International Journal of Psychological and Brain Sciences
2023; 8(2): 13-18
http://www.sciencepublishinggroup.com/j/ijpbs
doi: 10.11648/j.ijpbs.20230802.11
ISSN: 2575-2227 (Print); ISSN: 2575-1573 (Online)

Application of Generalizability Theory in Measurement Error in 2019 WAEC Mathematics Objective Examination in Benin Metropolis

Kennedy Imasuen[1], Praise Kehinde Adeosun[2]

[1]Institute of Education, University of Benin, Benin City, Nigeria

[2]Department of Educational Evaluation and Counseling Psychology, University of Benin, Benin City, Nigeria

Email address:

kennedy.imasuen@uniben.edu (Kennedy Imasuen), praise.adeosun@uniben.edu (Praise Kehinde Adeosun)

To cite this article:

Kennedy Imasuen, Kehinde Adeosun. Application of Generalizability Theory in Measurement Error in 2019 Waec Mathematics Objective Examination in Benin Metropolis. *International Journal of Psychological and Brain Sciences*. Vol. 8, No. 2, 2023, pp. 13-18.
doi: 10.11648/j.ijpbs.20230802.11

Received: April 2, 2022; **Accepted**: April 19, 2022; **Published**: May 29, 2023

Abstract: This study investigated measurement error in 2019 WAEC senior secondary school examination using generalizability theory. The study was specifically concerned with identifying and analyzing measurement error in the senior secondary school 2019 WAEC mathematics objective examination using generalizability theory, and also to determine the highest contribution of facets: students, items and teachers to measurement error. Four research questions were raised to guide the study. The study was survey which adopted a random effect two-facet fully crossed $s \times t \times i$ design for a generalizability (G) and decision (D) studies. The population consisted of fifty-six thousand, seven hundred and ninety-seven (5697) senior secondary three (SS3) students in the seventy-five (75) public secondary schools in Benin metropolis for the 2019/2020 academic session. The instrument for data collection was a fifty (50) multiple choice WAEC, mathematics 2019 examination. The instrument has been validated by the West African Examination Council (WAEC). The reliability of the items was ascertained using the Kuder – Richardson 20 (KR 20) to obtain internal consistency. It gave a value of 0.92. Data collected were analyzed using a software EduG version 6.0-e based on analysis of variance (ANOVA) and generalizability. The findings which emerged from the study were: the highest contribution to measurement error in examination scores was the students - teacher interaction which accounted for 68.9%, this was followed by the student factor (27.5%) and the residual, that is, interaction of student, teachers and items (3.6%). A generalizability coefficient of 0.97 high enough to rank order students according to their relative abilities in examinations was obtained when the number of teachers was increased to 78. Based on the findings, it was therefore recommended that generalizability analysis should be carried out by researchers, test developers and examination bodies so as to reduce or eliminate measurement error and hence maximize reliability.

Keywords: Error, Measurement Error, Generalizability, Variance Component

1. Introduction

Measurement permeates almost every aspect of modern society, because it is carried out by every individual both learned and unlearned. A great variety of things about individual which includes achievement, aptitude, intelligence, height and weights are measured by various people like teachers, doctors and so on. The results of measurement can have a profound influence on the individual's life; thus, it is important to understand how these scores are derived and the accuracy of the information they contain. Measurement plays an important role in the Education, Sciences, Engineering fields as well as many everyday activities. It involves assigning numbers to a particular type of measurement with refence to a specific rule. Omoroguiwa [19] defined measurement as the assigning of numbers to show the presence or absence of a trait in an individual, group or objects or to show the extent to which such a trait is present or absent, according to specified rules. Measurement can also be seen as the process of assigning numerical values to describe features or characteristics of objects, persons or certain events in a systematic manner. In education, traits like aptitude, opinion, scholastic ability or achievement can be measure using different instrument like test, questionnaire etc. These instruments have various purposes which includes description, prediction, assessing the individual differences, objective evaluation domain estimation, mastery decisions and diagnosis [14]. Another important usage of educational measurement is to classify examinees. Classification is regarded as high stakes because errors in classification may cause wrong decisions [6].

