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

Ogidi, Reuben .C.
Department of Educational Psychology, Guidance and Counselling
Faculty of Education, Ignatius Ajuru University of Education
Rumuolumeni, Port Harcourt, **NIGERIA**
reubenogidi90@gmail.com

ABSTRACT

The study utilized generalizability theory in estimation of variance components in National Examination Council Examination Council Essay Questions in Christian Religious Studies. The design of the study was two facet fully crossed G-study and D-study. The population of the study was 1228 senior secondary schools three (SS3) in Ogba/Egbema/Ndoni Local Government Area of Rivers State, Nigeria. Cluster sampling method was used to select 175 students for the study. The instrument for data collection was 2020 August/September National Examination Council in Christian Religious Studies Essay Questions. Data obtained was analyzed using computer software SPSS through General Linear Model via variance components MINQUE method. Results of the study showed that the largest contribution to error is person by item by rater (∂^2pir) (5.244) with the percentage variance of (46.654%) and person by item (∂^2pi) (4.361) with a percentage variance of (38.43%), person by rater (∂^2pr) (0.425) with a percentage variance of 3.76%, person (∂^2p) (0.383) with a percentage variance of 3.63%, rater (∂^2r) (0.77) with a percentage variance of 7.36%, item (∂^2i) (0.028) with a percentage variance of 0.24%, item by rater (∂^2ir) (-0.06*) with a percentage variance of -0.74%. Also, the universe score was 5.6466, relative error variance was 5.4381, absolute error variance was 0.4383. G-study coefficient was 0.5213 and index of dependability of 0.938 was also obtained which indicated that the instrument was adequate for certificate examination. Recommendations such as there should be more awareness on the use of generalizability theory or analysis of examinee scores for psychometric properties estimate of the instruments used by examination bodies amongst others was made.

Keywords; Christian Religions Studies, Generalizability Theory, Essay Test, Independability, Reliability.

INTRODUCTION

Christian Religious Studies is one of the subjects offered at the senior secondary school level of the educational system in Nigeria. It is also one of the subjects offered in the National Examination Council Examination, often written to mark the end of the senior secondary education in the country. The subject has always played a crucial role in the Nigerian society. One of the relevance of the subject is that it enhances the opportunities for the full development of students' cognitive, affective and psychomotor domains. The subject promotes an understanding of the Christian Religious Traditions, thought and practical life style among the students.

Christian Religious Studies enable students to be sensitive to the question posed by life and to the dimensions of mysteries and wonders that underlie human experiences (Dike, 2018). It serves as a foundation of moral and spiritual guidance for the students. In addition, it provides a sense of direction and ability to differentiate between what is right and wrong (NERDC, 2007). Christian Religious Studies is widely accepted as an important determinant of national development in modern society. The subject is a veritable tool geared towards the production of citizens who have attained an all-round personal development; educationally, morally, socially and spiritually. It helps in grooming citizens who are capable of carrying out the expected leadership assignment that would be entrusted to them in the future (Aboyomi, 2016).

Odeyemi (2019) posit that Christian Religious Studies have the following objectives: (i) to provide young people with new ways of coping with regard social change in a society that is constantly changing (ii) it is a useful tool in value-re-orientation. Thus, Christian Religious Studies helps to re-orient people to abandon preconceived ideas, prejudice and acts of nepotism. Such re-orientation will help to reposition society, and bring national unity and cohesion (iii) it provides opportunities to learn more about God and thereby develop their faith in God (iv) it equips young people and helps them in ensuring high level of morality in the society (v) it helps in restoring social and religious order in the society.

Christian Religious Studies can be combined with other subjects to further the career aspirations of the students. Thus, career or courses such as law and other courses in the Humanities requires a proficiency in Christian Religious Studies. Okunonyo (2019) assert that the moral decadence in the society which hampers the rapid development of the society, necessitate the relevance of Christian Religious Studies. It is a tool that enhances national rebirth. The current hue and cry of corruption in the society can be minimized by the effective implementation of Christian Religious Studies in the school system. It is essential in developing a new society where issues of crime will be reduced while emphasizing the value of human lives.

The Christian Religious Studies have two papers, the objective and the essay. This study concerns the essay part of the Christian Religious Studies in the National Examination Council. Inkor-Tariah and Ogidi (2017) posit that essay test requires testees to put their thoughts together and write down their responses in sentences that can at times take pages. They further explained that this type of test gives the opportunity to assess students' mastery not just of the content, but also the language of expression. Chikwe (2017) assert that essay test are tests in which students respond to test items in their own language or expression. In other words, in essay tests, students are allowed to express themselves freely in their answers to test items. Onunkwo (2002) affirm that essay tests allow students to organize their answers in a logical and comprehensive manner, and students are not meant to select from given answer options but are to supply the correct answers. Ukwuije and Opara (2012) suggest that essay test is the traditional type of examination in which the students are to describe, define, compare, contrast, illustrate, classify, enumerate or state. Asuru (2015) explain that essay test requires testee to supply rather than select the answer to a test item. He further explain that it is also called the free response or free answer test because it allows the testee to organize, integrate, synthesize and present his/her answers in series of inter-connected sentences. Thus, the testee is allowed to express himself in series of logically arranged sentences, in his own language. Anikweze (2018) posit that essay test measures the non-structured types of learning such as creative writing, critical thinking, problem solving, imagination and organizational ability. Also, Orluwene (2012) affirm that essay test is the type of test items which allows the students the freedom to supply their own responses, rather than select the correct answer. She further

