

Application of Generalizability Theory in Estimating Dependability of Public Examination Essay Questions in English Language in Rivers State

Ogunka, Richard Iheanyichukwu^{1*}, Dr. Ijeoma M. Opara¹

¹Department of Educational Psychology, Guidance and Counselling Faculty of Education University of Port Harcourt

*Corresponding author: Ogunka, Richard Iheanyichukwu | Received: 25.03.2021 | Accepted: 12.05.2021 | Published: 16.05.2021 |

Abstract: The study investigated the application of generalizability theory in estimating dependability of public examination essay questions in English language. The design is ex-post facto and two facet fully crossed G-Study and D-Study. The population was 36,705 senior secondary students. Taro Yamane formula was applied to estimate the sample from the population using multistage sampling procedure, simple random sampling techniques and purposive sampling technique to select 420 senior secondary school students who registered for both West African examination council and National examination council in Rivers Stat. The instruments were West African Examination Council and National Examination Council essay questions papers 2018. The data obtained were analysed using Variance Component; G-Study and D-Study calculated by applying Felaye's formula. Results showed items variance components contributed high, generalizability coefficient and index of dependability are low 2018 essay questions in English language West African examination council and Students variance components contributed high, generalizability coefficient and index of dependability relatively high in 2018 essay questions in English language conducted by National examination council. Recommendations, public examination bodies should store verified psychometric properties essay questions in English language in their item bank for sales to users, they should create community of data analyst especially with estimation of Generalizability theory. Students should be encouraged by both parents and teachers to study hard, since exit examination are for certification which is a criteria for tertiary institution admission. Test developers should be consistent with syllabus in arranging test items. Raters should be trained on rating skills to maintain rater agreement School administrators and education managers should provide conducive environment and support for conduct of examinations by public examination bodies.

Keywords: Generalizability Theory Dependability Examination.

INTRODUCTION

English language is a core subject both in primary and secondary level of education in Nigeria educational environment. It is a language of communication in both school and non-school setting. English language has to be learnt in schools because of the roles it plays in the educational system and the society. The teaching and learning of English language must not be taken for granted. It should be carefully and thoroughly studied. Adegbile (2002) points out that success at every level of educational system depends largely on competence in English language. Certificate awarded to candidates that pass English language examination set by West African Examinations Council and National Examinations Council could be used to obtain employment and gain the minimal requirement for admission into any tertiary institution of higher learning in Nigeria point at credit grade in English language. Akinnaso (2018) observed English language

usage in Nigeria would have noticed persistent errors in the spoken or written output of most Nigerians, regardless of status or occupation. He continued that the decline is reflected in the West African Senior School Certificate Examination and National Examinations Council results in which the failure rate in English language and literature in English has been on the rise. Issues that bother the teachers, the public, parents and other stakeholders is mass failure rate which they blame on the West African Examinations Council and National Examinations Council.

It is quite unfortunate that an important judgement concerning a public examination body should be based on unconfirmed opinions without reference to psychometric properties and other parameters to establish the assessment distribution and reliability of the examination validity, variability in performance or other factors like administration, supervision and item development process for such a

Quick Response Code



Journal homepage:

<https://crosscurrentpublisher.com>

Copyright © 2021 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY-NC 4.0) which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

Citation: Ogunka, Richard Iheanyichukwu & Ijeoma M. Opara (2021). Application of Generalizability Theory in Estimating Dependability of Public Examination Essay Questions in English Language in Rivers State. *Cross Current Int J Peer Reviewed J Human Soc Sci*, 7(4), 91-99.

decision to be taken, will be good (Anagbogu et al., 2011). This research is useful based on public complain and perception on the recorded mass failure that made candidates sat for the examinations refers English Language as ALMIGHTY ENGLISH hence the assessment of students performance on 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 possesses and demonstrate the trait can be adjudge for the best in relation to decisions to be subsequently taken for reliable oral, essay and objective item in English Language. Hence this investigation seeks to assess essay questions.

