ISSN: 2790-4407

Journal of Computer Adaptive Testing in Africa

2022; 1, 55-66

Application of Generalizability Theory in Estimating Variance Components in Senior Secondary School Certificate Examination Physics Essay Questions

[a]Blessing Iheanyichukwu Nwuchegbu & **[b]Richard Iheanyichukwu Ogunka***

[a,b] Ignatius Ajuru University of Education Port Harcourt, Nigeria

ABSTRACT

This study applied Generalizability Theory to estimate variance components in Senior Secondary School Certificate Examination (SSSCE) Physics essay questions using a two-facet fully crossed G-study and D-study design. The population comprised 1,767 Senior Secondary Three (SS3) students in Obio/Akpor, Rivers State. A multistage sampling technique selected 10 out of 25 schools, and 150 students were sampled using the Taro Yamane formula. The 2019 WASSCE Physics essay questions, rated on a 10-point marking scheme, were administered. Data analysis using SPSS under the Variance Components MINQUE method revealed that students contributed the most to measurement error (∂^2si = 1.030), followed by items (∂^2i = 0.851) and raters (∂^2r = 0.016). Interaction effects were highest for student-item (∂^2si = 2.957), while student-item-rater (∂^2sir = -2.868) was negligible. Reliability estimates indicated a universe score variance of 1.030, relative error variance of 2.981, absolute error variance of 4.878, a G-study coefficient of 0.26, and an index of dependability of 0.17, suggesting question difficulty. The findings highlight inconsistencies in student responses across items, which affect reliability. It is recommended that the West African Examinations Council (WAEC) establish a psychometric unit to validate essay questions, ensuring higher reliability and dependability before administration.

Keywords: essay questions, generalizability theory, variance components

* Corresponding author.

Ignatius Ajuru University of Education Port Harcourt, Nigeria, e-mail: iheanyiblessing1@gmail.com

To cite this article:

Nwuchegbuo, B. I. & Ogunka, R. I. (2022). Application Of Generalizability Theory in Estimating Variance Components in Senior Secondary School Certificate Examination Physics Essay Questions. *Journal of Computerized Adaptive Testing in Africa, 1*(1), 55-66.

Received: May 23 2021
Accepted: November 29 2021
Published: June 29 2022

INTRODUCTION

Physics is a natural science based on experiments, measurements, and mathematical analysis with the purpose of finding quantitative physical laws for everything from the nanoworld of the microcosmos to the planets, solar systems and galaxies that occupy the macrocosms. The laws of nature can be used to predict the behaviour of the world and all kinds of machinery. Many of the everyday technological inventions that we now take for granted resulted from discoveries in physics. The basic laws in physics are universal, but physics in our time is such a vast field that many subfields are almost regarded as separate sciences. Physics improves our quality of life by providing the basic understanding necessary for developing new instrumentation and techniques for medical applications, such as computer tomography, magnetic resonance imaging, positron emission tomography, ultrasonic imaging, and laser surgery. Physics - the study of matter, energy and their interactions - is an international enterprise, which plays a key role in the future progress of humankind. The support of physics education and research in all countries is important because:

1. Physics is an exciting intellectual adventure that inspires young people and expands the frontiers of our knowledge about Nature.
2. Physics generates fundamental knowledge needed for the future technological advances that will continue to drive the economic engines of the world.
3. Physics contributes to the technological infrastructure and provides trained personnel needed to take advantage of scientific advances and discoveries.
4. Physics is an important element in the education of chemists, engineers and computer scientists, as well as practitioners of the other physical and biomedical sciences.
5. Physics extends and enhances our understanding of other disciplines, such as the earth, agricultural, chemical, biological, and environmental sciences, plus astrophysics and cosmology - subjects of substantial importance to all peoples of the world.
6. Physics improves our quality of life by providing the basic understanding necessary for developing new instrumentation and techniques for medical applications, such as computer tomography, magnetic resonance imaging, positron emission tomography, ultrasonic imaging, and laser surgery.

Generally, for all these reasons, physics is an essential part of the educational system and of an advanced society. It is important to note that for teaching and learning of physics to be meaningful and relevant to the students and society, it must adequately reflect the nature of science; it must not only be process-oriented but should be emphasized on the product of science stressing honesty, open and critical mindedness, curiosity, suspended judgement and humility as a science and scientific enterprise (Akinbobola and Ado, 2007). Physics has to be learnt in secondary schools because of its central role in science and society generally. The teaching of physics therefore should be emphasized because, West African Examination Council (WAEC) uses the examination to assess students' performance or knowledge in physics and again, the certificate could be used for employment. Furthermore, the minimum requirement for admission into Nigerian universities to study medicine, engineering and other science related courses is credit grade.