explain that in essay test, students are required to organize and integrate information, interpret information, give arguments, give explanations, evaluate the merit and demerit of ideas, and conduct other types of reasoning that tap complex thinking. Essay test is one of the assessment tools in measuring students' achievements in several school subjects. The relevance of essay test is that usually, it measures higher levels of the cognitive domain which provide critical thinking and originality in students (Ogunka & Orluwene, 2019). Linn et al (2005) affirm that essay test items provide the freedom of response that is needed to adequately assess the ability of a student to formulate problems, organize, integrate and evaluate ideas and information and as well apply knowledge and skills. For information from essay test to be used in decision making about the students in Christian Religious Studies, there is the need to ascertain the dependability of behaviour measurement. Thus, the scores which form the basis of taking decision about the learner should reflect the ability or performance of the student. Also, the instrument from which scores are generated for making decisions about the students should be reliable. This consideration warranted the use of generalizability theory in this study.

Generalizability theory is a very important concept in the field of Measurement and Evaluation. It is also called G-theory. Falaye (2019) posit that generalizability theory is a statistical theory that is used for the evaluation of the reliability or dependability of behavioural measurements. Orluwene (2012) assert that it estimates multiple sources of measurement error and permit decision maker to design a measurement procedure that minimizes the error. Brennan (2001) suggest that, it is a statistical framework for conceptualizing, investigating and designing reliability observation.

Generalizability theory acknowledges that the reliability of an observation depends on the universe about which the researcher or measurement expert wants to draw inferences. This is because a particular measure may conceivably be generalized to many different universes, a measure may vary in how reliably it permits inferences about these universes and therefore, be associated with different reliability coefficients (Essien & Nbame, 2020).

Kin and Wilson (2009) posit that dependability of behavioural measures is the accuracy of generalizing from a person's observed score on a measure, or a test to the score that the person who have received average over all possible conditions. Such variation that are as a result of the measuring instrument rather than variables which are directly controlled by the examinee indicates uncertainty in the quantitative description of the individual on the basis of the measuring instrument. Shavelson and Webb (2005) assert that generalizability refers to the accuracy of generalizing from a person's observed score on a test or other measure to the average score that person would have received under all the possible conditions that the test user would be equally willing to accept.

Shavelson and Webb in Essien and Nbame (2020) identified the following assumptions of generalizability theory (i) the data examined in a generalizability analysis should be at interval level of measurement. In other words, data should be continuous (ii) generalizability theory assumes that a students' observed score is comprised of his/her universe score and/or more sources of error (iii) the errors are assumed to be independent of the universe score and uncorrelated. This indicates that all of the effects in the measurement model are independent (iv) generalizability theory assumes that the samples used to estimate the error variances and selected of students, items or occasions and comprise random samples from their respective populations. In other words, the concept of "randomness" states that even though conditions of a facet have not been sampled randomly, the face may be considered to be random if conditions not observed in the study can be exchanged with the observed conditions. (v) generalizability

theory assumes that the standard errors are the same at all score levels. This shows that the same standard error of measurement is often applied to all objects of measurement regardless of the underlying universe score.

Shavelson et al in Ogunka and Orluwene (2019) affirm that the theory of genealizability focuses on the magnitude of sampling out errors due to person, item, rater and occasions among others. In addition, such interactions provides estimates of the magnitude of measurement error in the variance components and also provide a summary dependability coefficient which reflect the generalizing samples core or profile to the much larger domain of interest.

Genralizability theory is a refinement of the classical test theory. Asuru (2015) posit that classical theory is a theory about test scores that introduces three concepts viz:

Test score   =   X (observed score)
True score   =   T
Error score   =   E

The undifferentiated errors in the above explanation provide too gross a characterization of the potential and actual sources of measurement error. This error which is part of the observed score is called an error component.