Essay question is one of the assessment tools in testing or assessing student's achievement in any given instruction. Essay tests measures the higher levels of cognitive domain which provide critical thinking and originality in students. Linn et al. (2005) stated that essay items provide the freedom of response that is needed to adequately assess the ability of students to formulate problems, organise, integrate and evaluate ideas and information and thus, apply knowledge and skills. Orluwene (2012) defined Essay test as the type of test item which allows the students the freedom to supply their own responses, rather than select the correct answer. The author in addition, 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 scopes, by requiring students 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 questions are data collection protocols customarily used in non-school and school settings. Iweka (2014) shows guidelines that essay questions should follow which includes the behaviour the student is expected to exhibit before writing the prompt and indicate the point value for each question. Opara (2016) affirmed that essay test items are known as free response tests items because they allow the testees the freedom to express their response in their own words, thought, manner and ideas.

Achievement tests (measures of cognitive domain) are usually designed as a regular terminal evaluation to determine the statue of an individual on completion of a course of study.

Generally, a student writing an essay must gather his/her thoughts about a given subject, create an

idea, and organise his/her thoughts. Essay-type examinations are more recognised compared to other types of examinations for measuring writing ability of a student (Atilgan et al., 2017; Schoonen, 2005). From this backdrop, essay-type examinations are considered to be essential measurement tools in the field of education. Because there are differences between writing abilities of various students, students are not expected to achieve identical ratings in essay-type examinations. Furthermore, ratings will vary from one student to another, thereby reflecting differences between their writing abilities. However, a student's rating is affected by several extraneous factors. With respect to writing, which is a complex ability, these extraneous factors include several variance sources such as the task, type of task, rater, rating tool, essay topic, student's interest in the topic, essay type (such as descriptive, analytical, narrative or argumentative), time constraint, administration instructions, rating process, interaction, and other such factors (Schoonen, 2005; Sudweeks et al., 2005). Moreover, changes in ratings that are obtained based on this variance are considered to be measurement errors.

In this study, the researchers made efforts to avoid the use of options in essay questions in English language in 2018 WASSCE and NECO as the instrument for this investigation because it is the assessment of levels of all the sampled population to estimate generalizability theory and dependability in West African Examinations Council and National Examinations Council. The senior secondary school certificate examination conducted by West African Examination Council and National Examination Council are taken by private and public secondary school candidates in the final year of secondary education. The West African Examination Council is an examination board that was formed out of the concern for Africa and established in 1952, whereas National Examinations Council was an opportunity that created choice of examination body for candidates for their indigenous conditions for the point of registration for the examination in order to profound an alternative examination platform to overcome the public outcry of poor performance of candidates for West African Examination Council. This examination body was created in 1999.

Generalizability (G) theory is a statistics theory that enables the assessment of reliability and dependability of behavioural measurements, and conceptualization of reliable observations (Brennan, 2001b; Shavelson et al., 1991). G Theory was first carried out by Cronbach et al. (1972) ; because of the shortcomings of classical test theory. Classical test theory considers the errors inherent in a measurement to be errors from single source. Whereas G theory considers the errors from all potential errors' sources together, as well as their interactions effects (Brennan,

2001a). The focus of G theory is to generalize the observed scores of measured subjects to the population scores accurately by defining and interpreting the measurement results and distinguishing different sources of variance. Orluwene (2012) indicated that, in the measurement of complex traits imperfect instruments are used so that the score observed for each person always differs from person's true ability or characteristics; it was also 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 raised above on the problem of incomplete certainty 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 researchers are required to empirically establish the reliability of the measuring instrument which he or she develops. The methods of estimating dependability of reliability in generalizability theory and classical test theory to determine standard error measurement varies. Atilla (2015) asserted that the use of classical test theory approaches to determine score reliability, however, are not capable of identifying and untangling this profusion of error which classical reliability was not conceptualized to do since it accounts for only one source of error at a time. Kpolovie (2010) asserted classical theory as reliability embedded in the true and error- score model. He defines reliability as the coefficient that predictable proportion of variance in observed scores from the true scores.

Whereas generalizability theory liberalizes classical theory by employing ANOVA methods that allows an investigator to disentangle multiple sources of error that contribute to the undifferentiated E in classical test theory. However, it is important to state that GENOVA, Statistical Analysis System (SAS), Statistical Package for Social Sciences 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 Analysis of variance method adopted for the estimation of variance components for generalizability theory.

