APPLICATION OF GENERALIZABILITY THEORY IN ESTIMATING VARIANCE COMPONENTS IN NATIONAL EXAMINATIONS COUNCIL PROBLEM SOLVING QUESTIONS IN MATHEMATICS

OGUNKA, RICHARD IHEANYICHUKWU[1] and DR. GOODNESS ORLUWENE[2]

[1,2]Department of Educational Psychology, Guidance and Counselling, Faculty of Education, University of Port Harcourt, Rivers State, Nigeria

Published: 29 April 2020

Cite this article: 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.

Abstract
Application of generalizability theory in estimation of variance components in National Examination Council Questions in Mathematics. Design is two facet fully crossed G – study and D – study. The population 723 senior secondary schools in Ikwerre Local Government Area. Simple random sampling technique via balloting to draw six schools out of fourteen senior secondary schools. Cluster sampling technique to sample 153 students. Instrument is 2018 November/ December National Examination Council Mathematics Problem Solving Questions. Data obtained analysed using computer software SPSS through General Linear Model via variance components MINQUE method. Result the largest contribution to error is student*item*rater (∂^2sir) (9.830) with percentage variance of 63.5 and student* item (∂^2si) (4.157) percentage variance of 26.7; item (∂^2i) (1.172) percentage variance of 7.5, rater (∂^2r) (.001) percentage variance 6.4, student (∂^2s) (.549) percentage variance 3.5, item*rater (∂^2ir). (-.062) percentage variance -0.4, student* rater (∂^2sr) (-.171) percentage variance -1.10. The universe score 148.704, Relative error variance 73.904, Absolute Error Variance 74.8, G- study coefficient 74.904 and Index of dependability 75.8 obtained. Recommendation, though these items are very difficult for the sampled students. Generalizability theory should be used in psychometric properties estimate to generate reliable test items.

Key Words: Variance, Estimating, Generalizability theory, Problem Solving.

Introduction
Mathematics is a core subject considered paramount in our Nigeria primary and secondary schools. The subject is taught with the aim of preparing students for the task ahead in future within the society they live for business and further studies. Hence the need for teachers to possess the right professional quality to help the students to understand the subjects right from elementary level. Teachers have been trained through strengthening Mathematical and education regarding to the aspect of change in professional quality from old method of teacher centeredness to student centred approach which involve using the new approaches activity, student-centred, active and self-directing strategies. Federal Republic of Nigeria (2004) National Policy on Education makes mathematics a compulsory subject throughout all primary school and secondary school in Nigeria. Every day a child goes to primary school or secondary school in Nigeria, the child must study mathematics. Again Azuka (2013b) observed that learning becomes more perfect when theoretical knowledge is implemented by learning all possible life supporting skills which can be enhance through Activity-Based learning method in Mathematics and science.

Problem solving is a fundamental means of developing mathematical knowledge at any level. Problem solving gives students a context to help them make sense out of the mathematics they are learning. Problems can be used to introduce new concepts and extend previously learned knowledge. The definition is the process of working through details of a problem to reach a solution. Problem solving may include mathematical or systematic operations and can be a gauge of an individual's critical thinking skills. In mathematics, problem solving entails words often indicate an operation, common clues for addition problems includes Sum, Total, Perimeter: Common clues words for subtraction problems are Difference, How much more, Exceed; For multiplication problems common clues word found are Product, Total, Area, Times and in division problems common clues words like Share, Distribute, Quotient, Average. Although clues words will vary from problem to problem, you will then learn to recognise which words mean what in order to perform the correct operation. Lester

and Kehle(2003) summarize themes and methodological shifts in problem solving includes what makes a problem difficult for students and what it means to be successful problem solvers; studying and contrasting experts and novices' problem solving approaches; learners' metacognitive, beliefs system and the influence of affective behaviours; and the role of context; and social interactions in problem solving environments. Again Teaching strategies also evolved from being centred on teachers to the active student's engagement and collaboration approaches (NCTM 2000). Lesh and Zawojewski (2007) propose to extend problem solving approaches beyond class setting and they introduce the construct" model eliciting activities '' to delve into learners idea and thinking as a way to engage them in the development of problem solving experiences. To this end, learners develop and constantly refine problem solving competencies as a part of a community that promotes and values modelling construction activities. In this study, the researcher will make efforts to avoid the use of options because it is the assessment of levels of all the sampled population to estimating variance components in National Examination Council Problem solving Mathematics Questions.The senior secondary school certificate examinations conducted by National Examinations Council have shown that many students fail ridiculously in Mathematics papers. Issues that bother the teachers, the public, parents and other stakeholders is mass failure rate which they blame on the National Examinations Council as it concern construction, validation, and administration of examination conducted by them that lack adequate psychometric analysis of its items; this research is needful based on public complain and perception on the recorded mass failure, hence development of a measuring instrument that accurately measures a particular characteristics of the examinee in the best perfect manner without any unsystematic or systematic error both in the instrument and characteristics under investigation is bound to pose problems to the examinee. It is only when the instrument measure in perfect manner that such unquestionable quantitative descriptions of the examinee in terms of the exact extent to which it possess and demonstrate the trait can be adjudge for the best in relation to decisions to be subsequently taken for reliable calculation, problem solving and objective item in Mathematics. Hence this investigation seek to validate problem solving questions in mathematics 2018 November\ December national examination council.