Muler and Linn in Esomonu and Okeaba [9] defined measurement error as a situation in which a student's true ability or achievement is either underestimated or overestimated. Also, Hofman in Egbulefu [8] stated that measurement error is defined as the difference between the distorted information and undistorted information about a measured product expressed in its physical quantity. An error is defined as real (untrue, wrong, false) value at the output of a measurement system minus ideal (true, good right) value at the input of a measurement system mathematically expressed as:

$$\Delta_x = X_r - X_i \qquad (1)$$

where

Δ_x is the measurement error,

X_r is the real untrue measurement value and

X_i is the ideal true measurement value.

A measurement under ideal condition has no error.

Error of measurement are categorized as random or systematic. Systematic errors are those errors which consistently affect an individual's score because of some particular characteristics of the person or the test that has nothing to do with the construct being measured. On the contrary, random errors of measurement affects an individual's score because of purely chance happenings. They may affect an examinee's score in either a positive or negative direction. Both random and systematic errors are a source of concern in score interpretation. Measurement (random) errors can result from the way the test is designed, or from the factors related to individual students or the testing situation and many other sources like the mood of examiner, the time of the test (occasion), the test environment, invigilators and the changing order of the questions, which may lead to higher or lower scores [13]. The need for estimating measurement error arises because of the inconsistencies in measurements especially those involving multiple sources of error. The low performance of students in examinations such as the senior secondary school examination calls for the estimation of multiple sources of error, so as to determine the contributions of error of the different facets in examinations and then see how these errors can be minimized or eliminated and hence increase reliability in examination scores.

More so, for subjects like mathematics whose relevance is so much, it is necessary to estimate measurement errors. Despite the significance of the subject. Bichi, et al [3] noted that over the years, the performance of students in senior secondary school mathematics in Nigeria have consistently been poor and unimpressive based on data from the two public examination bodies, that is, (the West African Examination Council [WAEC] and National Examination Council [NECO]) Secondary School Certificate Examination (SSCE) indicated that students' achievement in mathematics was low. According to him, about 71% of the candidates who sat for the October/November 2014 West African SSCE private failed to obtain five credits with English Language and Mathematics. A total of 241,161 candidates who sat for the examination, but only 72,522 candidates representing 29.37% got credits in five subjects including English Language and mathematics [1]. The West African Examinations council results in 2018 also showed that a total of 1.57m candidates sat for WAEC as public students, the results shows that 48.15% had 5 credits and above including English and Mathematics while 51.85% failed to do so. In the same year a total of 109,798 candidates sat for WAEC as private students but only 33.81% of them had 5 credits and above including English and Mathematics while 66.19% did not [17]. This low performance of students in examinations calls for the estimation of multiple sources of error, to determine the contributions of the different facets in examination to error and then see how these errors can be minimized or eliminated and hence increase reliability of examination scores [9].

The reliability of performance evaluation functions with methods is based on the three basic theories of measurement, classical test theory (CTT), item response theory (IRT), and generalizability theory (GT) [10]. Once the measurement error due to these sources of error are observed, the statistical frame work of generalizability (G) theory will be brought to bear on the technical quality of performance assessment scores. According to Lee [15], employing the generalizability theory approach can analyze more than one source of measurement error simultaneously in addition to the object of measurement. G theory is a structure for visualizing, scrutinizing and planning dependable observations. It is to ascertain how dependable a measurement is under a certain situation. It is also a statistical tool for evaluating results of psychometric tests such as objective test, computer adaptive test, among others. He also opined that to identify and reduce measurement error poses a major challenge in psychological testing. This has made most researchers to relied on the CTT for assessing reliability, even though many authors have advocated the use of generalizability theory in the estimation of the various sources of error in examination and assessment of performance.

