
Sources of Variance Component in Determining Reliability of West African Examination Council Essay Questions in Chemistry Using the Generalizability Theory

NWUCHEGBUO, BLESSING IHEANYICHUKWU PhD

Department of Educational Psychology, Guidance and Counselling,

Faculty of Education

Ignatius Ajuru University of Education Port Harcourt

iheanyiblessing1@gmail.com

Abstract

The study examines sources of score variance component in West African Examination Council (WASSCE, 2020) chemistry essay questions. The design is two facet fully crossed G–study and D–study. The population is made up of eight hundred and sixty-seven (867) senior secondary three (SS3) students in Etche Local Government Area of Rivers State. Purposive sampling technique was used to select eight (8) senior secondary schools out of sixteen senior secondary schools. Cluster sampling technique was used to sample one hundred and sixty-seven (167) Senior Secondary three Students (SS3) and the West African Examination Council (WASSCE 2020) chemistry question was administered the in their various schools. Instrument for data collection 2020 May/June chemistry questions West African Senior School Certificate Examination (WASSCE). Each item is rated on 10 points marking scheme, which was adapted by the raters. The data collected was analyzed using a computer software SPSS through General linear model under Variance components MINQUE method. The result shows students by item as the highest contribution to error variance (σ^2_{si} 1.755) followed by student (σ^2_s .141); rater (σ^2_r .064) students by rater (σ^2_{sr} .056) item (σ^2_i .034) and item by rater (σ^2_{ir} .005). G-study coefficient was 0.069 and Index of dependability was 0.064. The result showed that index of dependability was very low for West African senior school certificate examination hence the items were difficult. Recommendation includes that though; these items were very difficult for the sampled students. Generalizability theory should be used in psychometric properties estimate to generate reliable test items with high index of dependability.

INTRODUCTION

Generalizability theory (G) is a statistical or test theory about the dependability of behavioural measurement, (Cronbach, et al cited in Ogungka and Orluwene, 2019). Generalizability theory allows researchers to answer such questions as; is the sampling of a task or judges address the major sources of error? Can I improve the reliability of the measurement better by increasing number of tasks or the number of judges, or are some combinations of the two more effective? Are the test

scores adequately reliable to make decisions about the level of a student's performance for a certification? A test score or other measures in which the decision is based on is only one out of several scores that may serve the same purpose. Dependability of behavioural measures is the precision of generalizing from a student's observed test score in a measure or a test score of the person who have received averaged score in all possible conditions (Kin and Wilson, 2009). Variation that is as a result of the measuring instrument rather than factors that are directly controlled by the testee means uncertainty in the amount of description of the student on the basis of the test. The unsystematic error in the student's test scores over many repeated tasking means that in the behavioural sciences, one may not completely depend on the single test score obtained by each student in an attribute that is measured once.

According to Shavelson cited in Ogunka and Oriuwene, (2019), dependability refers to the precision of generalization from a student's observed score on a test or behaviour observation, opinion survey (other measures) to the average score that student may have scored under all the possible conditions that the test user may be equally accept willingly. Shavelson maintained that generalizability focuses on the magnitude of sampling out error as a result of person by item by rater and occasions etc. in addendum, the interactions give the estimates of the magnitude of error variance and provide a summary of dependability coefficient that reflects the generalizing samples representation of larger area of interest. This perception of dependability is the presumption that the student's knowledge, ability, attitude, skill, and other measured attribute is in a steady state. It is assumptive that any differences in scores obtained by a student in different occasions of measurement are due to one or more sources of error and not to systematic change in the individual due to maturation or learning. In generalizability theory, a behavioural measurement such as a test score is conceived as a sample from a universe of admissible observations. The universe is made up of all possible observations which decision makers may consider to be more acceptable substitutes (score sample on occasions 2 and 3) for observation in hand (scores in occasion 1). Every measurement situation has characteristic features in test form, test item, rater and test occasion and each characteristic feature is called a facet of a measurement, while a universe of admissible observations is defined by all possible combinations of the levels of the facets (items, occasions). Generalizability theory is a fineness of the classical theory. Classical theory is a theory about test scores that introduces three concepts; test score (X) (observed score), true score (T) and error score (E) (Asuru, 2015).

Comparison of dependability of reliability in generalizability theory and classical test theory to determine the standard measurement error varies. Atilla, (2012) stated that the use of classical test theory method in determining reliability of a score in whatever manner is not capable of identifying and disentangling the multiplicity of errors that classical test reliability could not conceptualized hence, classical test theory only accounted for one source of error at a time. Classical Test Theory (CTT) presumes student's true test score is the sum of the student's observed score and a single undifferentiated error term. Classical theory as reliability embedded in the true-score and error-score model defines reliability as the coefficient that is predictable proportion of variance in observed scores from the true scores (Kpolovie, 2010). Generalizability Theory liberalizes classical test theory by employing ANOVA, GENOVA, SAS, SPSS and Edu-G methods to separate multiple sources of error that contribute to the undifferentiated error in classical test theory. Generalizability theory (G) comprehensively assess sources of measurement error (variance components). G- Theory concerns itself about the relative and absolute dependability of behavioural measures.