The generalizability theory have been applied in various studies by scholars. Heitmen et al (2009) applied the geenralizability theory in estimating the reliability of Ankle in complex laxity measurement. Atilla (2012) investigated the dependability of job performance ratings based on the generalizability theory. Adegbile (2002) examined the relative effects of two model of advance organizer on performance in reading comprehension. Preuss (2003) used generalizability theory to develop clinical assessment protocols. Kim and Wilson (2009) carried out a comparative analysis of the ratings in performance assessment using generalizability theory. Also, Ikeh and Madu (2018) investigated the application of generalizability theory in estimating multiple sources of variation in Economics Essay test. Faleye (2019) demonstrated the use of one facet person (p) by item (i) design pxi design and two facets crossed design in carrying out generalizability theory analysis in CTBS Science Achievement Test. Ogunka and Orluwene (2019) applied the generalizability theory in the estimation of variance components in National Examination Council Essay questions in English Language.

The National Examination Council is a national examination particularly for senior school certificate students. It is therefore necessary that the construction, validation and administration of examinations conducted by such a body should have items that possess the psychometric properties or that such a body should carry out adequate psychometric analysis of items particularly in Christian Religious Studies in order to eliminate any unsystematic or systematic error especially in the instrument. Thus, this study apply the generalizability theory in the estimation of variance components in National Examination Council Essay in Christian Religious Studies.

Purpose of the Study
The main purpose of this study is to estimate the variance components of person by items by raters and scores dependability on the National Examination Council Essay Questions in Christian Religious Studies using the generalizability theory. The specific objectives of the study include the following:
i)   Investigate the relative contribution of person, item, rater and their interaction

ii)   Determine the dependability coefficient reliability of the essay questions in Christian Religious Studies.

Research Questions
This study answered the following research questions:
1.   What is the relative contribution of person, item, rater and their interactions?
2.   What is the dependability coefficient reliability of the essay questions in Christian Religious Studies?

Methods
The research design adopted in the study was the instrumentation with two-facet fully crossed G-study and D-study. This design was useful in minimizing undesirable variance and maximizing dependability reliability.

The population of the study was 1228 students in Senor Secondary School Three in the area. Cluster sampling method was used to select a sample of 175 students who were administered National Examination Council Essay Questions in Christian Religious Studies. The instrument for data collection was the August/September Essay Questions in Christian Religious Studies. The marking scheme was adopted by the raters. The data collected was analyzed. Computer software SPSS through general linear model via variance components MINQUE method aided the data analysis.

RESULTS
Table 1: Estimate of Variances Component on Generalized Linear Model Under Components (MINQUE Method)

Source	Type I Sum of Variances	Df	Mean Square	Variance Component	% Variance
Persons	2967.141	175	16.955	0.383	3.63
Items	92.211	5	18.442	0.028	0.24
Rater	58.518	1	52.518	0.77	7.36
Persons * Item	9853.168	875	11.261	4.361	38.43
Item * Rater	13.683	5	2.737	-0.06*	-0.074
Persons * Rater	1243.429	175	7.105	0.425	3.76
Persons * Item * Rater	4588.482	875	5.244	5.244	46.654
Error (Residual)	0.00	0			
Total	**18816.637**	**2111**		**11.151**	**100**
Components (MINQUE METHOD)					

Table 1 shows the estimated variance components and its interactions. The result indicates that the largest contribution to measurement error is person by item by rater (∂^2pir) (5.244) with percentage variance of 46.654. This is followed by person by item (∂^2pi) (4.361) with percentage variance of 38.43. Also, the person by rater (∂^2pr) (0.425) with percentage variance of 3.76, person (∂^2p) (0.383) with percentage variance of 3.63, rater (∂^2r) (0.77) with percentage variance of 7.36, item (∂^2i) (0.028) with percentage variance of 0.24 and item by rater (∂^2ir) (-0.06*) with a percentage variance of -0.074. The result of the study showed that the estimation of various components interacted very highly among person, item and rater (∂^2pir).

Table 2: estimate of G and D Reliability Coefficient

Relative Error Variance	Absolute Error variance	Universe Score	G-study Coefficient	Index of Dependability
5.4381	0.4384	5.6466	0.5213	0.938

Table 2 reveals the estimation of G and D study reliability index where; the universe score was 5.6466, relative error variance was 5.4381 and the absolute error variance was 0.4383. In addition, the G-study coefficient was 0.5213 and the index of dependability 0.938 were generated from the scores.

Discussion of Results

The results of the study indicated that the highest contribution to measurement error from the score generated was on the person, item and rater (∂^2pir) (5.244, 46.654%). The result showed that a proportion of the variance was as a result of the interaction of person by item by rater (∂^2pir). Nonetheless, this large variance component observed in this study was not only in relation to persons, items and raters but also to the undifferentiated error. The result also showed that the second largest source of variance is person by item (∂^2pi) (4.361, 38.43%). This is followed by person by rater (∂^2pr) (0.425, 3.76%). Again, the relative variation due to person (∂^2p) (0.383, 3.63%), followed by the rater (∂^2r) (0.77, 7.36%) and then item (∂^2i) (0.028, 0.24%). The item by rater (∂^2ir) variance component has (-0.06, -0.074), showing that error due to item by rater yielded negative estimate (-0.06) as a result of the degree of freedom for the residual (error) which was zero.

