

ESTIMATING THE VARIANCE COMPONENTS OF 2022/2023 SCIENCE, TECHNOLOGY AND SOCIETY TEST SCORES OF BAYERO UNIVERSITY, KANO NIGERIA

HAFSAT ABDULMALIK ANIKI¹; ABUBAKAR ADULLAHI Ph.D.²;
SADIYA MU'UZU SANI Ph.D.³ & BINTA ABBA Ph.D.⁴

¹Bayero University, Kano Staff Model Secondary School

^{2,3&4}Department of Education, Bayero University, Kano

abdulimalikhafsat62@gmail.com; aabdullahi.edu@buk.edu.ng

Babba.edu@buk.edu.ng

Abstract

The study estimated the variance components of 2022/2023 Science, Technology and Society Test Scores of Bayero University Kano, Nigeria. Three objectives along side with three research questions guided the study. One-facet random person by items crossed G-study design was used for the study. The population consisted of 1247 undergraduates who sat for the test. Sample of 600 was selected using multi-stage sampling technique. The responses of the examinees were the data used for the study. The data was analyzed using VARCOMP analysis via IBM SPSS (Version 27.0.1). The variance components obtained was used to answer the three research questions. The findings of this study revealed that the variance component due to; person is 0.013, item is 0.038 and person by item interaction is 0.199. It was concluded that, the largest contribution to measurement error occurs from residual and unsystematic sources of measurement error not measured in this study. It was therefore, recommended that the School of General Studies of Bayero University, Kano should maintain the current level of its item quality for 2022/2023 Science, Technology and Society Test. Similarly, generalizability theory should be embraced in analyzing measurement issues (Reliability).

Keywords: Generalizability and Variance Components.

Introduction

The identification and reduction of measurement errors is a major challenges in educational evaluation procedure which compelled most test specialists to rely solely on Classical Test Theory (CTT) for assessing reliability of assessment instruments, but some experts has recommended the use of Generalizability Theory (G-Theory) instead of CTT in order to breached the lapses, in terms of estimating undifferentiated single source of measurement error. This article provides synopsis introduction to G-theory; describes variance components as well as person by items effect of measurement errors.

CTT provides an assortment of useful individual methods for assessing reliability e.g., test-retest, alternate forms, internal consistence and inter-rater agreement. But it considered only one source of measurement error at a time and there is also no way of combining the numerous possible reliability indices to obtain an overall estimate of reliability; no way of determining the relative importance of the various sources of error and no way of assessing possible interactions between the sources of error (Conbach, Gleser, Nanda & Rajaratnam, 1972, Brennan, 2010).

G-Theory, on the other hand is more encompassing, informative and use full alternates to show that the error that exist in observe score can stem from different sources. G-Theory provides an all-at-once way of revealing and comparing the sources of error in a common metric (Brennan, 2010) by providing estimate index of the variance associated with interaction between the various sources. For instance, if an instrument is administered on two occasions, G-Theory analyze and estimate the variance

components contributed by persons (P), items (i), and occasion (o) of the measurement and the variance components of each four possible interaction between these facets (i.e., person by items [p x i], person by occasion [p x o], items by occasion [I x o] and person by item by occasion [p x i x o] Brennan, 2000a, 2000b, 2001). G-Theory liberalize CTT by employing Analysis of Variance (ANOVA) method that allow test analysts to untangle multiple sources of error (Brennan, 2010), while CTT note $X=T+E$ (where X is the observe score, T is the true score and E is the undifferentiated error), G-Theory allows for the exploration of multiple sources of error as, $X=T+(E1+E2+E3\dots)$, thus at the simplest level CTT is too limited in analysis of source of variance because it assumes only one source of error at a time, despite the fact that in reality there are many different definition of what this errors look like.