Chemistry is the branch of science concerned with the substances of which matter is composed, the investigation of their properties and reactions, and the use of such reactions to form new substances. It is a very important branch of science. Chemistry is the study of science that deals with constituents of matter like atoms molecules, ions etc.; and its properties, structure, behaviour, and interactions among them. Since everything is made up of atoms and molecules, we can see the

chemistry all around us. Today, chemistry has grown into a very diverse field. There is a significant overlap between chemistry and other branches of science, for example, biochemistry (chemistry and biology), physical chemistry (chemistry and physics), medicinal chemistry (medicine and chemistry), chemical engineering (chemistry and engineering) etc.

The contributions of chemistry and chemist to social, industrial and economics of life of the world in general and Nigeria in particular has been felt in all facets of life. Chemistry provides support structure to other branches of science. Adesoji (2008) stated three societal needs which could be accomplished through the knowledge of chemistry. These are needs for physical survival, living satisfactory economic life and raising the level of our socio-cultural life. The knowledge of chemistry has enabled us to produce good water, food and health care resources, varying materials for construction of industries, roads, automobile and home in addition to solving problems resulting from human interaction with environment. Even when the knowledge of chemistry to the society is of great importance, students' performance in chemistry as measured by their scores in senior secondary certificate examination is very poor. Effective science teaching is the gateway to attainment of scientific and technological greatness. Science and chemistry teaching can only be effective when teachers are willing to make use of appropriate methods and resources in teaching the students. Adesoji (2008) and Olatoye and Afuwape (2003) suggest that one of the urgent needs in Nigeria is how to improve the teaching and learning of science. Presently the condition of science teaching and learning in schools is very discouraging. It calls for all stakeholders of education to rise up to this challenge in the interest of national development. Akinbobola and Afolabi (2010) also suggest that for science teaching to be meaningful and relevant, the nature of science must be adequately reflected. This calls for a shift of emphasis from the traditional content and factual acquisition of scientific knowledge to those that actively involve the learner in learning by doing.

The empirical framework works that advanced this research includes, Atilla (2012) who studied the dependability of joy performance ratings using the generalizability theory and in (2015), Atilla examined the estimation of generalizability coefficient: application with different programs. Ikeh and Madu (2018) examined application of generalizability theory in estimating multiple sources of variation in economics essay test, Heitman, Kovalski 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. Faleye, (2019) demonstrated the use of one facet person by item design and two facet crossed design in carrying out generalizability theory in CTBS science achievement test. And application of generalizability theory in estimating variance components in national examinations council problem solving questions in mathematics (Ogunka, & Oruwene, 2020).

Moreover, generalizability theory is a design that relates with items crossed with raters in the universe of admissible observations. This means that any one of the N_i items might be rated by one of the raters. If the population is crossed with this universe of admissible observation, then the corresponding G-study design is $p \times i \times r$, in which the responses of N_p persons by N_i items are each evaluated by raters (Brennan 2001). Therefore, examining the variance components for this fully crossed design can be used to examine results for any feasible two-facet design. Therefore, generalizability theory is based on the magnitude of selecting out errors due to person, item, rater and occasions among others. Their interactions that judge the magnitude of measurement error in the variance components as well as providing a summary of dependability coefficient observing the generalization of a sample test score to a considerable area of interest (Shavelson, et al in Ogunka, & Oruwene, 2020).

This work is desired to examine person by item by rater variance of measurement error in chemistry questions in West African Senior School Certificate Examination (WASSCE, 2020). Johnson et al. (2002) affirmed that measurement error is a situation whereby students' true ability is either underestimated or overestimated. Hence, the need of estimating measurement error is

inevitable because of the inconsistencies that exist in measurements instruments are enormous especially variance components.

The word essay is derived from a French word “Essayer” which means “To try or to attempt”. Essay test typically requires a testee to write a response to fulfill the requirement of the Item. Essay tests is a test that requires the testee to structure a rather long written response up to several paragraphs, sentence or a longer passage. Essay test is also called free response or free answer test because it allows the test-takers to demonstrate their writing skills, correct spelling and grammar in presenting their answers in sequential and logical order.

Essay test take more time to score or grade but takes less time to construct. The grading process in essay test is subjective as non- tests related information may influence the process hence considerable efforts is required to minimize the subjectivity of the grading process. They are potentially unreliable in assessing the entire contents of a subject matter; hence it could be referred to as subjective.