Cronbach, and his associates introduced the theory of generalizability to make it possible to assess multiple sources of error in measurement, using some of the same principles of traditional analysis of variance (ANOVA). Generalizability theory uses variance components to represent the amount of error that comes from generalizing from a facet score to a universal score. In any measurement situation, there is a desire to obtain scores that are able to accurately separate the performance of different examinees while also minimizing the variability in the other factors (for example, items or invigilators). Variability in these other factors (facets) reduces the accuracy in the measurement of examinee performance.

Generalizability theory (GT) pinpoints the sources of measurement errors, disentangle them and estimate each one. In generalizability theory, a behaviour measurement say a test score is conceived of as a sample from a universe of admissible observation. This universe of admissible observation consists of all possible observations on an object of measurement which in most cases is a person. Each characteristic of the measurement situation, (for example, test form, test items, rater, or test occasion) is called a facet. The universe of admissible observation is usually defined by the Cartesian product of the levels (called conditions) in generalizability theory of the facet. In this study, the object of measurement is students (s), and the two instrumentation facets are items (i) and raters (teachers) (r).

The two different types of studies in generalizability theory are: generalizability study (G study) and decision study (D study). A generalizability study is done to determine or ascertain how well scores can be used for multiple situations. A generalizability study involves estimating variance components that might turn to be used in a D-study for computing generalizability coefficient. D-study on the other hand is conducted mainly for the purpose of determining the most efficient measurement procedure for a given situation. There are also two types of error variances, which corresponds to relative decisions and absolute decisions. According to Strube [21], relative decisions are decisions about individual differences between students, while absolute decisions are decisions about the absolute level of performance. The relative error variance ($\sigma_{\delta}{}^{2}$) is of primary concern when researchers are interested in decisions that involve the rank ordering of individuals. In this case, the error sources are limited to the interactions of the individuals with the facet(s) formed by random sampling of the conditions of measurements. This is because interactions involving the object of measurement reflect changes in relative standing across facet levels. The absolute error variance is of concern in decisions about whether a student can perform at a prespecified level. It reflects both information about the rank ordering of students and any differences in the average scores. All sources other than the object of measurement are a source of error for absolute decisions [21].

There are also two coefficients of reliability as generalizability (G) and dependability ($\emptyset$). The difference between the two coefficients is based on the definition of what constitute error for the type of decision to be made. It is on this premise that this study intends to apply generalizability theory in the estimation of measurement error and score dependability of the 2020 West African Examination Council Mathematics objective test.

1.1. Statement of the Problem

In measuring student's performance in a given examination, there are characteristics other than the students' factor that affect the scores made by them. These characteristics called sources of error such as test questions, invigilators, and so on contribute to error in measurement of students' achievement and they affect the score dependability of these measurement hence the need to find out their contributions to measurement error. Observed scores in any examination are affected by factors other than the students. Such specific factors (facet) as test questions, invigilators among others are likely to affect the reliability of an observed score in any examination. The impact of these factors leads to questions about the accuracy, precision, and ultimately, the fairness of the scores obtained by students in any given examination.

Since scores obtained by the objects of measurement, (students) in examination are affected by multiple sources of error and scores from the examinations are used in making relative and absolute decisions concerning students, there is the need to estimate measurement error and score dependability of examinations using generalizability theory, so as to determine the contributions of error of these facets in measurement situations in examinations with a view to minimizing and maximizing reliability of their scores. Estimating measurement error and score dependability of any given task involves multifaceted approach which the classical test theory cannot address as it addresses only one source of measurement error.

Generalizability theory has this added advantage over CTT as it can assess the effects of multiple sources of error or when more than one random facet is involved [5]. Also, generalizability theory provides generalizability estimates, which represent a raw proportion of the total variance accounted for by each included factor that are thought to systematically relate to the construct of interest. In the light of this, the present study seeks to assess measurement error and score dependability of WAEC 2019 Mathematics objective test using generalizability theory.

1.2. Research Questions

The following questions guided the study

1 What is the contribution of the facets: students(s), items (i), teachers (t) to measurement error in WAEC 2020 Mathematics objective test scores?

2 To what extent do the generalizability coefficients show the degree to which students maintain their rank order across facets: item (i), and raters (t) in WAEC 2020 Mathematics objective test scores?