Brennan (2003) elaborate further that, G-Theory allows the impact of variety sources of error such as items, occasions, forms or raters on the reliability of measurement to be examined within a unified framework. Similarly, G-Theory acknowledges multiple influences on measurement variance, whereas CTT partitions variance into only two sources. G-Theory like factorial ANOVA, partitions variance into many sources corresponding to a systematic variance among the object of measurement and to their interactions (Shavelson & Webb 1991, Hussain A, 2012), the authors further explained that, G-Theory informs the analyst the magnitudes of types of errors, distinguishes between relative and absolute decisions, as well as the possibility of treating sources of error as fixed or random.

Variance Components assume central importance in G-Theory, as they are the building block that provides a crucial foundation for all subsequent results. An ANOVA procedure is used for equating mean scores to their expected values and then solving for the estimated variance components. For the one facet person crossed by item design (p x i), the expected value of the mean squares (EMS equation) is;

$$EMS(p) = \sigma^2(pie) + ni\sigma^2(p)$$

$$EMS(i) = \sigma^2(pie) + np \sigma^2(i)$$

$$EMS(pie) = \sigma^2(pie) \text{ (Source: Brennan, 2001, Shavelson & Webb, 1991)}$$

Solving these equations for the variance components and using mean squares in place of their expected values, we obtain the ANOVA estimates:

$$\sigma^2(p) = [MS(p) - MS(pie)] / ni$$

$$\sigma^2(i) = [MS(i) - MS(pie)] / np$$

$$\sigma^2(pie) = MS(pie) \text{ (Source: Brennan, 2001, Shavelson & Webb, 1991)}$$

Considering the one facet design above where person as randomly selected, an observed score for a particular person on a particular item is decomposed into an effect for the grand mean, plus effects for the person, the item, one way interaction and a residual (one-way interaction plus unsystematic error). The contribution of each component or effect, except for the grand mean has a mean of zero and variance (σ^2) called variance component (Brennan, 2010). In practice, the parameter values and expected values are unknown therefore; G-Theory typically uses analysis of variance (ANOVA) procedure to compute terms that can be used to obtained estimates of variance components. However, unlike ANOVA, G-Theory is not much concerned itself with test of statistical significance but employs ANOVA sum of squares and mean squares to obtain estimates of variance components as the building block for estimating generalizability and dependability coefficients (Shavelson & Webb, 2005).

Estimating reliability of educational assessment instruments is one of important aspects of test development, administration and reporting procedure. However, estimating the contribution of each source of error to the overall reliability of an instrument cannot be over emphasize, as it will aid in adopting the best test procedure that tend to eliminate the possible measurement errors. According to

Brennan (2010), the contribution of error variance component of person, items, and persons by items interaction (i.e., one-facet design) provides the needed indices to determine the optimal test procedure. The few studies conducted empirically in this respect includes Bamidele (2018) applied multivariate G-theory to estimate the generalizability of 2015 National Examination Council School Certificate Examination (NECO) objective test in electrical installation and maintenance work, using one facet fully crossed design. The findings revealed that the variance component due to person is 0.2, item is 0.03 and person by item interaction is 0.20 for the 2015 NECO electrical installation test. In contrast to this study, Imasuen and Adeosun (2023) conducted study using generalizability theory to estimate the measurement error variance in 2019 WAEC Mathematics objective examination in Benin Metropolis reports that variance components of 0.38, 0.03 and 0.028 for person, item and person by item respectively for 2019 WAEC Mathematics objective examination in Benin Metropolis.

Bayero University, Kano as institution develops and administered test (Teacher-Made-Test) to its students that is of unknown psychometric properties, thereby limit the schools' knowledge in the best test procedure to adopt in order to minimize some of the possible measurement errors militating against the reliability of the results obtained from their assessment instruments. It is against this background, this article focused on estimating the variance component of 2022/2023 Science, Technology and Society Test of Bayero University, Kano Nigeria. Therefore, the article sought to achieve three objectives. To estimate the variance components due to; persons, items and person by interaction in 2022/2023 Science, Technology and Society Test of Bayero University Kano, Nigeria.