Essay test are tools used for evaluation. Despite the wider use of the short answer and objective test, there are certain outcomes of learning such as organizing, summarizing, integrating ideas and expressing in one’s own way which cannot be effectively measured by objective test hence the importance of essay test lies in the measurement of such instructional outcomes.

There are two categories of essay test namely; extended- response and restricted or short response item.

The aim of this study is to examine the variance components of persons by items by raters and test scores dependability in west African senior school certificate examination (WASSCE) essay questions in chemistry using generalizability theory What are the different sources of variance component in west African Examination council chemistry essay questions for 2020 in Etche local government area?

Research Questions

1. What are the different sources of variance component in West African council chemistry essay question for 2020 in Etche Local Government Area?
2. What is the generalizability coefficient of variance component in West African Examination Council (WASSCE, 2020) chemistry essay questions in Etche Local Government Area?
3. What is the index of dependability of variance component in West African Examination Council (WASSCE, 2020) chemistry essay questions in Etche Local Government Area?

METHODS

The design is two–facet fully crossed G–Study and D–Study. G–study is used to examine the partitioning of variance while D–study provides information for making decisions. D–Study generalizability theory uses the information obtained from G–Study to ascertain the measurement procedure in reducing unwanted variance and increase dependability reliability. The population is made up of eight hundred and sixty-seven (867) senior secondary three (SS3) students in Etche Local Government Area of Rivers State. Purposive sampling technique was used to select eight (8) senior secondary schools out of sixteen senior secondary schools. Cluster sampling technique used to sample one hundred and sixty-seven (167) Senior Secondary Students in (SS3). Instrument for data collection is 2020 May/June chemistry questions west African senior school certificate examination (WASSCE), each item is rated on 10 points marking scheme which was adapted by the raters. The instrument; West African Examination Council (WASSCE 2020) chemistry question was administered to the students in their various schools, retrieved and rated by the raters. The data collected was analyzed using a computer software SPSS through General linear model under Variance components MINQUE method.

Results

Research question 1: What are the different sources of variance component in West African council chemistry essay question for 2020 in Etche Local Government Area?

Table 1: Variance component estimate on generalized linear model (ANOVA MINQUE METHOD)

Source	Sum of Squares	ANOVA		
		df	Mean Square	Variance Estimate
Students	922.266	165	5.589	.141
Items	64.564	4	16.141	.034
Rater	54.330	1	54.330	.064
Items * Rater	4.828	4	1.207	.005
Students * Items	2550.741	654	3.900	1.755
Students * Rater	103.040	162	.636	.056
Students * Items * Rater	227.820	647	.352	-2.787 ^a

Table 1 presented seven sources of variance component and its estimate. The result shows that students by item is the highest contributor of measurement error (σ^2_{si} 1.755) followed by students (σ^2_s .141); followed by rater (σ^2_r .064); students by rater (σ^2_{sr} .056); item (σ^2_i .034) and item by rater (σ^2_{ir} .005). The estimation of various components interacted very high among students and item (σ^2_{si}).

Research Question 2: What is generalizability coefficient in West African council chemistry essay question for 2020 in Etche Local Government Area?

Table 2. Estimation of Generalizability Coefficient

Relative component	Variance	Universal Score	G-Study Coefficient
1.896	.141	0.069	

Table 2 shows the estimation of Generalizability coefficient of west African council chemistry essay question for 2020 in Etche local government area which indicates that; Relative Error variance 1.896; the Universe Score .141 and G-study coefficient 0.069.

Research Question 3: What is the index of dependability of variance component in West African Examination Council (WASSCE, 2020) chemistry essay questions in Etche Local Government Area?

Table 3. Estimation of index of Dependability

Absolute Error Variance	Universal Score	Index of Dependability
2.055	.141	0.064

Table 2 shows the estimation of Index of dependability of west African council chemistry essay question for 2020 in Etche local government area which indicates that Absolute Error Variance 2.055; universe score .141 and Index of Dependability 0.064 were obtained from the score.

DISCUSSION

The result of the study depended on estimating the variance component of students, items, raters and the interaction effects and determination of the coefficient of generalizability and index of dependability of 2020 May/June senior secondary school certificate examination in Chemistry for certification. The result of the study revealed that the highest contributor of measurement error from the score obtained is on student and item (σ^2_{si} 1.755). This indicates that a proportion of the variance was as a result of the interaction of students by item (σ^2_{si}). Although, the large variance component noticed here is not only with respect to student by item alone but to undifferentiated error. The second source of error variance is students (σ^2_s .141). Third rater (σ^2_r .064), fourth is students by rater (σ^2_{sr} .056); (item σ^2_i .034) and item by rater (σ^2_{ir} .005). But student by item by rater σ^2_{sir} variance has (-2.787^a) showing that error due to student by item by rater yielded negative estimate (-2.787) hence the degree of freedom for the residual error is zero. However, in variance component estimation negative estimate arises as a result of sampling error or model misspecification. According to Brennan cited in Ogunka and Orluwene (2020), the possible solution is to set negative estimate to zero but use of negative estimates in expected mean square equations for other components.

