West African Journal of Education, Vol. XXXIX 2019 Page 114-124

West African Journal of Education
Vol. XXXIX 2019 ISSN: 0043-2997
Website: http//www.instituteofeducui.com

Estimating Generalizability and Dependability Studies of Students' Scores in Teaching Practice Assessment in a Nigerian College of Education

[1]S. T. Bamidele , [2]S. B. Adebara, [3]Usman Muhammed and [4]Ameen Akeem
[1]bamidelets@gmail.com ; [2]adebarabolaji60@gmail.com;
[3]usmanmohammed3335@gmail.com ; [4]ameem.akeem@gmail.com
Department of General Studies Education, College of Education (Technical), Lafiagi Kwara State

Submitted February 4, 2020; Revised March, 23, 2020; Accepted June 9, 2020

Abstract

Teaching practice is a mandatory requirement for graduation from colleges of education in Nigeria. The objectives of the study was to find out the variance components for persons, occasions and persons crossed with raters and residual, generalizability coefficient and the dependability index (indices). Two facets crossed with ratters design were used. 1,393 students participated in the study. The data used were the scores of students' in 2016/2017 teaching practice course. The data were analysed using variance components (VARCOMP), relative error variance, absolute error variance and indices statistics. The findings revealed that in 2016/2017 session, the following scores were obtained in 2016/2017 teaching practice course; persons (o^2p) = 0.02 occasions, (0^2_0) = 0.03 interaction of persons with occasions, (0^2_{p0}) = 0.05 and residual (0^2_{prqe}) = 0.15, generalizability coefficient of 0.84 and a dependability index of 0.74 respectively. The study concluded that the indices of students' scores in 2016/2017 teaching practice were high or acceptable. This implies that the quality of students' scores in teaching practice course is standard. The study recommended that the sixteen weeks slated for teaching practice should be maintained.

Keywords: Generalizability, Dependability, Classical test theory, Students' scores, Teaching practice Assessment

Introduction

In measurement history, the leading theory for explaining latent trait underlying examinees' test performance is the Classical Test Theory (CTT) which describes how errors can influence observed scores. It is a simple model based on the true score theory that introduces three concepts - test scores or observed scores (X), true scores (T), and error scores (E). It recognizes that the characteristics of the testing situation can

contribute to measurement error and evaluates sources of error separately. This is the case of test-retest, alternate forms and internal consistency option of reliability of measures. The combined effect of error from the various threats to the reliability of an estimated score is rarely considered. Each method of reliability in CTT yields valuable information but provides only a slice of a bigger picture. Each piece of the source of error is estimated in isolation and fulfils a single objective. How all the various sources of error operate at one time and fit together to influence the overall reliability of the instrument cannot be estimated (Shavelson & Webb, 1991).

CTT deals with the reliabilities of relative decisions, where an individual's score is compared with a reference group and used to rank order the individuals as in norm-referencing although the measurement error is not considered. Absolute reliability occurs where an individual is compared to a well-defined standard and used to provide the absolute value of an attribute as in criterion-referencing; the measurement error is not considered (Shavelson & Webb 1991). *Generalizability Theory (GT) is a conceptual framework and a methodology that enables an investigator to disentangle multiple sources of errors in a measurement procedure. The roots of generalizability theory can be found in Classical Test Theory and Analysis of Variance (ANOVA).* In psychology and education, measurement issues have been addressed principally using CTT which postulates that observed scores can be decomposed into a "true" score and single undifferentiated random error term, *E* (Brennan, 2000).

Generalizability Theory (GT) is a statistical theory about the reliability of behavioural measurements. Reliability refers to the accuracy of generalizing from the person's observed score on a test or other measures (e.g. behaviour observation, opinion) to the average score that person would have received under all the possible conditions that the test user would be equally willing to accept. Generalizability theory provides an all-at-once way of revealing and comparing the sources of errors in a conjoint metric. It also provides estimates of the variance contributed by source and in addition, presents estimates of the variance associated with interface between the various sources. For instance, if an instrument is administered on two occasions, generalizability theory provides estimates of the variance contributed by persons, items, occasions and each of the four possible interactions (persons by items, persons by occasions, items by occasions and persons by items by occasions). It also provides helpful forecasts of the improvements in measurement. In this theory, reliability can be obtained by altering the number of persons, items, and occasions. This theory enables decision makers to determine how many occasions, test forms and administrators are needed to obtain reliable and valid scores.