2. Literature Review

Some authors have carried out various studies using generalizability theory. For example, Esomonu and Okeaba [9] estimated measurement error and score dependability of the inventory for students' integration into the University Academic Culture using Generalizability Theory. The results show that the highest contribution to measurement error in ISIUAC scores was the residual which accounted for 85.6% of the total variance. Ogidi [18] utilized generalizability theory in estimation of variance components in National Examination Council Examination Council Essay Questions in Christian Religious Studies. Results of the study showed that the largest contribution to error is person by item by rater (5.244) with the percentage variance of (46.654%) and person by item (4.361) with a percentage variance of (38.43%), person by rater (0.425) with a percentage variance of 3.76%, person (0.383) with a percentage variance of 3.63%, rater (0.77) with a percentage variance of 7.36%, item (0.028) with a percentage variance of 0.24%, item by rater (-0.06*) with a percentage variance of – 0.74%.

Iheanyichukwu and Orluwene [12] applied generalizability theory in estimation of variance components in National Examination Council Questions in Mathematics. The findings revealed that the largest contribution to measurement error from the score obtained was on the student*item*tater (9.830, 63.5%). The second largest source of variance was student and item (4,157, 26.8%) Third, item (1.172. 7.5%). Then rater (.001, 6.46%) followed by students (.549, 3.5%). Webb et al [22] in their study opined that more observers, gave a generalizability coefficient that rank ordered person in generalizability study of job performance measurement of Navy Machenist Mates. In the same vein, Brennan [4] study showed that with multiple raters, it was possible to differentiate between persons (Coeff. G relative 0.91). Therefore, to achieve a generalizability coefficient ($E\rho^2$) 0.91, the number of teachers were increased in order to rank order students relatively in examination. Bamidele, et al [2] carried out a study in estimating generalizability and dependability indices of students' scores in teaching practice assessment in a Nigerian College of Education. The finding revealed that residual has the largest contribution as it was responsible for 60% of the total variation in the students' scores in teaching practice course.

3. Methods

The study was a survey which adopted a random effect two-facet fully crossed $s \times t \times i$ design for a generalizability (G) and decision (D) studies. The fully crossed design in the G – study was used to estimate all the possible variance components in the measurement situation. The D – study used the information provided by the G – study to design the best measurement procedures minimizing undesirable sources of measurement error and maximizing reliability. The population of the study was all the senior secondary three (SS3) students of public secondary schools in Benin metropolis for the 2019/2020 academic session. Benin metropolis consist of four local government areas which are Egor, Oredo, Ikpoba-Okha and Ovia North – East. There are seventy – five (75) public senior secondary school in these four local government areas with student population of 5697 students. The sample for the study was 570 students which represent 10% of the total population of SS3 students in the four local government area. They were selected from thirty-eight (38) schools in the locality. The multi – stage sampling technique was adopted for the study. The instrument used for data collection was a fifty (50) multiple choice of the 2019 WAEC mathematics objective questions for the 2019 examination year. The fifty items cover the six major topics in mathematics namely; numbers and numeration, algebraic and arithmetic processes, mensuration, trigonometry, geometry, and statistics and probability. The objective items were constructed by WAEC and are assumed to have been validated and standardized before it was administered to the students. The items covered a range of topics in Mathematics showing that it is also content valid and considered appropriate for utilization in the study. The reliability of the instrument was established using a sample of 50 students and five teachers from public senior secondary (SS 3) who were not used in the main study. The reliability of the instrument was determine using the Kuder – Richardson 20 (KR 20) to obtain the internal consistency. It gave a value of 0.92. A computer software, EduG version 6.0-e based on analysis of variance (ANOVA) and generalizability theory was used to analyzed the data collected. It was therefore used to answer the two research questions raised.