The estimation of the G-Study Coefficient and Index of Dependability of generalizability theory was done using the universe score. In this study, the universe score is taken as student weight mean (.141). Generalizability coefficient is the ratio of universe score variance dividing itself plus relative error variance while index of dependability is the ratio of universe score variance dividing itself plus absolute error variance (Brennan, 2001). G-Study Coefficient; universe score (.141) by itself plus relative error variance estimate (1.896) and Index of Dependability is universe score (.141) by itself plus absolute error variance estimate (2.055). Therefore, the reliability coefficient of the instrument is 0.064, which is very low and not adequate for West African senior school certificate examination in Chemistry which indicates that the chemistry question contain difficult items in 2020.

The finding of this study disagreed with the finding of Ogidi (2021) who studied the application of generalizability theory in the estimation of variance components in national examination council essay questions in Christian Religious Studies and revealed that the highest contribution of variance components was from person by item by rater, followed person by item, rater, person by rater, person and then item. Ogidi found that the index of dependability in his study was high and reliable which also disagreed with the finding of this study, which revealed low index of dependability. However, Ogidi (2021) found a negative estimate of item by rater, which agreed with the finding of this study.

Ogunka et al. (2020) studied the application of generalizability theory in estimating variance components in National examination council problem solving questions in Mathematics and found that the largest contribution of variance components obtained was from student, item and rater, followed by student by item, then by item and rater which contradict the finding of this study but Ogunka et al revealed a negative estimate of student by item by rater which agreed with the funding of this study.

CONCLUSION

Based on the results of the generalizability study, it is concluded that the highest contributor of measurement error is the student-item followed by student, rater, item, student by rater and item by

rater while student by item by rater was negative due to degree of freedom for the residual error is zero. However, the index of dependability is low (0.064) and not reliable for West African senior school certificate examination. It indicates that the Chemistry contained difficult items hence the exclamation “Almighty Chemistry”

RECOMMENDATIONS

1. The study is generally to deduce the sources of error variance component arising from student scores in West African senior school certificate examination. All West African senior school certificate examination essay question papers should be subjected to Generalizability theory and analysis for psychometric properties estimate to generate reliable test items with high index of dependability.
2. Students that offer chemistry should be encouraged by their teachers and parents to study hard to cover their scheme of work since it is an important subject to aid science related courses in the universities.
3. Test developers should be moderate and skilful in selecting test items for chemistry examinations.

REFERENCES

- Adesoji, F. A. (2008). Students Ability Levels and Effectiveness of Problem—Solving Instructional Strategy. *Journal of Social Physics*, 17, 5-8.
- Akinbobola, A. O., & Afolabi, F. (2010). Analysis of Science process skills in West African senior secondary school certificate Physics practical examinations in Nigeria. *American-Eurasian Journal of Scientific Research*, 5, 234-240.
- Asuru, V. A. (2015). *Measurement and evaluation in education and psychology* (2nd ed). Port Harcourt: Pearl Publishers International Ltd.
- Atilla, Y (2015). Estimation of generalizability coefficient: an application with different program.
- Atilla, Y. (2012). Dependability of job performance ratings according to generalizability theory. *Journal of Education and Science*, 163(37), 157 – 348.
- Brennan, R.L (2001) *Generalizability Theory: Statistics for Social Science and Public Policy*. Springer-Verlag Berlin Heidelberg. New York.
- Faleye, B. A. (2019). Generalizability theory: A workshop presented at the annual conference Zuba, Abuja 2018. Association of educational researchers and evaluators of Nigeria (ASEREN).
- Heitman, R.J; Kovalski, 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 (Abuja 2018). Association of Educational Researcher and Evaluators of Nigeria.
- Johnson, S. Dulanay, C. & Bank, K. (2000). Measurement error. Retrieved from http://www.wcpss.net/evaluation_research/reports/2000/mment_error.pdf
- 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.
- Kpolovie, P. J (2010). *Advanced Research Methods*, Owerri: Springfield Publisher Ltd.
- Ogidi (2021) Application of generalizability theory in the estimation of variance components in national examination council essay questions in Christian Religious Studies.
- Ogunka, R. I. and Orluwene, G. W. (2020). Application of generalizability theory in estimating variance components in national examinations council problem solving questions in mathematics.
- Olatoye, R. A., & Afuwape, M. O. (2003). Test anxiety as a determinant of examination misdemeanour among some Nigerian secondary school students. *Ibadan Journal of Educational Studies*, 3(182), 32-39.
- Preuss, R.A. (2003). Using generalizability theory to Develop Clinical Assessment Protocols. *Physical Therapy*, 93(4), 562-569.