shapes, quantity and arrangement; but a comprehensive definition of mathematics should take into consideration all the branches of mathematics in the definition. The inability of scholars to capture all these branches of mathematics in its definition has resulted in different perceptions given to mathematics. Also, mathematics is a body of knowledge that is necessary for the achievement of scientific and technological nation (Anaduka & Okafor, 2013).

However, the researcher defines mathematics as a discipline which parades in all the activities of human endeavor and employs logic and algebra to foster creative thinking in an individual and ease some difficulties of life. Mathematics is very essential to this

present world as it is very vital in every discipline and in areas of Information and Communications Technology (Gimba & Agwagah, 2012). A closer examination of the objectives of studying mathematics in secondary school lends credence to the fact that Nigeria desires to harness the benefits. The objectives of teaching and learning mathematics include: alignment of the compulsory cross cutting subjects such as English language, General Mathematics, Civic Education, Trade/Entrepreneurship Studies; alignment of curriculum into four distinct fields of study (Nigerian educational research and development council (NERDC, 2012). Despite these importance and objectives of studying mathematics, students' scores in internal and external examinations still remain poor (Chief examiners report WAEC, 2016). Lassa (2012) lamented that if nothing positive is done to improve testers' efficiency in Mathematics, the state of the development in science and technology in Nigeria will be a mirage. More so, Rufa'i (2012) stated that something vital in mathematics must be done if the nation has to move ahead. Furthermore, Nalah and Daniel (2014) opined that the type of education given to the Nigerian child has become a source of treat due to measurement errors from sources.

Measurement error is the difference that exists in the measurement system when real value is subtracted from the ideal value of a measurement. Mathematically,  $E$  (measurement error) =  $X$  (ideal value) –  $T$  (real value). This implies that when real value is high, error increases and reliability decreases. According to Linn and Burton as cited in Ikeh (2016), error of measurement can be used to evaluate the minimum passing and maximum failing scores with a certain level of confidence. The need for estimating

error arises due to poor scores of students in WAEC examination and to see how these errors can be minimized to increase reliability of measurement. According to Nworgu (2015), reliability is the degree of consistency with which an instrument measure whatever it measures. According to Wagemaker (2010), the rich data obtained in every study must contain adequate information that allows the exposure of important relationships that was hidden.

Traditionally, methods of reliability that rely on Classical Test Theory (CTT) are in use which considers a single error source in a measurement and hence offers very few information. For instance, test-retest reliability. Some of these measurement errors that contribute to score variability include: raters, questions used, teachers, students among others. These are the sources of error that influence WAEC mathematics test scores. Once the measurement error due to these sources of error behind test scores are observed, the statistical frame of generalizability theory (GT) will be employed which has a mechanism of combining all these three reliability procedures into a single estimation procedure to produce only one reliability coefficient with its corresponding error.

According to Quansah (2020), Generalizability theory (GT) is a statistical measurement theory which expands CTT to involve many sources of error and explicitly links measurement operations to the purpose of measurement. Also, GT lets one to separate differences in measurement and the universe the researcher wants to draw inferences (Ene, 2015). Therefore, the researcher sees GT as statistical framework used to analyze measurement errors from various sources to produce a reliable result.

### Statement of the Problem

In an attempt to establish the WAEC mathematics test scores of secondary school students, there are characteristics other than students' factor which affect the scores made by them in examinations. These characteristics are called sources of error such as test questions; raters among others. These contribute to errors in measurement of students' test scores and affect score dependability of these measurements. Therefore, there is the need to find out the respective contributions of these multiple facets to measurement errors on WAEC test scores. To estimate measurement errors that influence students' test scores, involve a multifaceted approach that deals with multiple sources of error which then invalidate the use of CTT that considers only one source of measurement error. Generalizability theory is

more accurate when more than one error sources are involved. The observed scores in examinations are affected by factors other than students (learners). These facets like test questions, raters, among others are likely to affect reliability. The effect of these facets leads to ask about the exactness of scores obtained in schools as they were used in predicting the student's future.

Therefore, there is a need to find out and estimate the WAEC mathematics test scores within the context of GT. Hence, the problem of this study is what are the contributions of questions and scorers as errors and their interaction effects on WAEC scores?

### Purpose of the Study

The general purpose of this study was to apply generalizability theory in estimating the sources of error that influence WAEC mathematics students test scores over multiple sources (facets). The study was specifically designed to obtain:

- 1) the contributions of questions (q) and raters (r), to measurement errors in WAEC test scores.
- 2) the interaction effect of these facets in student test scores in mathematics.

### Research Questions

Two research questions were formulated to guide the study.

1. What are the contributions of questions and raters to measurement errors in WAEC mathematics test scores?
2. What are the interaction effects of these facets in student test scores in mathematics?

### Hypothesis

A single hypothesis was formulated to guide the study and was tested at 0.05 level of significance.

1. The contributions of questions, scorers (raters) and their interactions to measurement errors in WAEC mathematics test scores are not statistically significant.