G-theory considers two means of estimating reliability which are relative and absolute measurement. Therefore, in G theory two distinct coefficients are estimated: a generalizability coefficient for relative measurement and an index of dependability for absolute decisions. It is therefore particularly suitable for the analysis of reliability coefficient and index of dependability in cognitive data. The context of this study is a situation in which average of multiple sources

made on the same student across one measurement facets, which includes the facet of item and rater, are of interest and that the precision (or standard error) will be estimated for such means. (Brennan, 2001a) acknowledged the difficulty of handling the facet of time, but from a different perspective. He emphasized the importance of conceptualizing the replication of a measurement procedure that was the heart and soul of the concept of measurement reliability. If a test taken on one occasion and rated by two independent raters was considered the measurement procedure and then variance of scores across the two independent raters would be an adequate estimate of random fluctuation across the administration of the measurement procedure. Data obtained from the two independent raters would not likely be affected by changes in true score, but data obtained from one rater that was too far apart might confound random error with changes in true scores.

This study seems to thoroughly investigate student by item by rater, its interaction effects error variance, relative and absolute errors, generalizability coefficient and index of dependability in 2018 essay questions in English language conducted by West African Examinations Council and National Examinations Council.

It is known fact that the results of students' performance in the public examination conducted by West African examinations and National examinations council are determinants of the quality of those who enter tertiary institutions. This is highly related to the students' result in English language since without a credit pass in English language candidates for admission into tertiary institution are denied despite it is found out that students experience poor performance in English language. The syllabus consists of three sections; comprehension and summary, lexis and structure and oral form (phonetic and phonology) conveying the four key areas of the Use of English course which are listening, speaking, writing and reading. Both West African examinations council and National examinations council English language examination question papers involve objective, essay and oral questions. It is obvious the difficulties students face while answering questions that bothers on critical thinking, sequence arrangement of argument, mechanical and accuracy. Hence the failure rate in English language is mostly from the essay items. This study seeks to apply generalizability theory in estimating dependability of essay questions in English language 2018 senior secondary school certificate examination conducted by West African examinations council and National examinations council.

The following research questions guided the study

1. What are the different sources of error that contribute to error variance component in

- essay questions in English language conducted by West African examination council in 2018 in Rivers state?
2. What are the different sources of error that contribute to error variance component in essay questions in English language conducted by National examination council in 2018 in Rivers state?
 3. What is the contribution of students to variance component in essay questions in English language conducted by West African examinations council and National examinations council in 2018 in Rivers State?
 4. What is the contribution of items to variance component in essay questions in English language conducted by West African examinations council and National examinations council in 2018 in Rivers State?
 5. What is the contribution of raters to variance component in essay questions in English language conducted by West African examinations council and National examinations council in 2018 in Rivers State?
 6. What are their interaction effects to variance component in essay questions in English language conducted by West African examinations council and National examinations council in 2018 in Rivers State?
 7. What is the comparison of generalizability coefficient in essay questions in English language conducted by West African examinations council and National examinations council in 2018 in Rivers State?
 8. What is the comparison of index of dependability in essay questions in English language conducted by West African examinations council and National examinations council in 2018 in Rivers State?

METHODS

The research design are Ex-post facto and generalizability theory. Ex-post facto design is adopted

for empirical comparison of the group’s differential independent variables on the given dependent variable. While generalizability theory design, is two-facet fully crossed G-study and D-study. This is because G-study generalizability theory estimates multiple sources of variance components and the information obtained from G-study is use in determining D-study measurement procedure in minimizing undesirable variance and maximize generalizability coefficient and index of dependability. The population of the study was 36,705 senior secondary three students in Rivers State. (Source Rivers State ministry of education 2019/2020). A sample of 420 senior secondary school three students that registered in both West African examination council and National examination council in 2019\2020 session was used for the study. This sample size is adequate, because Taro Yamen’s formula (Baridam, 2001) was applied to obtain minimum sample size of 396 which is lower than 420. Multistage sampling procedure of simple random sampling technique, purposive and accidental sampling techniques were also used in drawing the sample size. Two instruments were used for the study. They are 2018 Senior School Certificate Examination in essay questions in English language conducted by West African Examinations Council and National Examinations Council respectively. The two examination were already valid and reliable. The data were analysed using variance components which disentangled the multiple sources of error variance and also formula related to G-study and D-study as in Faley (2019).