Table 1. *Variance components of the contributions of the facets of the study to measurement error in the 2019 WAEC senior secondary examination.*

Sources	Sum of squares	df	Mean square	Variance Components	Estimate	% of total variability
Students (s)	591198.822	14	42228.487	$\sigma^2 s$	21.278	27.5
Teachers (t)	96515.024	37	2608.514	$\sigma^2 t$	0.000	0.0
Items (i)	153.806	48	3.204	$\sigma^2 i$	0.000	0.0
s × t	1351210.340	518	2608.514	$\sigma^2 st$	53.234	68.9
s × i	2153.280	672	3.204	$\sigma^2 si$	0.012	0.0
t × i	4876.083	1776	2.746	$\sigma^2 ti$	0.000	0.0
s × t × i	68265.158	24864	2.746	$\sigma^2 sti, e$	2.746	3.6
Total	2114372.513	27929				100.0

4. Results

Table 1 showed that the highest contribution to measurement error in 2019 WEAC senior secondary school examination score came from the interaction of the students and the teachers (invigilators), accounting for 68.9% of the total variability. This was followed by the student factor which accounted for 27.5%. The interaction of the students, teachers and the items contributed only 3.6% of the total variability. However, the facets, teachers, items, and the interaction of the students and items, teachers and items did not contribute to measurement error in the study.

Table 2. *Estimated generalizability coefficient ($E\rho^2$) for a fully crossed $s \times t \times i$ D-study Design with different number of teachers.*

Number of teachers	$E\rho^2$
38	0.86
48	0.90
58	0.91
68	0.94
78	0.97

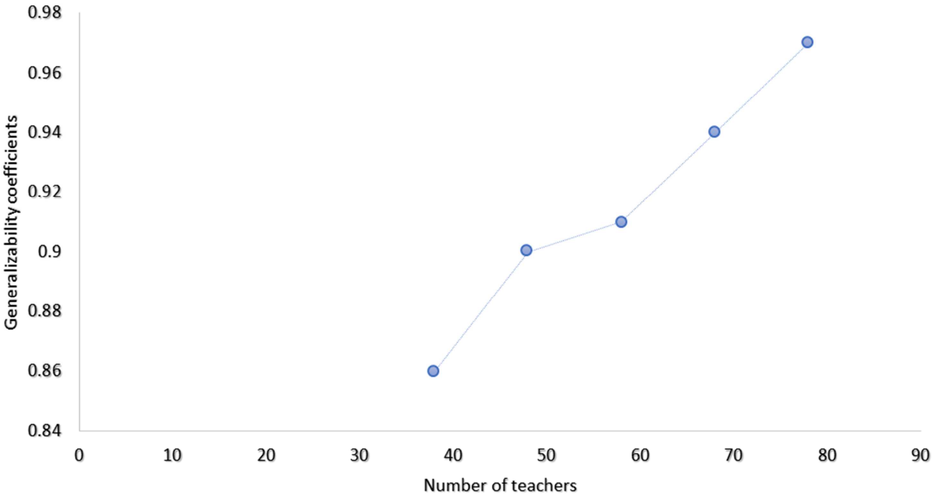

Figure 1. *Generalizability coefficients resulting from relative decisions for different teachers.*

Table 2 and figure 1 showed the impact of increasing the level (numbers) of teachers (invigilators) on students (object of measurement) in terms of their relative standing. They showed a steady but gradual increase. When the number of teachers was increased from 38 to 48, generalizability coefficient ($E\rho^2$) was increased from 0.86 to 0.90. When the number of teachers was increased to 78, generalizability coefficient ($E\rho^2$) became 0.97. At this point the students are ranked ordered relatively. This showed that an increase in the number of teachers ranked ordered students according to their relative ability in the 2019 WAEC senior secondary school examination scores.

5. Discussion of Findings

The study revealed that the highest contribution to measurement error in examination scores was the student × teacher (invigilators), accounting for 68.9% of the total variability. The second largest source of variation to measurement error was the student factor $\sigma^2 s$ (due to differences among students) variance which accounted for 27.5%. This implied that the items did distinguished somehow among the students. The 3rd highest contributor to measurement error was the residual $\sigma^2 sti, e$ which acounted for 3.6% of the total variability. This showed that a proportion of the variance was due to the interaction of students × items × teachers and other unsystematic or systematic sources of variance that most probably were not measured in the study.