Generalizability (G) theory is a statistical theory about the dependability of behavioural measurement, Cronbach,Gleser, Nanda, and Rajaratnam (1972). The score (on a test or other measures) on which the decision is to be based is only one many scores that might serve the same purpose. Kin and Wilson(2009) refers dependability of behavioural measures as 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 averaged over all possible conditions. This type of variation that is mainly due to the measuring instrument rather than factors which are directly controlled by the examinee denotes uncertainty in the quantitative description of the individual on the basis of the test. The unsystematic error or fluctuation in the individual's scores over several repeated testing's mean that in the behavioural sciences, one cannot completely depend on the single score obtained by each student on an attribute that was measured once.

According to Shavelson and Webb(1991) dependability refers to the accuracy of generalizing from a person's observed score on a test or other measure (behaviour observation, opinion survey) to the average score that person would have received under all the possible conditions that the test user would be equally willing to accept. This notion of dependability is the assumption that the person's knowledge, attitude, skill, or other measured attribute is in a steady state; it is assume that any differences among scores earned by an individual on different occasions of measurement are due to

one or more sources of error, and not to systematic changes in the individual due to maturation or learning. Orluwene (2012) indicated that, in the measurement of complex traits imperfect instruments are used so that the score observed for each person almost always differs from person's true ability or characteristics; she further affirmed that the discrepancies between the true ability and the observed ability results from measurement error, which implies some inaccuracy in the measurement exist because measurement error may inflate or depress any subject's score in an unpredictable or predictable manner. To satisfactorily control the whole issue raise above on the problem of incomplete certainty in and dependence on a score obtained by an individual from a single administration of a test for accurate quantitative description of the person with respect to a given psycho- social construct, the researcher is required to empirically establish the reliability of the measuring instrument which he or she develops. The comparism of dependability of reliability in generalizability theory and classical test theory to determining standard error measurement varies. Atilla(2012) asserted that the use of classical test theory approaches to determining score reliability, however, are not capable of identifying and untangling this profusion of error which classical reliability was not conceptualized to do since it account for only one source of error at a time. Similarly, Ikeh and Madu cited Tavako and Brennan (2013) states Classical Test Theory (CTT), assume the student's true score is the sum of the student's observed score and a single undifferentiated error term. Kpolovie (2010) asserted classical theory as reliability embedded in the true- score and error- score model defines reliability as the coefficient that predictable proportion of variance in observed scores from the true scores.

Generalizability Theory liberalizes classical test theory by employing ANOVA methods that allows an investigator to untangle multiple sources of error that contribute to the undifferentiated E in classical test theory. But is important to assert that GENOVA, SAS, SPSS and Edu-G program are computer software used for statistical analysis, data mining and predictive analysis. In this study SPSS computer program via General Linear Model on Variance Components (MINQUE) method was fully adopted for the estimating the variance components for generalizability theory. Generalizability Theory is a statistical theory for estimating the reliability of behavioural measurements which offers researchers an opportunity to assess comprehensively sources of measurement error (variance components). G-Theory concerns itself about the relative and absolute dependability of behavioural measures. The goal of this study is to design an assessment protocol and in most other cases as well, the data should represent a random sampling of conditions. In other words, the measurements within the data set should be considered interchangeable with measurements of the same parameters taken from others members of the target population. In this therefore, a random-effects (random-facets) model of ANOVA should be used. There are two facets item and rater for this study design. The steps we need is a factorial model for the ANOVA. Specifically, this study design requires a two facets, random-effects, tworaters measures model of the ANOVA, to be run for each of the measures of interest. The results of the variance components that can be attributed to the object of measurement, the study facets, and the interactions.