RESULTS AND DISCUSSION

Research Question 1: What are the sources of error that contribute to error variance component in essay questions in English language conducted by West African examination council in 2018 in Rivers state? This research question was answered using the Variance Components, this was done on yearly bases. The results obtained are displayed table 1.

Table-1: Summary of ANOVA table on the sources of error variance on essay questions in English language of West African senior school examination certificate

ANOVA estimator for 2018				
Source	Type 1 sum of square	df	Mean. Square	Variance components
Students	5454.257	419	13.017	0.24
Item	142673.143	6	23778.857	28.20
Rater	5.178	1	5.178	-0.027
Item*Rater	524.778	6	87.463	0.18
Student*Item	30841.225	2514	12.268	0.19
Student*Rater	3884.311	419	9.270	-0.37
Student*Item*Rater	29836.670	2513	11.873	11.87
Error	.000	1	.000	0.000
Total	773143.000	5880		

Table 1 shows the components that contributed to error variance. The table highlighted the multiple sources, sum of square (SS₁), Degree of freedom (DF), mean square (MS) and variance components on the columns; student, item, rater, item*rater, student*rater and student*item rater on the rows.

Whereas WAEC 2018 result of this study shows, the highest contribution to error variance is item $\partial^2_i = (28.20)$ and student* item* rater $\partial^2_{sir} = (11.87)$, student $\partial^2_s = (0.24)$ student* item $\partial^2_{si} = (0.19)$, rater $\partial^2_r = (0.18)$, negative error variance for

student* rate $\partial^2_{sr} = (-0.37)$ and rater $\partial^2_r = (-0.027)$.

Research Question 2: What are the sources of error that contribute to error variance component in essay questions in English language conducted by National examination council in 2018 in Rivers state.

This research question was answered using the Variance Components, this was done on yearly bases. The results obtained are displayed in table 2.

Table-2: Summary of ANOVA table on the sources of error variance on essay questions in English language in National examination council

ANOVA estimator for 2018				
Source	Type 1 sum of square	df	Mean. Square	Variance components
Students	26761.693	419	63.870	4.50
Item	7427.791	5	1485.558	1.76
Rater	3.377	1	3.377	-0.001
Item*Rater	24.162	5	4.832	0.006
Student*Item	17660.157	2095	8.430	2.99
Student*Rater	1614.781	419	3.854	0.24
Student*Item*Rater	5088.148	2079	2.447	1.2
Error	19.500	16	1.219	1.2
Total	526799.000	5040		

Table 2 shows the components that contributed to error variance. The table highlighted multiple sources, sum of square (SS), degrees of freedom (DF), mean square (MS) and variance components on the columns and student, item, rater, item*rater, student*item, student*rater, student*item*rater on the rows.

NECO 2018 error variance estimate student contributed very high with $\partial^2_s = (4.50)$ followed by student * item $\partial^2_{si} = (2.99)$ item error variance $\partial^2_i = (1.76)$ then student * item * rater has $\partial^2_{sir} = (1.2)$ student * rater $\partial^2_{sr} = (0.24)$ item * rater $\partial^2_{ir} = (0.006)$ and rater $\partial^2_r = (-0.001)$.

Research Question 3: What is the contribution of students to variance components in essay questions in English language conducted by West African examinations council and National examinations council 2018 in Rivers State?

This research question was answered using ANOVA estimator table as indicted by contribution of student error variance. This was done in table 3.

Table-3: Students contributions to variance components in essay question in English language conducted by West African Senior School Certificate Examination and National Examination Council

Public examination 2018 Students variance components.	
WASSCE	.239
NECO	4.50

Table 3 shows the student error variance student $\partial^2_s(.239)$ for 2018 Again student $\partial^2_s = (4.50)$ for 2018 contribution to error variance in essay questions in English language conducted by West African Examination Council and National Examination Council respectively in Rivers State.

Research Question 4: What is the contribution of items to variance components in essay questions in English language conducted by West African examinations council and National examinations council in 2018 in Rivers State?

This research question was answered using ANOVA estimator table as indicted by contribution of item variance components. This was done in table 4.