Decision study, which is the second study in GT according to Wan, Li, Fan, Yang, & Pan (2014), gives information about which protocols are optional for particular measurement situations by generating generalizability coefficient that could be considered as reliability coefficient for various facets of the study. Decision studies

make use of information provided by generalizability study in arriving at its own reliability coefficient called index of dependability. It is the index of dependability that is used to determine how dependable measurement behaviour is. Dependability is the accuracy of generalizing an individual's average score which he would have received under all the possible conditions that the examiner would be equally willing to accept (Nie, Yeo, & Lau 2007). Therefore, the dependability of a measurement depends on the accuracy of its generalization. Dependability subsumes all other aspects of Generalizability theory. This is because the result obtained from generalizability study is used to access dependability.

GT, as noted by Gerbil (2013), is a powerful tool for estimating consistency in test scores. It also helps test developers to make correct decisions related to test development and administration, effective investigation of the relative contribution of various facets to test precision, provision of information about the impact of increasing or decreasing the number of test tasks, and test validation in order to optimize measurement precision. The importance of teaching practice in teacher education programme and its usefulness for student teachers makes it necessary that such an important programme is validated.

Table 1 shows the sources of variability of the univariate two-facets crossed persons by occasions by raters (p x o x r) design.

Table 1: Sources of Variability in Two -Facet Measurement Error

Source of variability	Type of Variability	Variance Notation
Persons (p)	Universe score	σ^2p
Occasions (o)	Conditions	σ^2o
Raters	Conditions	σ^2r
Persons by Occasions by raters	Residual	pro,e

Source: Shavelson & Webb (1991)

The students' scores during teaching practice exercise has always been used as part of their final grades. The scores as noted earlier, usually differ from one lecturer to the other. This difference is a major part of what GT uses to determine the dependability level of scores. It is therefore necessary to examine the dependability of students' scores in teaching practice programme using GT. This was why this study examined and analysed the dependability of the students' scores in teaching practice courses in Nigeria colleges of education.

Statement of the Problem

Nakpodia (2011) examined teachers and the student practice teaching programme in Nigerian educational system; Jekayinfa et.al' (2012) examined the lecturers' assessment of teaching practice exercise in Nigerian Universities; Fasasi (2013) investigated the

duties of students' teachers and supervisors in teaching practice; and Ekundayo, Alonge, Kolawole and Ekundayo (2014) examined the challenges and the way forward for education students in Nigerian colleges of education in teaching practice. However, none of these researchers examined the dependability of the scores given to students by the supervisors particularly with the use of generalizability theory.

Also, researchers such as Shavelon and Webb (1991), Brennan 2000, 2001, 2010), Burns (1998) and Gerbil (2013), have worked on the application of generalizability theory to test the dependability of measurement behaviour (instruments, scores). But GT has not been applied to test the dependability of students' scores in teaching practice courses particularly in Nigerian colleges of education. All these constitute research gaps parts of which this study intends to fill.

The purpose of this study is to investigate the dependability level of the sources of regular Nigerian Certificate of Education students' teaching practice in the 2016/2017 academic session in the sampled universities in Nigeria. Specifically, this study intends to:

i. determine the five variance components for persons, occasions, and persons crossed with raters and residual.
ii. investigate the generalizalibility coefficient of the students' scores in teaching practice.
iii. find out the dependability (phi) coefficient of the students' scores in teaching practice.

Research Questions

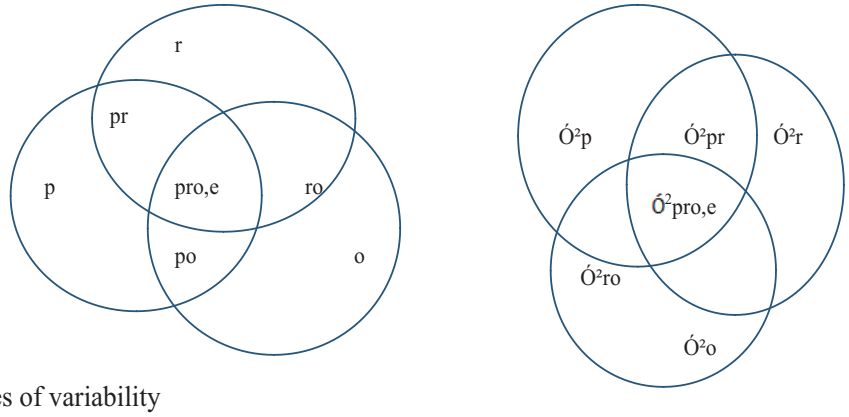
The following research questions were generated to guide the study:

1. What are the variance components for persons, occasions, persons crossed with rater and residual in the undergraduate students' scores in teaching practice course?
2. What is the generalizability coefficient of the students' scores in teaching practice course?
3. What is the dependability (phi) coefficient of the students' scores in teaching practice course?

