GENERALISABILITY AND DEPENDABILITY OF STUDENTS' WORK EXPERIENCE PROGRAMME ASSESSMENT FOR QUALITY UNIVERSITY ENGINEERING GRADUATES IN NORTH-CENTRAL NIGERIA

[1]OLADELE, J. I., [2]OWOLABI, H. O. & [3]O'CONNOR, B. P.

[1]Department of Social Sciences Education,
Faculty of Education, University of Ilorin
oladele.ji@unilorin.edu.ng

[2]Department of Adult and Primary Education,
Faculty of Education, University of Ilorin
henryowo@unilorin.edu.ng

[3]Department of Psychology,
University of British Columbia-Okanagan, Canada
brian.oconnor@ubc.ca

Abstract

This study ascertained the dependability of Engineering Students' Work Experience Programme (SWEP); a period where students are given practical exposures in engineering fields of Tractor Operation/Maintenance, Central Workshop operations, Mechanical Services/Automobile, Chemical, Biomedical, Fabrication/Welding, Materials/Metallurgy, Electrical/Electronics, Water Resources/Environmental Operations, Building/Plumbingand area of specialisation after which they are assessed by technologists. The design adopted for the study was a one-facet nested fixed design. The design specifically has assessors nested within persons. The target population for the study was all the 200 level students in the Faculty of Engineering and Technology in a University in North-central Nigeria and all the technologists who took part in their assessment. A total of 517 students that were assessed in each of the engineering fields were purposely sampled for the study. Assessment scores were collated using a Profoma while the data obtained were analyzed using ANOVA option for obtaining Variance Components using GENOVA Programme. Findings revealed that generalisability and dependability coefficients was (0.10) andrelative/absolute error variances were equal (7.06) as with fully nested designs. At least, five assessment sessions for each of the twelve engineering assessment areas is required for an acceptable G-coefficient and dependability index of 0.8.The study concluded that the generalisabilty coefficient of University Engineering SWEP assessment scores in the sampled University was low and as such not dependable. It was recommended that the five assessment sessions in each of the engineering fields is required for ensuring quality in university engineering training among others.

Keywords: Generalisabilty theory, Assessment quality, University SWEP, Variance Components, Engineering students' assessment

Introduction

The main goal of educational institutions is to make learning possible. To determine whether learning has taken place and the extent to which it has, institutions of learning need to carry out several measures of learners. The quality of such measurements has been subject of interest to test experts, teachers, examination bodies and other stakeholders in education. The goals of tertiary education as stipulated in the Nigerian National Policy on Education is to contribute to national development through high level relevant manpower training, develop and inculcate proper values for the survival of the individual and society, develop the intellectual capability of individuals to understand and appreciate their local and external environment as well as acquire enabling physical and intellectual skills for self-reliance and to become useful members of the society, among others (Federal Government Nigeria-FGN, 2013).

The purpose of tertiary education is therefore to prepare young people for the job market; to have sustainable employment by enhancing technical skills and competence in a chosen field, and life skills such as problem solving and analytical skills, effective communication and literacy skills, interpersonal and team skills etc (Anho, 2011). Therefore, the quality of tertiary education system is reflected in its products and this is a pointer to the fact that education acquired is only relevant to the extent to which it makes notable impact on the lives of the individual and society. University education is at the center of human resource development; turning out the nation's graduates who are expected to be highly skilled personnel such as teachers, engineers, administrators/managers, accountants, and surgeons among other professionals. Engineering undergraduate education has evolved in ways that improve the readiness of graduates to meet the challenges of the twenty-first century while National and International organizations continue to call for change (Litzinger, Lattuca, Hadgraft & Newstetter, 2011). The engineers of the year 2020 will need to learn much new technical information and techniques. They must be conversant with and embrace a whole realm of new technologies (National Academy of Engineering, 2004). It becomes imperative to ensure that trainee engineers are thoroughly groomed in professional development.

The National Universities Commission (NUC) (2007) benchmark for minimum academic standards in Nigeria is in place to ensure that undergraduates learn through rigorous professional development courses in their various fields. The goal is to produce graduates with high academic standard and adequate practical background for self-employment as well as being of immediate value to industry and the community in general. Students' Work Experience Programme (SWEP) is an essential component of undergraduate engineering educational programme that exposes trainee engineers to on-the-job or practical experience and gives them the opportunity to put what they have learnt in the classroom to practice in real life situations. This enables them to have sound engineering skills in aspects relating to design, analysis, instrumentation, installation, maintenance, experimental and other onsite activities (Faculty of Engineering and Technology SWEP Handbook (Undergraduate), 2016a).