Table-4: Items contributions to variance components in essay question in English language conducted by West African Senior School Certificate Examination and National Examination Council

Public examination	2018 Item variance components.
WASSCE	28.20
NECO	1.76

Table 4 shows the item $\partial^2 i(28.20)$ for 2018, item. And items variance components $\partial^2 i=(1.76)$ for 2018. item variance contribution to error variance in essay questions in English language conducted by West African Examination Council and National Examination Council respectively in Rivers State.

Research Question 5: What is the contribution of raters to variance component in essay questions in

Table-5: Raters contributions to variance components in essay question in English language conducted by West African Senior School Certificate Examination and National Examination Council

Public examination	2018 Raters variance components.
WASSCE	-0.027
NECO	-0.001

Table 5 shows the rater $\partial^2 r = (-0.027)$ for 2018. And rater $\partial^2 r = (-0.001)$ for 2018, raters contribution to error variance in essay questions in English language conducted by West African Examination Council and National Examination council respectively in Rivers State.

Research Question 6: What are their interaction effects to variance components in essay

Table-6: Interaction contributions to variance components in essay question in English language conducted by West African Senior School Certificate Examination and National Examination Council

Public examination body	2018 Interactions variance components.
WASSCE	
Item*Rater	0.18
Student*Item	0.19
Student*Rater	-0.37
Student*Item*Rater	11.87
NECO	2018
Item*Rater	0.006
Student*Item	2.99
Student*Rater	0.24
Student*Item*Rater	1.23

Table 6 shows the interaction effects of item *rater $\partial^2 ir = (0.18)$, student*item $\partial^2 si = (0.19)$, student* rater $\partial^2 sr = (-0.37)$, student* item* rater $\partial^2 sir = (11.87)$ for 2018, Again, item* rater $\partial^2 ir = (0.006)$, student * item $\partial^2 si = (2.99)$ student* rater $\partial^2 sr = (0.24)$, student * item * rater has $\partial^2 sir = (1.23)$ for 2018. Interaction effects contribution to error variance in essay questions in English language conducted by West African

English language conducted by West African examinations council and National examinations council in 2018 in Rivers State?

This research question was answered using ANOVA estimator table as indicted by contribution of raters variance components. This was seen in table 5.

questions in English language conducted by West African examinations council and National examinations council in 2018 in Rivers State?

This research question was answered using ANOVA estimator table as indicted by contribution of effects variance components. These were done in table 6.

Examination Council and National Examination Council respectively in Rivers State.

Research Question 7: What is the comparison of generalizability coefficient in essay questions in English language conducted by West African examinations council and National examinations council in 2018 in Rivers State?

This research question was answered using estimation of G-coefficient formula.

Table-7: G-coefficient of essay questions in English language conducted by West African Senior School Certificate Examination and National Examination Council

Year	Relative error Variance	Universes score	G-coefficient
WASSCE 2018	2.99	.239	0.07
NECO 2018	0.99	6.75	0.87

Table 7: Compare: $E^{p^2} = (0.07)$ for 2018
 And $E^{p^2} = (0.87)$ for 2018 as the calculated generalizability coefficient in essay questions in English language conducted by West African Examination Council and National examination council respectively in Rivers State.

Research Question 8: What is the comparison of index of dependability in essay questions in English

language conducted by West African examinations council and National examinations council in 2018 in Rivers State?

This research question is answered using estimation of index of dependability (D reliability coefficient) formula.

Table-8: Index of Dependability in essay questions in English language conducted by West African Senior School Certificate Examination and National Examination Council

Year	Absolute error Variance	Universes score	Index of dep.
WASSCE 2018	17.21	.239	0.01
NECO 2018	3.96	6.75	0.63

Table 8: compare $\phi = (0.01)$ for 2018.
 And $\phi = (0.63)$ for 2018. The index of dependability in essay questions in English Language conducted by West African Examination Council and National Examination Council respectively in Rivers State.

DISCUSSION OF FINDINGS

The result of this study is anchored on estimating the variance components of students, items, raters and their interaction effects, and determining the coefficients both generalizability and index of dependability of public examination conducted by West African examination council in National examination council every year for senior secondary students in the exit class for certification.