Methodology

A two-facet random design of persons crossed, occasions and raters was employed in this study. In this design (P x rxo), there was one object of measurement (persons – these were the NCE students that participated in the teaching practice programme for the 2016/2017 academic session) and two facets of observations (raters – the lecturers that assess the students during teaching practice and occasions – the periods of teaching

practice exercise). This is in agreement with Shavelson and Webb's (1991) statement that two facet random design has one object of two other facets of measurement (person) observed by different conditions of two other facets of observations (raters and occasions).



a. Sources of variability

b. Variance components

Figure 1. Venn Diagrams for a Two -Facet, Crossed (p• x r• x o•) Design

The population for this study consists, of 1,393 students of College of Education (Technical), Lafiagi Kwara State. Ninety (90) Lecturers participated in the exercise during the 2016/2017 academic session. Purposive sampling technique was used to carry out the study. A Proforma titled "Scores of Students in 2016/2017 Teaching Practice course" was used to obtain samples scores in Teaching Practice in the 2016/2017 academic session. More than one lecturer usually assess students during teaching practice programme. They graded the students on percentage and the scores of students from the lecturers' were averaged to represent each students' performance. In this study, the raw scores turned in by each lecturer was used. The results of teaching exercise were used for students' grading and certification which means that the results could be said to be valid and reliable. This study was carried out using generalizability theory.

The data collected were analysed using variance (VARCOMP) to answer research question one, while the remaining research questions were answered using the formulas of relative error variance, generalizability coefficient and phi coefficient.

Results

The data used for this study were the sampled scores of students' in the 2016/2017 teaching practice course of a Nigerian college of education. Research question one was answered using VARCOMP while research questions two and three were answered using the relative variance error variance (σ^2rel) absolute variance error (σ^2abs), generalizability co-efficient and phi co-efficient formula

Research Question 1: What are the variance components for persons, occasions, and person crossed raters and residual in the students' scores in teaching practice course?

The raw scores of students from the ten assessors were input into the SPSS (IBM version 21) after which the data were subjected to syntax analysis to obtain the variance components. The results obtained are presented in Table 2.

Table 2: Estimated variance components for students' scores in 2016/2017 Teaching Practice scores

Source of variability	Variance	estimated variance	% of total variance
Student (P)	σ^2p	0.02	8%
Occasion (o)	σ^2o	0.03	12%
Person x occasion (po)	σ^2po	0.15	20%
Residual (pro,e)	σ^2pro,e	0.25	60%
Total			100%

The research question was aimed at showing how much variance component for persons in the observed scores is due to students' differences in their performance during the teaching practice assessment; the difference could be among persons' characteristics. From the above table, it was revealed that the estimated variance component for persons (σ^2p) is 0.02 which accounts for 8% of the total variance in the students' scores during teaching practice. The estimated variance component or persons crossed with occasions (σ^2po) is 0.05 and this accounted for 20% of the total variation in the students' scores. While the estimated variance component for residual (σ^2pro,e) is 0.15, which accounted for 60% of the total variance in the students' scores at the teaching practice. The largest estimated variance component of (0.15) is that of the residual as it accounted for 74% of the total variance in the students' scores.

Research Question 2: What is the generalization co-efficient of the student scores in teaching practice course?

Relative error variance and generalizability coefficient were used to answer this research question. The estimated variance component for persons (σ^2p) and person occasion by ratters residual (σ^2Pr, Pro,e) were used to determine the relative error variance while estimated variance component for person and relative error were also used to obtain generalizability coefficient. The result are presented in Table 3.

Table 3: Relative Error Variance and Generalizability Coefficient for 2016/2-17 Students' Teaching Practice Scores

Source	Generalizability co efficient
σ^2Rel	0.0053
σ^2P^2	0.8400

Table 3 shows that the estimated relative error of 0.051 and generalizability coefficient of 2016/2017 teaching practice scores is 0.84. Therefore the generalizability obtained is high and acceptable, since the value obtained is not less than the acceptable value of 0.070.

Research Question 3: What is the dependability (phi) co efficient of the student scores in 2016/2017 teaching practice course?

The decision study or dependability index for the 2016/207 teaching practice was obtained using absolute error variance and dependability equations respectively. Table 4 revealed the results obtained.