In addition, the estimate of the G and D study of generalizability theory was realized. The relative error of variance was 5.4381, the absolute error variance was 0.4384 while the universe score was 5.6466. On the other hand, the G-study coefficient was 0.5213 while the index of dependability was 0.938. Interesting, the estimate of the index of dependability which indicated the reliability coefficient of the NECO Christian Religious Studies August/September 2020 was 0.94, which was deemed to be high and adequate for the instrument to be used for certificate examination.

CONCLUSION

The result of the generalizability analysis showed that the highest contribution of variance components generated is from person by item by rater, followed by person by item, rater, person by rater, person and then item. However, the item by rater showed a negative estimate. This is due to the fact that the degree of freedom for the residual was zero. In addition, the index of dependability is high and reliable showing that the instrument Christian Religious Study in NECO August/September, 2020 was good to be used for certificate examination.

RECOMMENDATIONS

Based on the results of the study, the following recommendations were made:
1. There should be more awareness on the use of generalizability theory/analysis of examinees scores for psychometric properties estimate of the instruments used by examination bodies.
2. Test experts should be trained on the use of generalizability theory
3. Test developers should also subject the scores generated from the instruments developed by them to generalizability analysis to further determine the psychometric properties of such instruments.

REFERENCES

Ababomi, K.J (2016). *Christian Religious Studies Curriculum.* Capgrace publishers.

Adegbile, J.A (2002). The relative effects of two model of advance organizer on performance in reading comprehension. *Journal of Pedagogy and Educational Development, Rivers State College of Education, Port Harcourt, 9*(2), 22-36.

Anikweze, C.M (2018). *Measurement and evaluation for teacher education (4th edition).* Shiloh press associates.

Asuru, V.A (2015). *Measurement and evaluation in education and psychology* (2nd edition). Pearl publishers.

Atilla, Y. (2012). Dependability of job performance ratings based on generalizability theory. *Journal of Education and Science, 163*(37), 157-179.

Brennam, R.L (2001). Generalizability theory. Springer vering.

Chikwe, C.K (2017). *Fundamentals of test, measurement and evaluation in education.* Emmanest ventures.

Dike, M.G (2018). Christian Religious Studies for national development. *Journal of Religious Studies, 3*(2), 39-52.

Essien, E.J. & Nbame, L. (2020, Dec. 16th). Generalizability theory. A seminar paper presented to the Department of Educational Psychology, Guidance and Counselling, Ignatius Ajuru University of Education. Port Harcourt, Nigeria.

Faleye, B.A (2019). Generalizability theory. *Nigerian Journal of Educational Research and Evaluation, 18*(2), 1-12.

Heitman, R.J., Kovaleski, J.E & Pugh, S.F (2009). Application of generalizability theory in estimating the reliability of Ankle- complex laxity measurement. *Journal of Athletical Teaching*, https://www.ncbi,nih.gov.

Ikeh, E.F & Madu, B.C (2018). Application of generalizability theory in estimating multiple sources of variation in economics essay test. A paper presented at National Annual Conference of Association of Educational Researchers and evaluators of Nigeria held at Abuja.

Inko-Tariah, D.C & Ogidi, R.C (2017). *Fundamental of psychological testing for psychologists, counsellors and educationists.* Rodi printing and publishing company.

Kim, S.C & Wilson, M. (2009). A comparative analysis of the ratings in performance assessment using generalizability theory. *Journal of Allied Measurement, 10*(4), 408-423.

Linn, R.L., Miller, M.D. & Gronlund, N.E (2005). *Measurement and assessment in teaching (9th edition)*. Pearson Prentice Hall.

NERDC (2007). Christian Religious Studies Curriculum. Government press.

Odeyemi, B.H (2019). *Christian Religious Studies Education.* Toti publishers.

Ogunka, R.I & Orluwene, G.W (2019). Application of generalizability theory in estiamtion of variance components in National Examination Council Essay questions in English Language.

Okunonyo, F.K (2019). Christian Religious Studies in modern Nigerian society. *Journal of Humanities, 3*(2), 55-68.

Onunkwo, G.I.N (2002). *Fundamentals of educational measurement and evaluation.* Cape publishers.

Orluwene, G.W (2012). *Fundamentals of testing and non-testing tools in educational psychology.* Harey publishers.

Preuss, R.A (2003). Using generalizability theory to develop clinical assessment protocols. *Physical therapy, 93*(4), 562-569.

Shavelson, R.J. & Webb, N.M (2005). *Generalizability theory: A primer.* Sage.

Ukwuije, R.P.I & Opara, M.I (2012). *Test and measurement for teachers.* Chadik printing press.