This study further shown that the sample senior secondary students were administrated, 2018 essay questions in English language in West African examination council and National examination council respectively to determine the error variance components; which each contribute to relative error variance and absolute error variance.

It is important to note, that where negative notation occurs in error variance, the number is adjusted to zero. The reason is because to calculate generalizability coefficient and index of dependability with the negative number it can cause exponential increase. The contribution of each of the various sources of variance will depend on the number of unique levels for each source. Example, the contribution of students equals the ratio of the study on variance component to the number of students (Brennan, 2001a; Haertel, 2006) . In the same vein, Ogunka et al. (2019) assented negative variance

component 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 arise 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 equation for other components (Brennan, 2001a).

This study design is two- facet fully crossed and the calculated generalizability coefficient and index of dependability is low for West African Examination Council. Now, generalizability theory, particularly the two-facet design, is arguably the most sophisticated design for estimating reliability, yet a review of literature revealed some studies that reported low levels of reliability which agrees with Cannings et al. (2005) and Anatol et al. (2009). Potential reason for the low coefficients may include limited error variance among students themselves. This study observed low coefficient because students use shorthand, wrong spelling and Pidgin English and above all lack of vocabulary identified as the other reasons deduced as characters of students writing essay questions in English language conducted by West African Examination Council. The public relations officer of the council in Ghana, Agnes Teye Cudjoes said the 2018 Chief Examiner’s report shows that students failed the subject because they used shorthand and Pidgin English in their exams. She further asserted lack of vocabulary as reasons secondary school students fail the subject. Cudjoes said “you would agree with me that these days our candidates hardly write a good letter, our students hardly read, everybody is busy doing WhatsApp, Facebook and all that”.

However, for the estimated item variance component been large, it suggests that much variability is in observed scores; also, it appears that the items are quite similar in average difficulty in West African examination council essay questions. It's an obvious fact among students who sat for West African examination council that refers English language as ALMIGHTY ENGLISH, this study result confirms that assertion.

National Examination Council NECO results showed that student variance component was significantly contributed to the error variance. This means that high variance indicates that the data points are very spread out. The variance is the average of the squared distances from each point to the mean. The generalizability coefficient and index of dependability of essay questions in English language senior secondary school examination conducted by National Examination Council in 2018 showed high G- coefficient and index of dependability. The students discriminate appropriately with the item for National examination council causing relatively high coefficient. The result agreed with Iheanyichukwu et al. (2020). Application of generalizability theory in estimating variance components in National Examination Council problem solving questions in mathematics; the study design is two facet research design, instrument is from same public examination body and it has an acceptable reliability. Again, this study result agrees further with Atilla (2015) estimation of Generalizability coefficient; an application with different programs. The software is GENOVA, EDU G, SPSS, SAS program. The program software was paired as GENOVA and EDU G, SPSS and SAS. And the design is two facet G study that estimation generated person, task and evaluator, and its interactions. The generalizability coefficient and index of dependability was close to each other with high common and double reliabilities coefficient.

Application of generalizability theory is useful in estimating dependability of reliability in determining how many measurements are needed across the facet items and raters in order to obtain reliable scores from essay items in English language in examination conducted by both West African Examination and National Examination Council. The significantly contributed amount of measurement error variance was items in West African Examination Council and measurement error variance due students in National Examination Council. Rater agreement was consistent. Although the generalizability coefficient and index of dependability of West African Examination Council had low reliabilities in 2018 and National Examination Council had acceptable reliabilities. The number of samples for this study made the estimate of the dependability robust and exact.

RECOMMENDATIONS

The following recommendations were therefore made:

1. Public examination bodies should store verified psychometric properties essay questions in English language in their item bank for sales to users. They should create a community of data analyst especially with estimation of Generalizability theory.
2. Students should be encouraged by both parents and teachers to study hard, since exit class examination are for certification which is a criteria for tertiary institution admission.
3. Test experts should be consistent with syllabus when developing test items.
4. Raters should be train on rating skills to maintain rater agreement.
5. School administrators and education managers should provide conducive environment and support for conduct of examinations by public examination bodies.