The teachers who prepare physics students for senior secondary school certificate examinations, parents, students and other stakeholders in education are highly worried over the poor student's performance in physics over the years. Students' poor performance in physics in senior secondary school certificate examinations has been blamed on the construction, validation and administration of examination by senior secondary school certificate examination council as lacking in psychometric properties of its items; this study is imperative based on students and public outcry on the mass failure of students in physics. Therefore, developing a measuring instrument that is accurate, consistent and have high index of dependability that can measure in perfect manner with reliable quantitative description of the examinee in terms of exact extent in which it possesses the trait to be adjudge the best in respect to decision to be reliable essay question in physics. The purpose of this study is to validate essay question in physics.

Essay test is an assessment tool used in assessing students' performance in a given instruction. Essay questions elicits high levels of cognitive domain for critical thinking and student's originality in answering questions. Essay items provide the freedom of response that is needed to adequately assess the ability of students to formulate problems, organize, integrate and evaluate ideas and information and apply knowledge and skills (Linn, Miller and Gronlund, 2005).Orluwene, (2012) defined essay test as the test item which allows students the freedom to supply their own responses rather than select the correct answer. Orluwene went on to say that essay tests are presented with a narrative or question form and the students are required to compose a reply which presents a complete response in at least one sentence since the item allows for task with larger scope by requiring students to organize and

integrate information, interpret information, give arguments, give explanations, evaluate to merit and demerit of ideas and conduct other types of reasoning that tap complex thinking. Asuru, (2015) defined essay test as a test in which the testee supplies rather than selects the answer; he further stated that essay test is free answer test because it allows the testee to organize, integrate, synthesize and present his/her answers in series of inter-connected sentence. In this investigation, the researcher will try to avoid the use of options in physics essay questions since it is the assessment of level of all the sampled population of estimating variance components in Senior Secondary School Certificate Examination (SSCE, 2019) essay question.

Generalizability theory (G) is a statistical or test theory about the dependability of behavioral measurement, (Cronbach, et al cited in Ogunka & 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? Atest 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 behavioral 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 & 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 tastings, means that in the behavioral 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 Orluwene, (2019), dependability refers to the precision of generalization from a student's observed score on a test or behavior 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 behavioral 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; Oladele, 2022).

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 behavioral measures.

This study seeks to entirely examine student by item by rater variance components of measurement error in cognitive domain based on Senior Secondary School Certificate Examination (SSCE, 2019) essay question. Johnson, Dulanay and Banks, (2002) affirmed that measurement error is a situation in which students true score ability is either underestimated or overestimated. Thus, the need of estimating measurement error is unavoidable since the inconsistencies that are noticeable in

measurements instruments numerous mostly variance components. The purpose of this study is to estimate the variance components of students by items by raters and index of the dependability on Senior Secondary School Certificate Examination (SSCE, 2019) essay question in physics with the application of generalizability theory.

Research Questions

1. What are the different sources of variance component in west African Examination council (WASSCE, 2019) physics essay questions for 2020 in Obio/Akpor local government area?
2. What is the student contribution variance component in West African Examination Council (WASSCE, 2019) Physics essay questions in Obio/Akpor Local Government Area?
3. What is the items contribution variance component in West African Examination Council (WASSCE, 2019) Physics essay questions in Obio/AkporLocal Government Area?
4. What is the rater contribution variance component in West African Examination Council (WASSCE, 2019) Physics essay questions in Obio/AkporLocal Government Area?
5. What isthe interaction effect of variance component in West African Examination Council (WASSCE, 2019) Physics essay questions in Obio/AkporLocal Government Area?
6. What is the generalizability coefficient of variance component in West African Examination Council (WASSCE, 2019) Physics essay questions in Obio/AkporLocal Government Area?
7. What is the index of dependability of variance component in West African Examination Council (WASSCE, 2019) Physics essay questions in Obio/AkporLocal Government Area?

MATERIAL AND 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 one thousand, seven hundred and sixty-seven (1,767) senior secondary three (SS3) students in Obio/Apkor Local Government Area of Rivers State. Multistage sampling technique was used, simple random sampling technique and purposive sampling technique were used to select ten (10) senior secondary schools out of twenty-five senior secondary schools. Taro Yamane formula is applied to estimate the sample size of one hundred and sixty-seven (150) Senior Secondary three

Students (SS3) from the population and the west African examination council (WASSCE 2019) physics question was administered to the students in their various schools. Instrument for data collection 2019May/June physics 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.