Also, the variance due to items did set to zero, the variance associated with the interaction of students × items yielded no source of variance. Therefore, the findings suggested that while the items were on average, comparable in difficulty, they were not uniformly difficult for all students in examination. However, by contrast, the variance due to the variance components $\sigma^2 t, \sigma^2 i$, $\sigma^2 si$, $\sigma^2 ti$ were all set to zero, which suggest that items, teachers (invigilators) and the interactions of items × teachers contributed minimally or did not contributed at all to variability in examination scores. From the findings, the interaction of students × items (questions) which ought to be the most important contributions to measurement error in an educational context, contributed very little to measurement error in examination scores. This showed that the magnitude of an effect is not inferred from statistical significance in generalizability theory. More so, the interaction of students × teachers (invigilators) which yielded the highest contribution to measurement error could be attributed to the fact that the teachers were the ones who scored the students. The findings of this study was supported by Shavelson, Baxter and Gao in Egbulefu [8] who reported that person × task interaction contributed immensely in performance based assessment. This study was also in line with the findings of Hintze and Pettite [11] on performance-based assessment. This study was in agreement with Lombardi et al [16] who indicated that the generalizability analyses of the Koppitz scores, shows that the variance components for rater and the interaction with both person and occasion were very small, suggesting that very little measurement error was associated with raters. But the result of the study was not supported by Shavelson and Webb [20] and Egbulefu [8] whose studies revealed that the residual was the highest contributor to measurement error.

The study further showed that an increase in the number of teachers to 78, rank ordered students according to their relative standing or ability in examination. A change in the number of teachers from 38 to 78 increased the generalizability coefficient ($E\rho^2$) to 0.97. This result is in conformity with the findings of the study by Webb, et al [20] who opined that more observers, gave a generalizability coefficient that rank ordered person in generalizability study of job performance measurement of Navy Machenist Mates. In the same vein, the result agreed with Brennan [4] whose study showed that with multiple raters, it was possible to differentiate between persons (Coeff. G relative 0.91). Therefore, to achieve a generalizability coefficient ($E\rho^2$) 0.91, the number of teachers were increased in order to rank order students relatively in examination.

6. Conclusions

Generalizability theory provides an integrated frame work for evaluating multiple sources of variability in examination scores and for deriving implications for test-development and test scores interpretation. Apart from the student factor, other sources (facets) affects the scores students obtain in examination. In this study, the interaction of students and teachers contributed to measurement error. Also, the interaction of student and teachers had a large effect to score dependability in examination. Above all, an increase in the number of the facet -teachers (invigilators) showed that with a high generalizability coefficient ($E\rho^2$), was high enough to rank order student relatively. To minimize error and minimize reliability in examinations, there is the need to estimate as many sources of error.

7. Recommendations

Based on the findings of this study, the following recommendations were made.

Generalizability analysis should be carried out by test developers and examination bodies in the estimation of reliability so as to estimate multiple sources of error and to reduce or eliminate measurement error and hence maximize reliability.

In generating items, item writers should endeavor to develop items that will discriminate among students of different achievement levels. This will in no small way reduce error in measurement.

References

[1] Ameh, V. G. (2014). WAEC releases Nov/Dec 2014 results, says 70% of students failed. Daily Post Nigeria. Retrieved from http://www.google.com/amp/s/dailypost.ng/2014/12/19/waecreleases-nov/dec-2014-results-says-70-students-failed/%3famp.

[2] Bamidele, S. T., Gana, A. Y., Kehinde, A., & Adekunle, A. R. (2021). Estimating generalizability and dependability indices of students' scores in teaching practice assessment in a Nigerian College of Education. *Sapientia Foundation Journal of Education, Sciences and Gender Studies (SFJESGS)*, 3 (2); 7 – 15 ISSN: 2734-2522 (Print); ISSN: 2734-2514 (Online).

[3] Bichi, A. A, Suleiman, A. H & Ali, H. (2019) Students' achievement in Mathematics: Analyzing the influence of Gender and School Nature. *Contemporary Educational Researches journal*. 9 (3) 50-56.