REFERENCE

- Adegbile, J. A. (2002). The Relative Effects of Two Models of Advance Organiser on Performance in Reading Comprehension. *Journal of Pedagogy and Educational Development*, Rivers State College of Education, 3(2), 22-36.
- Akinnaso, N. (2018). The failing standard of English in Nigeria *Punch Newspaper*.
- Anagbogu, G. E., and Damkor-Ikpa, M. A. (2011). Comparative Analysis of Students' Performance in NECO and WAEC Examinations: Instruments Appraisal for Research and Development. *Benue State University. Journal of Education*, 11, 215-228.
- Anatol, T., & Hariharan, S. (2009). Reliability of the Evaluation of student's awareness to Essay-type Questions. *West India medical Journal*, 58(1), 13-16.
- Atilgan, H. (2013). Sample Size for Estimation of I and Phi Co-Efficient in Generalizability Theory. *Egitim Arastirmalari-Eurasian Journal of Educational Research*, 51, 215-228.
- Atilgan, H., Kan, A., & Aydin, B. (2017). *Egitimde Olcme ve Degerlendirme [Measurement and Evaluation in Education]*. Ankara: Anı Yayıncılık.
- Atilla, Y. (2015). Estimation of generalizability coefficient: an application with program.
- Brennan, R. (2001a). *Generalizability Theory*. New York, USA: Springer-Verlag.
- Brennan, R. (2001b). GENOVA (Version 2.1) Computer Software and Manual. www.education.uiowa.edu/casma/
- Cannings, R., Hawthorne, K., Hood, K., & Houston, H. (2005). Putting Double marking to the test: A framework to Assess if it is worth the trouble. *Medical Education*, 39(3), 299-308.

- Cronbach, L. J., Gleser, G., Nanda, H., and Rajaratnam, N. (1972). *The Dependability of Behavioural Measurements: Theory of Generalizability of Scores and Profiles*. New York, NY: Wiley.
- Faleye, B. A. (2019). *Generalizability Theory: A Workshop Presented at Annual Conference Zuba, Abuja 2018*. Association of Educational Researchers and Evaluators of Nigeria (ASEREN) Journal of Educational Research and Evaluation.
- Graham, S. (2006). Writing. In P. Alexander, & P. Winne (Ed.) *Handbook of educational psychology*. NJ: Erlbaum: Mahwah.
- Iheanyichukwu, O. R., & Orluwene, G. W. (2020). Application of Generalizability Theory in Estimating Variance Components in National Examination Council Problem Solving Question in Mathematics. *European International Journal of Science and Technology*, 9(4), 61-69.
- Iweka, F. (2014). *Comprehensive Guide to Test Construction and Administration*. Omoku Chifas Nigeria. .
- Johnson, S., and Johnson, R. (2009). *Conceptualising and Interpreting Reliability*. Coventry: Of qual.
- Kpolovie, P. J. (2010). *Advanced Research Methods*. Owerri: Springfield Publishers Limited.
- Linn, R. L., Miller, M. D., & Ground, N. E. (2005). *Measurement and Assessment in Teaching* (9th ed.) Upper Saddle River, N.J: Pearson Education, Inc.
- Ogunka, R. I., and Orluwene, G. W. (2019). Application of Generalizability Theory in Estimation of Variance Component in National Examination Council Essay Question In English Language The African Journal of Behavioral and Scale Development Research, 1(2), 43-48.
- Opara, I. M. (2016). *Test Construction and Measurement*. Owerri: Career Publisher.
- Orluwene, G. W. (2012). *Introduction to Test Theory and Development Process*. Port Harcourt: Chris-Ron Integrated Services.
- Schoonen, R. (2005). Generalizability of Writing Scores: An Application of Structural Equation Modeling. *Language Testing*, 22(1), 1-30, <https://doi.org/10.1191/0265532205lt0265532295oa>.
- Shavelson, R., and Webb, N. (1991). *Generalizability Theory*. Newbury Park, USA: Sage Publications.
- Sudweeks, R. R., Reeve, S., & Bradshaw, W. (2005). A Comparison of Generalizability Theory and Many-Facet Rasch Measurement in An Analysis of College Sophomore Writing. *Assessing Writing*, 239-261, <https://doi.org/210.1016/j.asw.2004.1011.1001>.