RESULTS

Research Question 1: What are the different sources of variance component in west African Examination council (WASSCE, 2019) physics essay questions for 2020 in Obio/Akpor local government area?

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

Source	Sum of Squares	df	Mean Square	Variance Estimate
Students	2509.037	149	16.839	1.030
Items	1065.523	4	266.381	.851
Rater	17.797	1	17.797	.016
Items * Rater	23.250	4	5.812	.036
Students * Items	3780.981	596	6.344	2.957
Students * Rater	84.861	149	.570	.024
Students * Items * Rater	269.787	579	.466	-2.868[a]

Presented different sources of variance component and its estimate. The result shows that students by item is the highest contributor of measurement error (∂^2si 2.957) followed by students (∂^2s 1.030); followed by item (∂^2i .851); item by rater (∂^2ir .036) followed by students by rater (∂^2sr .024) and rater (∂^2r .016); The estimation of various component interacted very high among students and item (∂^2si) in West African Examination Council (WASSCE, 2019) Physics essay questions in Obio/Akpor Local Government Area.

Research Question 2. What is the student's contribution variance component in West African Examination Council (WASSCE, 2019) Physics essay questions in Obio/Akpor Local Government Area?

Table: Student Variance component estimate on generalized linear model

source	Sum of squares	df	Mean square	Variance estimate
student	2509.037	149	16.84	1.030

Table 2 shows that student contributed 1.030 variance component in west African council physics essay question for 2019 in Obio/Apkor local government area.

Research Question 3. What is the item contribution variance component in West African Examination Council (WASSCE, 2019) Physics essay questions in Obio/Akpor Local Government Area?

Table 3: Item Variance component estimate on generalized linear model

source	Sum of squares	df	Mean square	Variance estimate
Item	1065.52	4	266.38	.851

Table 3 shows that item contributed .851 variance component in west African council physics essay question for 2019 in Obio/Apkor local government area.

Research Question 4: What is the rater contribution variance component in West African Examination Council (WASSCE, 2019) Physics essay questions in Obio/Akpor Local Government Area?

Rater Variance component estimate on generalized linear model (ANOVA MINQUE METHOD)

source	Sum of squares	df	Mean square	Variance estimate
Rater	17.797	1	17.797	.016

Table 4 shows that item contributed .016 variance component in west African council physics essay question for 2019 in Obio/Apkor local government area.

Research Question 5: What is the interaction effect of variance component in West African Examination Council (WASSCE, 2019) Physics essay questions in Obio/Akpor Local Government Area?

Table 5: Interaction effect of Variance component estimate on generalized linear model

source	Sum of squares	df	Mean square	Variance estimate
Item * Rater	23.250	4	5.812	.036
Student * Item	3780.981	596	6.344	2.96
Student * Rater	84.861	149	.570	.024
Student * Item * Rater	268.787	579	.466	-2.868

Table 5 show the interaction effect of Variance component estimate in west African examination council in physics essay questions in 2019 in Obio/Apkor local government area. It shows that Student*Item (∂^2si= 2.96) interacted more and better followed item*Rater (∂^2sr = .036) and student*Rater (∂^2ir =.024). Student*Item*Rater (∂^2sir = -2.87) show negative estimate hence the residual error is zero.

Research Question 6. What is the generalizability coefficient of variance component in West African Examination Council (WASSCE, 2019) Physics essay questions in Obio/Akpor Local Government Area?

Table 6. Estimation of Generalizability coefficient

Relative Variance component	Universal Score	G-Study Coefficient
2.981	1.030	0.26

Table 6 shows the estimation of Generalizability coefficient of west African council physics essay questions for 2020 in Obio/Apkor local government area which indicates that; Relative variance component 2.981; the Universe Score 1.030 and G-study coefficient 0.26.

DISCUSSION

The result of this study relied on estimating the variance components of students, items, raters and the interaction effects; and determination of the coefficient of generalizability and index of dependability of 2019 May/June's senior secondary school certificate examination in physics in exist class for certification. The result shows that student is the highest contribution of measurement error (∂2s1.030) followed by item (∂2i .851); followed by rater (∂2i..016). The interaction effect of the variance components shows that student by item (∂2si = 2.957) interacted very well followed by item by rater (∂2ir .036) and student by rater (∂2sr .024). The estimation of variance component interacted very high among students and item (∂2si) which implied that student interacted well with the items. However, the result show negative error variance for students by items by rater hence the degree of freedom for residual error is zero. It is of great important to note that when there is negative value in error variance, the number is adjusted to zero. This is because calculating generalizability coefficient and index of dependability with negative number may cause exponential increase.