### Method

The research design was a random effect two-facets fully crossed  $s \times q \times r$  design. It was used to estimate all variance components that exist in the measurement. The researcher wishes to use this design because the universe of admissible observations includes all the possible combinations of the facets. The area of the study was Udi Education Zone of Enugu State which was made up of Udi Local Government and

Ezeagu Local Government area. The population used for the study was sixty-four (64) mathematics teachers as raters and eight thousand nine hundred and seventy-six (8976) mathematics students ((SS 3)) in the fifty-four (54) government secondary schools in the Education Zone according to its Statistic (2018/2019). The sampled eight hundred and ninety-eight (898) students and 6 mathematics teachers. This is ten percent (10%) of both students and teachers' population. Similarly, 27 public secondary schools which is fifty (50%) percent of the 54 secondary schools was used for the study. Moreover, this study was done using two stage sampling. In the first stage, the researcher sample

the schools and sampling of SS3 students to be included in the study as the second stage. The instrument of test MAET was adopted from WAEC questions. The items were drawn from the questions based on the five major mathematics topics that were included in the WAEC examination years (2013 – 2016) academic sessions. The data was administered and collected by the researcher with the help of the mathematics teachers. The data collected was analyzed using EduG version 6.1. Also, the hypotheses were tested at 95% confidence interval using standard error.

## Result

**Table 1: The result of the variance component estimates for students, questions, raters and interaction on MAET.**

Source	Variance Component	SS	Df	MS	%
S	$\sigma_s^2$	523280.0231	311	1682.5724	72.2
Q	$\sigma_q^2$	0.0000	4	0.0000	0.0
R	$\sigma_r^2$	362.5321	2	181.2660	0.1
SQ	$\sigma_{sq}^2$	85741.0667	1244	68.9237	9.3
SR	$\sigma_{sr}^2$	16293.4679	622	26.1953	0.0
QR	$\sigma_{qr}^2$	0.0000	8	0.0000	0.0
SQR	$\sigma_{sqr}^2$	68027.3333	2488	27.3422	18.4
Total		693704.4231	4679		100%

Table 1 showed that the largest source of variation component recorded is the variance associated with the student [ $\sigma^2(S)$ ], which accounted for 72.3% of the total variance in the G-study. Also, the result indicated that there is other source of error due to student that was not measured. This shows that students' performances across task are inconsistent.

What are the interaction effects of the facets to measurement error in student test scores in mathematics?

From the table above, 18.4 % of student-by-question-by-rater [ $\sigma^2 (SQR)$ ] interaction was observed as the

second largest variance component recorded in the G-study. Moreover, 9.3% of student-by-question [ $\sigma^2 (SQ)$ ] interaction was seen as the third largest variance component in the study. This portrays that a large number of the questions were not attempted by the students while raters are unique in rating the items. Furthermore, very small error of 0.1% of raters' variance component were seen. More so, zeros (0.00) variance components were associated with question [ $\sigma^2 (Q)$ ], student-by-rater and question-by-rater interaction [ $\sigma^2 (QR)$ ] were recorded. This suggesting that error which influence rating of items were not from them.

## Discussion

The results of this study have reveals that there are various sources of measurement errors in any test items. Also, the findings of this study are in agreement with the findings of Ikeh (2016) and Guler and Gelbal (2010) who opines that enlarging the items will give a better reliability coefficient and that increase in the number of raters increases the reliability ( $\Phi$  or Phi) coefficient. The examination bodies, curriculum planners and researchers will know directions of their facts and generalizations. More so, teacher factors such as gender, attitude, personality, teaching methods among others were not considered. Hence, these induce score variations from different schools and may have affected the result of the study. Furthermore, in testing the hypothesis at 95% CI using standard error, we observed that there was an overlap of their collective variance components in both relative and absolute error variances which shows that there was no significant difference existing between their test scores over multiple facets and their interactions at 0.05 level of significance.

## Conclusion

Based on the findings of the study, the following conclusions were drawn.

1. The variance 72.2 % obtained from the students [ $\sigma^2(S)$ ], which were seen as the highest contributor of error followed by 18.4% from the residual factor [student by question by rater [ $\sigma^2 (SQR)$ ]]. The third largest variance component of 9.3% of student by question [ $\sigma^2 (SQ)$ ] followed by 0.1 of rater in error contribution. The variances zero (0.00) each were found in variances of question [ $\sigma^2 (Q)$ ], student-by-rater [ $\sigma^2 (SR)$ ] and question-by-rater interaction [ $\sigma^2(QR)$ ].
2. Increasing the number of items produces a higher reliability coefficient.

## Recommendations

Based on the limitations and findings of the study, the researcher suggests a replication of this study using a wider geographical scope; using different D-studies, reliability for different conditions which include a greater number of raters and items. Also, it suggested that further studies should be carried using the same number of raters on Joint Admission and Matriculation Board (JAMB). Furthermore, Test item writers, researchers and Examination bodies should embark on

generalizability analysis when dealing with multiple error sources.