Research Questions

This study was specifically conducted to find answers to the following research questions;

1. What is the variance component due to person in 2022/2023 Science, Technology and Society Test Scores of Bayero University, Kano?
2. What is the variance component due to items use in 2022/2023 Science, Technology and Society Test Scores of Bayero University, Kano?
3. What is the variance component of 2022/2023 Science, Technology and Society Test Scores of Bayero University Kano?

Methodology

One-facet random of person by items (p x i) crossed G-study design was used for this study. Brennan (2010) stated that, univariate G-theory design has only one universe score for the object of measurement, similarly Shavelson and Webb (2005) opine that random design is necessary and more appropriate if the conditions of a facet can be exchanged with any other condition from the same universe, as it allows conditions in the sample to be exchanged with another set of the same size conditions from the universe as well as when the researcher intends to generalize the outcome of the research to the universe and beyond. Therefore, the study focuses on only one universe score (2022/2020 Science, Technology and Society Test Scores) for its analysis and intend to generalize the outcome to cover all undergraduates that sat for the tests.

Table 1: Shows Source of Variability in one-facet design

Source Notation	Type of Variability	Variance
Persons (P)	Universe Score	$\sigma^2 p$
Items (i)	Condition	$\sigma^2 i$
P x Interaction i (pi e)	Residual	$\sigma^2 pie$

Source: (Shavelson & Webb, 1991)

The population of the study consists of 1,247 undergraduates that registered and sat for 2022/2023 Science Technology and Society examination in Bayero University, Kano. They included 558 students from seven programme in Faculty of Arts and Islamic Studies, 47 students from Theatre and Performing Art in Faculty of Communication, as well as, 642 students from twelve programme in Faculty of Education. Out of this population a sample of 600 students' scores were sampled in line with Smith, (1978), Charter, (2003) and Hakan, (2013).

Multi-stage sampling procedure was used to select the scores of 600 students as sample size out of 1,247 students that registered for and wrote 2022/2023 Science, Technology and Society test for this study. In stage1 Proportionate sampling technique was used to select 268 students' scores form faculty of Art and Islamic studies, 24 students' scores from faculty of communication and 308 students' scores from faculty of Education. Stage2, Proportionate sampling technique was equally used to select students' scores from each program in each faculty according to sample sizes proportionate to them. While in Stage3 Simple random sampling technique without replacement was used to select students' scores from each programme in their respective faculty to constitute the sample sizes.

The data use for the study were the students' responses to 100 multiple-choice tests for 2022/2023 Science, Technology and Society that was collected from the School of General Studies and Entrepreneurship Programme therefore, the use of instrument is not required. The students' response was subjected to VARCOMP analysis via IBM SPSS (Version 27.0.1) to obtain the variance components.

Results

Table 2. Summary of Analysis of Variance showing the Estimate Variance Components and Percentage of Score Variation For 2022/2023 Science, Technology and Society Test of Bayero University, Kano.

Source Variance	DF	SS	MS	Variance	% of Variance
Corrected Model	698	3209.027	4.597		
Intercept	1	14760.960	14760.960		
Person (P)	599	905.520	1.512	0.013	5.2
Item (i)	99	2303.507	23.268	0.038	15.2
Residual (pie)	59301	11790.013	0.199	0.199	79.6
Total	60000	29760.000			100

As shown in Table 2, the estimated variance component for person (p) is 0.013 which account for 5.2% of the total variance, while the estimated variance component for item(i) is 0.038 accounting for 15.2% of the total variance and the estimated variance component due to person by item(pie) or residual is 0.199 indicating 79.6% of the total variance. This implies that, person by item interaction (residual) contributed more to the measurement errors in 2022/2023 Science, Technology and Society Test Scores of Bayero University Kano, Nigeria.

Research Question One: What is the variance component due to persons in 2022/23 Science, Technology and Society Test Scores of Bayero University, Kano?

Table 3. Estimate Variance Component for Persons in 2022/2023 Science, Technology and Society Test of Bayero University, Kano.