The generalizability coefficient and index of dependability of senior secondary school certificate examination (SSCE, 2019) is very low. The generalizability coefficient and index of dependability in this result agrees with Canning, R., Hawthrone, K., Hood, K. and Houston, H., (2005) and Anatol, T.,

and Hariharan, S., (2009). The reason for low coefficient could be as a result of limited error variance among the students. the universe score is taken as student weight mean (1.030). Generalizability coefficient is the ratio of universe score variance divide itself plus relative error variance while index of dependability is the ratio of universe score variance divide itself plus absolute error variance (Brennan, 2001). G-Study Coefficient; universe score (1.030) by itself plus relative error variance estimate (0.26) and Index of Dependability is universe score (0.17) by itself plus absolute error variance estimate (2.981). Therefore, the reliability coefficient of the instrument is 0.17 which is very low and not adequate for senior secondary school certificate examination in physics, it indicates that the physics contain difficult items hence the exclamation "Almighty physics".

CONCLUSION

Relying on the result of the findings of this study, it is concluded that students by item is the highest contribution of measurement error (∂2si2.957) and that negative error variance for students* items* and rater hence the degree of freedom for residual error is zero.it is also concluded that the reliability coefficient of the instrument is 0.17 which is very low and not adequate for senior secondary school certificate examination in physics, it indicates that the physics contain difficult items hence the exclamation "Almighty physics". The generalizability coefficient and index of dependability of senior secondary school certificate examination (SSCE, 2019) is very low. The generalizability coefficient and index of dependability in this result agrees with Canning, R., Hawthrone, K., Hood, K. and Houston, H., (2005) and Anatol, T., and Hariharan, S., (2009). The reason for low coefficient could be as a result of limited error variance among the students. The universe score is taken as student weight mean (1.030). Generalizability coefficient is the ratio of universe score variance divide itself plus relative error variance while index of dependability is the ratio of universe score variance divide itself plus absolute error variance (Brennan, 2001). G-Study Coefficient; universe score (1.030) by itself plus relative error variance estimate (0.26) and Index of Dependability is universe score (0.17) by itself plus absolute error variance estimate (2.981). Therefore, the reliability coefficient of the instrument is 0.17, which is very low and not adequate for senior secondary school certificate examination in physics, it indicates that the physics contain difficult items hence the exclamation "Almighty physics".

The following recommendations were made based on the finding of the results of the study:

1. West African examination council (WAEC) should have a unit of psychometricians to verify their essay questions to establish if it has good index of dependability before sending it out for use.
2. West African examination council (WAEC) should use generalizability theory for reliability test of all their essay questions before storing in their item banks to reduce the difficulty index of their items.
3. Raters should be train on rating skills in order to maintain rater agreement.

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.

Anatol, T. and Hariharan, S. (2009).Reliability of evaluation of student's awareness to essay-type questions.West India medical journal, 58 (1), 13-16.

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 Heidlberg. New York.

Cannings, R., Hawthrone, K., Hood, K. and Houston, H., (2005). Putting double marking to the test.A framework to assess if it worth the trouble.Medical education, 39 (3) 299-308.

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; 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 (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.

Linn, R. L., Miller, M. D. and Ground, N. E., (2005). Measurement and assessment in teaching (9thed) upper saddle river N. J: Pearson education, inc.

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.

Oladele, J. I. (2022). Estimating Measurement Error in University Students' Work Experience Programme (SWEP) Assessment using Generalizability Theory: Implications for Transdisciplinarity in Engineering Training. Transdisciplinary Journal of Engineering & Science, 13. https://doi.org/10.22545/2022/00185

Olatoye, R. A., &Afuwape, M. O. (2003). Test anxiety as a determinant of examination misdemeanor among some Nigerian secondary school students. Ibadan Journal of Educational Studies, 3(182), 32-39.

Orluwene, G. W., (2012). Introduction to test theory and development process. Port Harcourt: Chris-Ron integrated services.

Preuss, R.A. (2003). Using generalizability theory to Develop Clinical Assessment Protocols. Physical Therapy, vol. 93 issue 4, pp 562-*Examples of reference entries:*