The empirical works that support this investigation includes, Atilla (2015) estimation of generalizability coefficient: application with different programs, Ikeh and Madu (2018) applications of generalizability theory in estimating multiple sources of variation in economics essay test, Heitman, Kovaleski and Pugh (2009) Application of generalizability theory in estimating the reliability of Ankle- Complex Laxity Measurement and Preuss (2003) Using generalizability theory to Develop Clinical Assessment Protocols. Further generalizability theory design that deal with items – crossed – with raters in the universe of admissible observations. This means that any one of the Ni items might

be rated by one of the Nrraters. If the population is crossed with this universe of admissible observation, then the corresponding G-study design is p X i X r, in which the responses of Np persons by Ni items are each evaluated by Nrraters (Brennan 2001). Therefore, estimating the variance components for this fully crossed design can be used to estimate results for any possible two facet design. The theory of generalizability focuses on the magnitude of sampling out errors due to person, item, rater and occasions among others, then their interactions which provides estimates of the magnitude of measurement error in the variance components as well provide a summary dependability coefficient reflecting the generalizing a sample score or profile to the much larger domain of interest (Shavelson, Baxter and Gao, 1993).

This study seeks to investigate person by item by rater variance of measurement error in the problem solving questions in mathematics National Examination Council. Johnson, Dulanay and Banks (2002) asserted measurement error is a situation in which students true ability is either underestimated or overestimated. Hence, the need of estimating measurement error is inevitable because the inconsistencies that exist in measurements instruments is enormous especially variance components.

The aim of this study is to estimate the variance components of persons by items by raters and scores dependability in problem solving question in mathematics National Examinations Council with the application of generalizability theory.

Research Question:

1. What are the relative contribution of person, item, rater and their interactions in National Examinations Council Mathematics Problem solving question?
2. What is the generalizability coefficient in National Examination Council Mathematics Problem solving question?
3. Index of dependability of National Examinations Council Mathematics Problem solving question?

Methods

The design is two – facet fully crossed G – Study and D- Study. D- Study generalizability theory uses the information obtained from G-Study to determine the measurement procedure in minimizing undesirable variance and maximize dependability reliability.

The population is made up of 723 senior secondary school three students in Ikwerre Local Government Area of Rivers State, Nigeria. Simple random sampling technique via balloting used to draw six schools out of fourteen senior secondary schools. Cluster sampling technique used to sample 153 SS3 Students and was administered the problem solving question in mathematics National Examinations Council.

Instrument for data collection 2018 November/December Problem Solving Questions in Mathematics National Examinations Council, each item is rated on 10 points marking scheme which was adapted by the raters.

The data collected was analysed using a computer software SPSS through General linear model under Variance components MINQUE method

Results

1. What are the relative contribution of person, item, rater and their interactions in National Examinations Council Mathematics Problem solving question?

Estimating the Variances component on Generalized Linear model under variance components (MINQUE METHOD)

Table 1: The result presented the estimated variance components and its interactions.

Source	Type 1 sum of variances	Df	Mean. Square	Variance component	% variance
Student	3463.484	152	22.786	.549	3.5
Item	1469.063	4	367.266	1.172	7.5
Rater	.628	1	.628	.001	6.4
Item* Rater	1.604	4	.401	-.062	-0.4
Student *Item	11032.137	608	18.145	4.157	26.86
Student * Rater	1364.472	152	8.977	-.171	-1.10
Student* Item* Rater	5976.796	608	9.830	9.830	63.5
Error (Residual)	0.00			15.47	

The result shows that the largest contribution to measurement error is student*item*rater (∂^2 sir) (9.830) with percentage variance of 63.5 and it is followed by student* item (∂^2si) (4.157) percentage variance of 26.8; item (∂^2i) (1.172) percentage variance of 7.5, rater (∂^2r) (.001) percentage variance 6.4, student (∂^2s) (.549) percentage variance 3.5, item* rater (∂^2ir). (-.062) percentage variance -0.4, student* rater (∂^2sr) (-.171) percentage variance (-1.10). Student*item*rater contributed very high in variances component.

Question 2. What is the generalizability coefficient in National Examination Council Mathematics Problem solving question?

Question 3. Index of dependability of National Examinations Council Mathematics Problem solving question?

Table 2: The result shows estimation of G and D study reliability coefficient

Relative Error variance	Absolute Error variance	Universe score	G-Study Coefficient	Index of Dependability
73.904	74.8	148.704	74.904	75.8

The universe score 148.704, Relative error variance 73.904, Absolute Error Variance 74.8, Generalisability coefficient 74.904and Index of dependability 75.8, were obtained from the scores.