Source Variance	Variance Comp.	Estimated Variance Comp.	% of Variance
Person (P)	$\sigma^2 p$	0.013	5.2

From the result in Table 2 and 3, the variance component for person is 0.013 which account for 5.2%

of the total variance and it is the lowest variance for 2022/2023 Science, Technology and Society Test of Bayero University, Kano. This implies that the universe score among the examinees vary relatively small from one person to another corresponding to the largely similar scores obtained by the examinees in 2022/2023 Science, Technology and Society Test. Therefore, since persons represent the object of measurement and not error, it is then deduced that the level of variance component obtained from the examinees represent the systematic individual differences in 2022/2023 Science, Technology and Society Test of Bayero University, Kano.

Research Question Two: What is the estimate of the variance component due to items used in the 2022/2023 Science, Technology and Society Test Scores of Bayero University, Kano?

Table 4. Estimate Variance Component for Items used in 2022/2023 Science, Technology and

Source Variance	Variance Comp.	Estimated Variance Comp.	% of Variance
Item (i)	$\sigma^2 i$	0.038	15.2

The result in Table 2 and 4 shows that, the variance component for items is 0.038, indicating 15.2% and second to the largest of the total variance for 2022/2023 Science, Technology and Society Test of Bayero University, Kano. This implies that, most of the items used in the test are moderately difficult and none of the examinees scored the entire item correct or incorrect. It is generally desired that the numbers of moderate difficult items in a test should be higher and the numbers of easy and difficult items relatively less.

Research Question Three: What is the estimate of the variance component in the 2022/2023 Science, Technology and Society Test Scores of Bayero University, Kano, due to the interaction of persons by items?

Table 5: Estimate Variance Component for Persons by items interaction in 2022/2023 Science, Technology and Society Test Scores of Bayero University, Kano.

Source Variance	Variance Comp.	Estimated Variance Comp.	% of Variance
Residual (pie)	$\sigma^2 pie$	0.199	79.6

The result in table 5 revealed that, the variance component due to person by item interaction is 0.199 accounting for 79.6% of the total variance and it is the largest variance component for 2022/2023 Science, technology and society Test Scores of Bayero University, Kano. This reveals that the larger proportion of measurement error was due to person by item interaction and other unsystematic or systematic sources of variance not measured in the study.

Discussion of Findings

The findings of this study revealed that the variance component due to person is the lowest component of the total variance; it stands at 0.013 accounting for 5.2% of the total variance for 2022/2023 Science, Technology and Society Test Scores. This finding is in line with the study of Bamidele (2018) who reported variance component of person in 2015 NECO electrical installation and maintenance work’s test to be 0.2 accounting for 8% of the total variance and it is the lowest. The finding is also in consonant with the study of Akindahunsi and Afolabi (2017) who study 2012 NECO English examination and reported variance component of person to be 0.0142 accounting for 6%.

The finding of study revealed that variance component due to item used in 2022/2023 Science, Technology and Society Test of Bayero University Kano has a value of 0.038 accounting for 15.2% of the total variance and second to the largest variance. This finding is in agreement with the finding of Bamidele (2018), who reported that, 2015 NECO objective test in electrical installation and maintenance work has variance component due to item of 0.03 accounting for 12%. Imasuen and

Adeosun (2023) who assessed the variance components of 2019 WAEC mathematics objective test in Benin metropolis, reports variance due to items at 0.028 accounting for 24% as second to the largest variance.

The finding of the study revealed that the highest contribution to measurement error in 2022/2023 Science, Technology and Society Test was the residual (person by item interaction) effect with the value of 0.199, accounting for 79.6% of the total variance. This finding is in line with Bamidele (2018), who reported that, the variance due to person by item interaction for 2015 NECO objective test in electrical installation and maintenance work account for the largest variance with a value of 0.20 representing 80%. Equally Akindahunsi and Afloabi (2017)'s study of 2012 NECO English examination in Nigeria, reported variance component due to person by item interaction to be 0.1472, accounting for 62.3% of the total variance.