Table 4: Relative Error Variance and Dependability Index for 2016/2017 of Students' Teaching Practice Scores

Absolute error (σ^2abs)	dependability index (Φ
0.0028	0.7400

Table 4 revealed that the absolute error variance is 0.0028 while, the decision study or dependability index of 2016/2017 teaching practice scores is 0.7400. This shows that dependability index of the 2016/2017 teaching practice scores are acceptable since the obtained value does not fall below the acceptable value of 0.70. From the result of this study, it was revealed that 0.80 generalizability coefficient and 0.74 dependability index were obtained. This shows that there is no need to increase the number of occasions and raters. Forecasting could have been carried out assuming that the coefficient and dependability index are not above the acceptable value of 0.70

Discussion of Findings

It was observed from the result obtained that the variance component due to persons (σ^2p) accounted for 8% of the total variance in the 2016/2017 students' scores in teaching practice course percentage and this indicated that students performed differently in during the assessment. The difference in the performance of the students

could be because students normally try to improve their performance steadily after the first three weeks assessment during teaching practice programme. This is an indication that learning has taken place. The variance components for occasion (σ^2o) accounted for 12% of the total variance in the students' scores in 2016/2017 teaching practice course percentage. This is a little higher than the variance component for persons which suggested that the sixteen weeks of teaching practice led to the difference in the students' scores during the 2016/2017 teaching practice programme.

The finding also revealed that residual (σ^2pro,e) has the largest contribution as it was responsible for 60% of the total variation in the students' scores in the teaching practice course. This variance component represents persons crossed with occasion and raters, the interaction effect between persons, occasions and raters' other systematic and unsystematic sources of variation that were not part of the design. It was also suggested that the combination of students' period of teaching practice and raters' variable contributed seriously to the difference in the students' scores. This interpretation, given to the variance components, is supported by Shavelson and Webs (1991a) assertion that when the estimated variance components are summed together, the results give the total variance components. The variance component which has the larger percentage contributed to the largest proportion of the total variation while the variance component with the least percentage contributed the smallest proportion of the total variation. The finding of the current study is in line with Tunde (2015) who reported that the residual has the major sources of measurement error. The finding was equally corroborated by Shalveson and Webs (1991a); Hintze and Petite' (2001) whose study found that the residual has the largest contribution to measurement error. Findings in this study revealed that the generalizability co efficient of the 2016/2017 of teaching practice scored was 0.8400. This indicated that the estimated relative error variance (σ^2o) which the estimated relative error persons observed deviation scores and their universe scores (Brenman 2002a) obtained through the variance and estimated variance components that were used to obtain the generalizability coefficient of the 2016/2017 teaching practice scores was high or acceptable. Therefore if generalizability coefficient of 0.8 is to be classical test theory it will be regarded as a high reliability.

The result also revealed that the dependability coefficient/index of the 2016/2017 teaching practice scores and that the estimated absolute error variance (σ^2abs) for the design of this study was the difference between person observed score and the variance score (Brennan 2000). Three out of four variance components in this study contributed to the absolute variance error. It was only the variance component for persons that did not contribute to the absolute variance error (σ^2abs). The obtained D study or dependability index was high (0.74) considering the 0.70 level of acceptability value. Therefore, the dependability index of the 2016/2017 teaching practice is high as the dependability index is the parameter used to determine the dependability level of an instrument (Nie et al 2007).

The high dependability index level of the 2016/2017 teaching practice scores may be due to the contributions of the four sources of measurement errors and the difference in the persons' performance evident in the high level of commitment of students during the 2016/2017 teaching practice programme. The interpretation given the variance components were in line with Shalveson and Webs (1991b) who opined that when estimated variance components are added together, the outcome reveals that the total variance components and the variance components that has the largest percentage contributed the largest part of the total variation, while the variance component with the lowest percentage contributed the least part of the total variance. This finding was not in line with that of Turner et al (2010) who studied the effect of ventilation on segmental bio-impedance spectroscopy measures using generalizability theory. They found that the largest source of variation was persons (P) variance component unlike this study which has residual (Pro,e) as the largest source of variation. The findings of this study are not also in line with findings of Turner's (2010) study. This may be due to the fact that the study was not carried out in Nigeria. This could be an indication of location differences in research - 100 participants participated in Turner's study while 1,393 participated in this study. These factors could be responsible for the difference.

Conclusion

It could be concluded from the findings of this study that the value of generalizability and dependability coefficients of the 2016/2017 is high or acceptable. This means that the students were hardworking and adequate supervision took place in the sampled college of education and the participating schools. This could be due to the contribution of the four sources of measurement errors to the difference in the 2016/2017 teaching practice scores used for this study.