Adhering to the benchmark minimum academic standard for undergraduate engineering, SWEP is a 6 credit unit, compulsory course offered at 200 Level with duration of 12 weeks (NUC, 2007). During this time, students are given practical exposures in various fields of engineering such as Tractor Operation and Maintenance, Central Workshop, Mechanical Services and Automobile, Chemical, Biomedical, Fabrication and Welding, Materials and Metallurgy, Electrical and Electronics, Water Resources and Environmental

Operations, Building and Plumbing in line with the specific demands of each of their programme. After this, they are assessed in practical sessions under the supervision of technologists. This exposure is geared towards ensuring that students get the most relevant and effective experience after which they are assessed in each of these fields as a way of ascertaining their progress made.

Teachers like other professionals involved in measurement, are confronted with the task of assessing behavioral outcomes of learners consistently. This indicates reliability; one of the psychometric properties of a set of test scores that indicates how consistently a score reflects students' competence; popularly approached using Classical Test Theory (CTT). Marcoulides (2000) stated that the fundamental axiom of the CTT approach to reliability is that an observed score (X) obtained through some measurement procedure can be decomposed into the true score (T) and a random error (E) components as represented in equation 1.

$$X = T + E \qquad \text{……………. (Equation 1)}$$

The E in the equation connotes error which could be due to the test form (Parallel Forms Reliability), item (Internal Consistency Reliability), rater (Inter-rater Reliability) or occasion (Test-retest Reliability) exclusively in each analysis.

The better an assessment procedure is at providing an accurate indication of an individual's true score, the higher and more accurate the T component will be and the smaller the E component. In practical applications of CTT, reliability is typically defined as the ratio of the true score variance to the observed score variance (Junker, 2012). This indicates the proportion of the observed score variability that is attributable to true score variability across individuals, and the reliability coefficient increases as error decreases. In essence, the CTT technique provides a variety of useful individual methods for assessing reliability. As useful as these methods are, measurement error in CTT is an undifferentiated random variation and so, the theory does not distinguish among various possible sources. Resultantly, the assumptions of a single undifferentiated random error that follow the fundamental axiom of the CTT approach are often unrealistic (Brennan, 2001a). Marcoulides (2000) further stressed that the existence of a single undifferentiated random error in the measurement procedure is quite questionable and this is a deficiency Generalisability theory (G-theory) remedies. G-theory recognizes that multiple sources of error; attributable to the test items, testing occasions, examiners and examinees among others may occur simultaneously in a measurement procedure and as such, an observed test score is the sum of an unobservable true score T and multiple error components each denoted E_k.

$$X = T + E_1 + E_2 + \ldots\ldots + E_k \qquad \text{……………. (Equation 2)}$$

Where $E_1 + E_2 + \ldots\ldots + E_k$ are multiple sources of error which could be due to form (Parallel Forms Reliability), item (Internal Consistency Reliability), rater (Inter-rater Reliability) or occasion (Test-retest Reliability) among others that could be accounted for in a single analysis.

This stand as evidence that G-theory extends CTT by providing a flexible and practical framework for estimating the effects of multiple sources of error in a single analysis, thereby providing more detailed information that researchers and administrators can use in deciding how to improve the usefulness of an assessment procedure for making decisions or drawing conclusions. Thus, the focus of the CTT concept of reliability is expanded in G-theory to address a broader issue of how accurately a researcher can generalize about a persons' behavior in a defined universe from observed scores. This concept is construed as dependability in G-theory (Egbulefu, 2013). According to Shavelson and Webb (1991), dependability refers to the accuracy of generalizing from a person's observed score on a test; that is, the generic term for a measure such as behavior observation or opinion survey to the average score the person would have received under all possible test conditions that the test user would be equally willing to accept having put into consideration all the possible sources of error. As such, G-theory estimates multiple sources of error separately in a single analysis which guarantees its dependability.

In G-theory, the sources of variation are associated with the persons being measured referred to as the objects of measurement or differentiation facet as it is not seen as a source of measurement error and potential source of error arising from the testing situation, such as questions, invigilators, examiners, occasions; referred to as instrumentation facets, being that they contribute to error variance. These testing situations are called facets in G-theory and each facet is composed of one or more levels or conditions. Although, the choice and number of facets in a G-study may vary according to the interests of the researcher, the object of measurement is always being included as a distinct facet. In G-theory, objects of measurement are usually persons that are measured. However, some testing contexts have other entities such as classrooms or schools as the objects of measurement (Shavelson& Webb, 2005).