## References

- Abdullahi, O. E. (2013) Inter-Relationship Between Personal Factor and Academic Achievement in Mathematics. Ebira Secondary School Students in Kogi State. *Ife Psychologic An Inter-national Journal*, 5(1), 154-155. Retrieved June 15, 2013 from [www.unilorin.ed.ng/publication/abudullahioe/interation](http://www.unilorin.ed.ng/publication/abudullahioe/interation) between personal factor academic achievement.pdf)
- Anaduka, U.S. and Okafor, C.F. (2013). *Poor performance of Nigerian students in mathematics in senior secondary examination, (SSCE); what is not working* (on line). Ret rival from [www.agol.info/journals/iorinal](http://www.agol.info/journals/iorinal).
- Caldwell, D. J; Sampognaro, I. & Pate, A. N. (2015). Effect of question formats on student and item performance. *American journal of pharmaceutical education*, 79(6); 87.
- Cronbach, I.J; Rajarantnam, N; Nanda, H & Gleser, G.C. (1972). *The Dependability of Behavioral Measurement: Theory of Generalizability*. New York, John Wiley.
- Egbulefu, C.A. (2013). *Estimating Measurement Error and Score Dependability in Examination Using Generalizability Theory*. An unpublished thesis. University of Nigeria Nsukka.
- Elaine, J.H. (2013). *Definition of mathematics* [on line]. retrieve 15 August from <http://M.linescience.com> 138.
- Ene, C.U. (2015). Comparison of three test theory in measurement. *Journal of general studies* 2(1), 53-61.
- Gimba, R. W. & Agwagah U. N. V. (2012). Importance of Mathematics to Science and Technology. *Journal of Science, Technology, Mathematics and Education (JOSTMED)* Volume 8(2). April, 2012.
- Han, T. (2013). The Impart of Rating Methods and Rater Training on the Variability and Reliability of EFL Students' Class room Based Writing Assessment in Turkish Universities. *An Investigation of Problems and Solutions*. (Unpublished doctorate dissertation) Turkey: Atatilrk University.
- Ikeh, E.F. (2016). *Estimating Multiple Sources of Variation and Score Dependability in Economics Essay Test Using Generalizability*





- Theory*. An unpublished thesis. University of Nigeria Nsukka.
- Justice, E; Osei, K.A & Daniel, N. (2015). Factors Influencing Students Mathematics Performance in Some Selected Colleges of Education in Ghana. *An International Journal of Education Research Learning and Development*, vol 3, no 3 pp.68-74 (publish by European Centre for Training and Development Uk).
- Lassa, P.M. (2012). *The teaching of mathematics for Nigerian secondary schools*, Fab Anieh.
- Lewis, W.D. and Young, T.V. (2013). The Politics of Accountability: Teacher Education Policy. *Education Policy*, 27(2), 190-216.doi:10.1177/0895904812472725.
- Lombardi, A; Seburn, M; Conley, D. & Snow, E. (2010). *A Generalizability of Cognitive Demand and Rigor Ratings of Items and Standards in an Alignment Study*. Presented at the annual conference of American educational research association. Denver, co. educational empowerment Centre. 720E.13<sup>th</sup> Ave, suite 202.
- Murunga, F; Kilaha, K. & Wanyonyi, D. (2013). Emerging Issue in Secondary School Education in Kenya. *Int. J. Adv. Res*, 1(3): 231-240.
- Nalah, A.B. and Daniel, L.I. (2014). *Discourse Journal Of Education Research*. Factors Mitigating Ducational Stability of Student in central Nigerian and the way. ISSN; 2346-7045, vol 2(1); 1-5.
- NERDC (2012). *The Senior Secondary Education Curriculum Structure*. Unpublished position paper from the curriculum Centre, NERDC. Nigerian limited Jos, e-mail: [fabniehpress@yahoo.com](mailto:fabniehpress@yahoo.com)
- Nworgu, B.G. (2015). *Educational Measurement & Evaluation*, Theory and Practice. Third Edition. University trust publishers. Nsukka, Enugu State, Nigeria.
- Quansah, F. (2020). Students' Evaluation of the Quality of Teaching Using Generalizability Theory: A Case of a Selected University in Ghana south African *journal of higher education*. Doi: 10.20853/34-5-4212. Retrieved from <https://www.semanticscholar.org.s...>
- Rufa'march, A.R. (2012). Nigerian lament poor performance of students in mathematics. *Daily Champion Newspaper of March 2<sup>nd</sup>*. Retrieved 20/01/2014 from <http://www.africasti.com/headline/Nigeria-lament-poor-performance-of-students-in-mathematics>.
- Shavelson, R.J. & Webb, N.M. (2005). Generalizability theory. 1973-1980. *British journal of mathematics and statistical psychology*, 34, 133-166. [http://dx. Doi.org/10,1111/j,2044-8317. Tb00625.x](http://dx.doi.org/10.1111/j.2044-8317.1500625.x). *Journal of Psychology*, 15, 72-101.
- Wagemaker, H. (2010). IEA; *International Studies, Impact and Transition*. Paper presented in the 4<sup>th</sup> IEA international research conference. July1-3. Gothenburg, Sweden. Retrieved from <http://www.iea.u/irc-2010.html>.
- West African Examination Council (2016). Chief Examiner's Report. Lagos: WAEC.