Discussion

The result revealed that the largest contribution to measurement error from the score obtained is on the student*item*tater ∂^2sir (9.830, 63.5%) the result indicated that a proportion of the variance was due to the interaction of student, item and rater (∂^2sir). However, this large variance component observed is not only in relation to student but also to undifferentiated error. The second largest source of variance is student and item ∂^2si(4,157, 26.8%) Third, item ∂2i (1.172. 7.5%). Then rater∂2s (.001, 6.46%) followed by student∂2s (.549, 3.5%) Again item* rater ∂2ir (-.062, -0.4%) Lastly the item*rater ∂ir variance component has (-.171,-1.10%) indicating that error due to item by rater and student*rater sryielded negative estimate (-.62, -0.4% and -.171, -.1.10%) respectively, as a result of the degree of freedom for the residual (Error) amounted to zero. However the concern with variance component estimation is when a negative estimate arises because of sampling error or model misspecification, the possible solution is to set negative estimate to zero but use of negative estimates in expected mean square equations for other components (Brennan 2001).

Interestingly, the estimation of the G and D study of generalizability theory was achieved by using the universe score. A generalizability coefficient is the ratio of universe score variance to itself plus relative error variance, and Index of dependability is the ratio of universe score variance to itself plus absolute error variance (Brennan 2001). The universe score 148.704 on itself plus relative error variance estimate G. study coefficient 74.904. Again the universe score of 148.704 on itself plus Absolute Error variance estimate Index of dependability 75.8Therefore, the index of dependability of the instrument is 75.8 which is 76 percent

Conclusion

Based on the results of the generalizability analysis, the largest contribution of variance components obtained is from student, item and rater followed by student and item, then by item. Againrater, while item and rater, student and rater indicated a negative estimate as a result of the degree of freedom for the residual which amounted to zero. Hitherto, the Index of dependability is high and reliable based on the performance of the sampled population and the instrument could be use for certificate examinations.

Recommendation

This study is basically to reduce the influence of variance components that arise from scores by students. Generalizability Theory/analysis should be used to subject students' scores for psychometric properties estimate in order to generate reliable test item for examination bodies

Reference

[1] Atilla, Y. (2012). Dependability of job performance ratings according to generalizability theory. Journal of Education and Science, 163(37), 157 – 348.

[2] Atilla, Y (2015). Estimation of generalizability coefficient: an application with different program.

[3] Adegbile, J. A (2002). 'The Relative Effects of two model of Advance organiser on performance in Reading Comprehension '. In Journal of pedagogy and Educational Development, Rivers State College of Education, Port Harcourt.Vol.9, No 2.pp, 22-36.

[4] Azuka, B.F.(2013b) Activity- based learning strategies in the mathematics classrooms. Journal of Education and Practice,4(13)

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

[6] Brennan, R.L. (2001). An essay on the history and future reliability from the perspective of Replications. Journal of Educational Measurement, 38, 295 – 317.

[7] Cronbach, L.J, Gleser, G.C., Nanda, H. &Rajaratnam, H. (1972). The dependability of behavioural Measurements. New York: Wiley.

[8] Federal Republic of Nigeria (2004) National Policy on Education (4th ed.) Lagos NERDC Press.

[9] 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

[10] 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 (Abuja 2018). Association of Educational Researcher and Evaluators of Nigeria.

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

[12] 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.

[13] Kpolovie, P. J (2010). Advanced Research Methods, Owerri Springfield Publisher ltd.

[14] Lester, P. E and KehleF.K (2003). From problem-solving modelling: The Evolution of thinking about research on complex mathematical activity. In R. Lesh and H.M. Doerr (Eds) Beyond constructivism: Models and modelling perspectives on mathematics teaching, learning and problem solving. Mahwah, NJ: Lawrence Erlbaum Associates, Inc

[15] Linn, R. L., Miller, M. D. &Gronlund, N. E. (2005). Measurement and Assessment in teaching 9th Edition USA Pearson Prentice Hall.

[16] Shavelson, R. J. & Webb, N. M. (2005). Generalizability Theory https://www.researchgate.net>publication

[17] Shavelson, R. J., Baxter, G. P., &Gao, X. (1993). Sampling Variability of Performance Assessments. Journal of Educational Measurement, 3(30), 215 – 232. Retrieved from http://links.jstor.org/sici?sici=00220655%28199323%2930%3A3%3C215%3ASVOPA%3E2.0.C0%3B2-M

[18] Preuss, R.A. (2003). Using generalizability theory to Develop Clinical Assessment Protocols. Physical Therapy, vol. 93 issue 4, pp 562-569. Canada

[19] Zawojeewski, J. S and Lesh, R.(2007) A models and modelling perspective on problem solving. In R,Lesh and H.M. Doerr (Eds) Beyond constructivism: models and modelling perspectives on mathematics teaching , learning and problem solving. Mahwah,N.J: Lawrence Erlbaum Associates,Inc.