Conclusion

Based on the finding, it is therefore concluded that the persons who are object of measurement in 2022/2023 Science, Technology and Society Test of Bayero University Kano contributed the lowest variance component and are regarded as measurement free error, it shows relative difference among the test-taker. Similarly, items used in the test contributed moderate effect to the measurement error, showing high items quality. The main contribution to the measurement error comes from the residual (person by item interaction), this is regarded as both systematic and unsystematic sources of error not measured in this study, it's classified as the measurement error.

Recommendations

Based on the findings of this study, the following recommendations are made;

1. The management of Bayero University, Kano and the School of General Studies should make test analysis one of important aspects in test development process.
2. The management of Bayero University, Kano and the School of General Studies and Entrepreneurship Programme should maintain the current level of 2022/2023 Science, Technology and Society Test item quality and extend it to other tests items, as well as device a means to identify more sources of measurement error in order to maximize the reliability of its assessment instrument.
3. Educational measurement community (test experts, test developers, test specialists e.t.c), researchers, students of Tests and Measurement as well as other disciplines should adopt G-theory methodology and procedure in estimating measurement error, as it provides an all-at-once way of revealing and comparing sources of error in a common metric.

References

- Akindahunsi, O. F., & Afolabi, E. F. (2017). Using Generalizability Theory to Investigate the Reliability of scores Assigned to students in English Language Examination in Nigeria. *Journal of Measurement and Evaluation in Education and Psychology*: 12(2): 147-162 ISSN: 1309-6575.
- Bamidele, S. T. (2018). *Multivariate Generalizability of 2015 National Examinations Council School Certificate Examination objective test in electrical installation and maintenance work.*[Unpublished doctoral thesis], University of Ilorin, Ilorin Nigeria.
- Brennan, R. L. (2003). *Coefficients and Indices in Generalizability Theory* (CASMA Research Report Number 1). Center for Advanced Studies in Measurement and Assessment.
- Brennan, R. L. (2001b). *GENOVA (Version 2.1)* [Computer software and manual]. American College Testing, Inc.
- Brennan, R. L. (2010a). *Generalizability theory- Statistics for Social Sciences and public policy.*

Springer-Verlag Inc.

- Charter, R. A. (2003). Study Sample are too small to produce Sufficient Precise Reliability Coefficients. *The Journal of General Psychology*, 130(2), 117-129.
- Cronbach, L. J., Gleser, G. C., Nandam, H., & Rajaratnam, N. (1972). The dependability of behavioral measurements: Theory of generalizability scores and profiles. John Wiley.
- Hakan, A. (2013). Sample Size for Estimation of G and Phi Coefficients in Generalizability Theory. *Eurasian Journal of Educational Research*. 13(51), 215-228.
- Hussain, A. (2012). Generalizability Theory: An Analysis of Variance Approach to Measurement Problems in Educational Assessment. *Journal of Studies in Education*. 2(1). ISSN: 2162-6952.
- Imasuen, K., & Adeosun, P. K. (2023). Application of Generalizability Theory in Measurement Error in 2019 WAEC Mathematics Objective Examination in Benin Metropolis. *International Journal of Psychological and Brain Service*, 8(2), 13- 18. ISSN: 2575-2227(Print).
- Shavelson, R. J., & Webb, N. M. (1991b). *Generalizability theory. A primer (concepts in generalizability)*. Park, CA: Sage.
- Shavelson, R. J., & Webb, N. M. (2005). *Generalizability theory*. Green, 36, 599 -612.
- Shavelson, R.J., Webb, N.M., & Rowley, G. (1989). *Generalizability theory*. *American psychologist*, 49, 922 -932.
- Smith, P. L. (1978). Sampling Error Variance of Components in Small Sample Multifaceted Generalizability Studies. *Journal of Educational Statistics*, 3(5), 319-346.