G-theory can be carried out in two stages which are Generalisability studies (G-studies) and Decision studies (D-studies). G-study is the first stage in applying G-theory and it plays a similar role to the investigation of reliability in a traditional measurement. G-study is used to isolate and estimate as many facets of measurement error as is reasonably and economically feasible expressed as variance components and it yields a generalisability coefficient. It specifies the universes of admissible observations; that is, all possible combinations of the levels of the facets which are denoted using single lower case letters of alphabets such as p for persons, a for assessor, r for raters, t for task, o for occasion, f for forms of test etc. (Brennan, 2001a). D-study deals with the practical application of a measurement procedure within a universe of generalization; that is, a set of facets and their levels (e.g., items and occasions) to which a decision maker wants to generalize. D-study makes use of variance component information from a G-study to design a measurement procedure or possible application of an instrument that minimizes error for a particular purpose and it yields a dependability coefficient expressed as Ø (phi) (Shavelson& Webb, 2005).

According to Shavelson and Webb (2005), G-theory could either adopt a crossed design with the notational symbol 'x' (crossed with) or nested design with the notational symbol ':' symbol (nested within). These designs could be random or fixed. Similar to the indices of reliability coefficients, Generalisability and Dependability coefficients range from 0 to 1.0., with acceptable coefficients ranging from .80 or higher (Brennan, 2001a). Though, dependability and reliability can be used interchangeably when referring to what will later be labelled as "person variance", dependability also refers to variances associated with many sources of potential measurement error (Shavelson& Webb, 1991).G-theory is useful not only for understanding the relative importance of various sources of error but also for designing efficient assessment procedures. This strengthens the fact that G-theory does not only encompasses CTT as a special case, but goes far beyond it in clarifying conceptual confusions and providing more powerful statistical analysis (Brennan, 2011). The onus is therefore on assessment experts to employ the use of G theory in ensuring institutional quality among others.

Egbulefu (2013) estimated measurement error and score dependability in examinations in order to determine the extent of the contributions of these sources of error (facets) to examination scores. Findings revealed that the residual made the highest contribution to measurement error. Also increasing the number of invigilators to 90, increased the Generalisability coefficient and Index of dependability which rank ordered students and classified them based on their performance, irrespective of the performance of other students.Bamidele (2015) conducted a study on analysis of multivariate Generalisability of national examinations council's 2014 senior school certificate examinations objective tests in electricity, in order to determine the generalisability coefficient and the index of dependability of the relative variance error and absolute variance error for decision making on every item. Results revealed that persons crossed with (x) items interactions had the highest percentage than other variance components while the G coefficient was 0.54; another indication that the coefficient of the test related to persons was low. Kassab, Fida, Radwan, Hassan, Abu-Hijleh and O'Connor (2016) applied G-theory to Problem-Based Learning (PBL) where students construct concept maps that integrate different concepts related to the PBL case and are guided by the learning needs generated in small group tutorials. The goal was to develop an instrument to measure students' concept maps in PBL programmes as well as assess its psychometric properties. Results revealed that students' concept map scores (universe scores) accounted for the largest proportion of total variance (47%) across all score comparisons while the conducted D-study suggested that a dependability level of 0.80 can be achieved by using three raters who each score two concept map domains, or by using five raters who each score only one concept map domain. Mahmud (2017) used G-theory to determine the dependability level of undergraduate students' teaching practice scores conducting both G-study and D-study. The G-study revealed that the largest variance was accounted for by the residual while the Generalisability and Dependability coefficients; being low with six occasions and five raters being the least combination required obtaining 0.80 dependability level for the Generalisability and dependability coefficients. Aside Mahmud (2017) whose study was on dependability of teaching practice programme, none of these studies centered on institutional assessment quality and thus a dart of literature in this area. Applying G-theory statistics to analyzing the Generalisability of students' SWEP scores becomes necessary as this will aid determining the dependability of SWEP assessment as well as aid designing a more effective assessment procedure that could improve the dependability of scores produced and this stands as the main purpose of this study. Specifically, this study was designed to:

1. obtain the generalisability coefficient of university engineering SWEP scores;
2. obtain the dependability index of university engineering SWEP scores; and
3. estimate the number of assessors sufficient to get a generalisability coefficient and dependability index of 0.80 respectively in university engineering student's SWEP assessment scores.

Research Questions

The study was tailored to answering the following research questions in carrying out this study:

1. What is the generalisability coefficient of University Engineering SWEP scores?
2. What is the dependability index of University Engineering SWEP scores?
3. How many assessors are sufficient to get a generalisability coefficient and

dependability index of 0.80 respectively from University Engineering student's SWEP assessment?