[4] Brennan, R. L (2001) *Generalizability Theory: Statistics for Social Science and Public Policy.* Springer-Verlag Berlin Heidlberg. New York.

[5] Cetin, B., Guler, N., & Sarica, R. (2016). Using generalizability theory to examine different concept map scoring methods. *Eurasian Journal of Educational Research*, 66, 212-228 http://dx.doi.org/10.14689/ejer.2016.66.12

[6] Eckes, T. (2017). Guest editorial rater effects: Advances in item response modeling of human ratings–Part I. Psychological Test and Assessment Modeling, 59 (4), 443–452.

[7] Edu G. English program, IRDP. Institut de recherche et de documentation pedagogique. Accessed from https://www.irdp.ch/institut/english-program-1968.html on 05.08.2021.

[8] Egbulefu, C. A. (2013). Estimating measurement error and score dependability in examinations using generalizability theory. (Unpublished doctoral dissertation) University of Nigeria, Nsukka.

[9] Esomonu, N. P., & Okeaba, J. U. (2021) Estimating Measurement Error and Score Dependability of the Inventory for Students' Integration into the University Academic Culture (ISIUAC) Using Generalizability Theory. *Rivers State University Journal of Education (RSUJOE)*, 24 (1): 35-46.

[10] Güler, N. (2009). Generalizability Theory and Comparison of the Results of G and D Studies Computed by SPSS and Genova Packet Programs. *Education and Science*, 34, 154.

[11] Hintze, J. M. & Pettite, H. A. (2001). The Generalizability of CEM Oral Reading Fluency Measures Across General and Special Education. *Journal of Psycho Educational Assessment*, 19 (1), 52-68.

[12] Iheanyichukwu, O. R. & Orluwene, G. (2020). Application of Generalizability Theory in Estimating Variance Components in National Examinations Council Problem Solving Questions in Mathematics. *European International Journal of Science and Technology*, 9 (4), 61-69.

[13] Johnson, S. Dulanay, C. & Bank, K. (2000). *Measurement error.* Retrieved from http:/www.wcpss.net/evaluationresearch/reports/2000/mment_error.pdf

[14] Kaya Uyanik, G., & Guler, N. (2016). Investigation of concept map scores' reliability: Example of crossed mixed design in generalizability theory. *Hacettepe University Journal of Education,* 31 (1), 97-111.

[15] Lee, Y. W. (2005). Dependability of scores for a new ESL speaking assessment consisting of integrated and independent tasks. *Language Testing*, 23 (2), 131-166.

[16] Lombardi, Seburn, Conley and Snow (2010) A Generalizability Investigation of Cognitive Demand and Rigor Ratings of Items and Standards. Educational Policy Improvement Center Eugene, Presented at the annual conference of the American Educational Research Association Denver, National Bureau of Statistics (2019).

[17] National Bureau of Statistics, 2019.

[18] Ogidi, R. C (2021) Application of generalizability theory in the estimation of variance components in national examination council essay questions in Christian religious studies in Ogba/Egbema/Ndoni local government area of Rivers State, Nigeria. *European Journal of Research and Reflection in Educational Sciences*, 9 (2): 1-8.

[19] Omorogiuwa, O. K. (2019). *An Introduction to Educational Measurement and Evaluation*. Benin City: Mase-Perfect.

[20] Shavelson, R. J., & Webb, N. M. (1991). *Generalizability Theory*: A Primer. USA: SAGE Publications.

[21] Strube, M. J. (2002). *Reliability and generalizability theory*. In L. G. grimm & P. R. Yarnold (Eds.), Reading and understanding more multivariate statistics (pp. 23-66). Washington, DC: American Psychological Association.

[22] Webb, N. M., Shavelson, R. J., & Haertel, E. H. (2007). *Reliability coefficients and generalizability theory.* Handbook of statistics 26: Psychometrics (C. Rao and Sinharay (Eds.) 81, 1-124, the Netherlands: Elsevier.