Recommendations

Based on the findings and conclusions drawn in this study, it is recommended that:

1. The sixteen weeks for the teaching practice exercise should be maintained.
2. raters should be given adequate orientation before going out for assessment.
3. The schools administrator of every participating school should always assess students' lesson plans on a daily basis.
4. raters should be highly remunerated in order to improve their commitment.

References

Akindutire, I.O. (2001). *The teaching profession*. Lagos: Universal Publishers.

Akindutire, I. O. and Ekundayo, H. T. (2012). Teacher education in a democratic Nigeria: Challenges and the way forward. International Research Journals *Review.* Vol. 3(5) pp. 429-435, May 2012

Brennan, R. L. (2000). Performance assessments from the perspective of generalizability theory. *Applied Psychological Measurement*, Vol.24, No. 4, 339-353.

Brennan, R. L. (2001a). *Generalizability theory*. New York: Springer.

Brennan, R. L. (2001b). *GENOVA (Version 2.1)* [Computer software and manual] Iowa City, IA: American College Testing, Inc.

Brennan, R. L. (2010a). *Generalizability theory*- Statistics for Social Sciences and public policy: New York: Springer-Verlag Inc.

Brennan, R. L. (2010b). *Generalizability theory*. International Encyclopaedia of Education, 61-68 New York: Springer.

Burns, N. & Grove, S. (1998). *The practice of nursing research; conduct, critique and utilization.* 3rd edu. WB. Saunders Company Philadelphia

Ekundayo, H. T., Alonge, H. O., Kolawole, A. O. & Ekundayo, S. K. (2014). Teaching Practice exercise for education students in Nigerian universities: Challenges and the way forward. *Mediterranean Journal of Social Sciences, 5(9), 486-492*

Fasasi, Y. A. (2013). Teaching Practice: Duties of student teachers and supervisors, *Ilorin Journal of Health, Physical Edducation and Recreation, 6, 143-150*

Gebril, A. (2013). Generalizabilty theory in language testing. *The Encyclopaedia of Applied Linguistics*, 1-7

Hintze, J.M. & Pettite, H.A. (2001). The generalizability of CEM oral reading fluency measures across general and special education. *Journal of Psycho Educational Assessment,* 19, 158-170.

Jekayinfa, A. A., Yahaya, L. A., Yusuf, A., Ajidagba, U. A., Oniye, A. O., Oniyangi, S. O., & Ibraheem, T. O. (2012) Lecturers' assessment of teaching practice exercise in Nigerian universities. *Journal of Education and Practice, 3(4), 79-86*

Mahmud, J. O. (2017). Analysis of Dependability of Undergraduate Students' Scores in Teaching practice Courses in a Nigerian University. A Ph. D. Thesis presented to the Department of Social Sciences Education, Faculty of Education, University of Ilorin; Ilorin, Nigeria.

Nakpodia, E. D. (2011). Teacher and the students' teaching practice programme in Nigerian educational system. *International Journal of Educational Administration and Policy Studies 2 (3), 33-39*

Nie, Y., Yeo, S.M. & Lau, S. (2007). Application of Generalizability Theory in the Inestigation of the quality of Journal Writing in Mathematics Studies. Educational Evaluation, 33, 371-383

Shavelson, R.J. & Webb, N.M. (1981). Generalizability theory: 1973–1980, *British Journal of Mathematical and Statistical Psychology* **34**, 133–166.

Shavelson, R.J. & Webb, N.M. (1991). *Generalizability Theory: A Primer*, Sage Publications, Newbury Park.

Shavelson, R. J. & Webb, N. M. (1991a). *Generalizability theory: A primer*. Park, Newbury CA: Sage.

Shavelson, R. J. & Webb, N. M. (1991b). *Generalizability theory: A primer (concepts in generalizbility)*. Park, Newbury CA: Sage.

Tunde, S. B. (2015). Multivariate Generalizability of NECO's 2014 SSCE objective test in electricity; Unpublished M.Phi. Dissertation presented to the Department of Social Sciences Education, Faculty of Education, University of Ilorin, Ilorin Nigeria.

Wan, C., Li, H., Fan, X., Yang, R. & Pan, J. (2014). Development and validation of the coronary heart disease scale under the system of quality of life instruments for chronic diseases QLICD-CHD: combinations of classical test theory and generalizability theory. *Health and Quality of life Outcomes Volume 12(1), 1-11*