Methodology

The design adopted for this study was a one-facet nested fixed design of assessors nested within persons (a:p). The design is one-facet as though, there are two annotations in the design, here persons and assessor, persons are the objects of measurement and so are not a source of error and, therefore, is not a facet leaving assessor as the only facet in the study. There are twelve assessment areas during the SWEP engineering programme. This design is seen as nested as a unique assessor rates persons on each of the twelve assessment areas. The design is also fixed because the conditions of the facets exhaust the conditions of the universe to which the researcher wants to generalize; as such, all conditions are included in the measurement design. This design was employed so as to estimate the contributions of the facet in the study to measurement error in the university SWEP assessment which covers the G-study aspect of the design.

The nested research design employed in this study had two (2) distinct effects and variance components (VERCOMS) analysis procedure was employed to ascertain the contribution of these effects. The variance components for a one-facet nested design $\sigma^2_{(a:p)}$, is given in Equation 3.

$$\sigma^2 X_{pa} = \sigma^2_p + \sigma^2_{a,pa,e} \qquad \text{............... (Equation 3)}$$

Where:

$\sigma^2 X_{pa}$ = Variance component of Grand mean

σ^2_p = Variance component for person effect (students)

$\sigma^2_{a,pa,e}$ = Variance component for the assessor effect confounded within person by assessor interaction and residual

The one-facet nested design is represented with the Venn diagram in Figure 1.

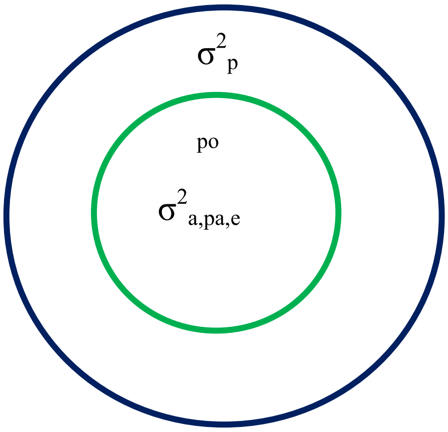

Figure 1: Venn diagram for (a:p) design

The one-facet nested design has no separate term for the assessor effect.

Rather, it is part of the residual term. Different assessors rated each person and as such, the assessor effect cannot be estimated independently of the person-by-assessor interaction and thus assessor and residual are confounded. This is why the full form of the residual effect shows the assessor effect as part of the residual term (Shavelson & Webb, 1991).

The population of the study was University Engineering students while the target population for the study was all the 200 level students in the Faculty of Engineering and Technology in a University in North-central Nigeria who underwent the SWEP in 2016/2017 academic session; a total of 634 and all the technologists who took part in each of the 12 assessments they were exposed to. A total of 517 students that were assessed in each of the twelve engineering fields were purposely sampled for the study. Purposive sampling was to ensure that students who had complete score in the 12 engineering assessment fields participated in the study. Assessment scores on the 12 assessment fields of Tractor Operation and Maintenance, Central Workshop, Mechanical Services, Automobile Services, Chemical, Fabrication and Welding, Materials and Metallurgical, Electrical and Electronics, Water Resources & Environment, Building and Plumbing, Biomedical and area of specialization assessments in line with the specific demands of each of their programme. were collated using a Profoma which is a collation rate of 82% while the data obtained were analyzed using Analysis of Variance (ANOVA) option for obtaining Variance Components using GENOVA Programme (Version 3.1).

Results

Research Question One: *What is the generalisability coefficient for University Engineering SWEP scores?*

To answer research question one, the G-coefficient is arrived at in 3 stages.

Stage 1: Determine the contributions of the identified sources to measurement error in university engineering SWEP assessment

The first stage is to determine the contributions of the identified sources to measurement error in university engineering SWEP assessment based on the estimated variance components and percentage of total variance produced of person (p) and assessor (a,pa,e). The result of the analysis as given by GENOVA output is shown in Table 1.

Table 1: Estimated Variance Components of Person (p) and Assessor (a,pa,e)

Source of Variation	Sum of Squares	df	Mean Square	Estimated Variance Components	Percentage of Total Variance (%)
Persons (p)	0.38	516	45.82	0.06	0.1
Assessor (a,pa,e)	1.97	5687	461896.28	84.66	99.9
Total	**2.35**	**6203**			**100**

As shown in Table 1, the estimated variance component for persons (σ^2_p) is 0.06 which accounted for 0.1% of the total variance in students' SWEP scores while the estimated variance component for assessor, confounded with person by assessor interaction and the residual ($\sigma^2_{a,pa,e}$) is 84.66 accounted for 99.9% of the total variance. This shows that the assessor effect confounded with person by assessor interaction as well as the residual ($\sigma^2_{a,pa,e}$)

contributed more to measurement error in university engineering SWEP scores.

Stage 2: Determine the relative and absolute error variances

The relative and absolute error variances for a one-facet nested design are computed by dividing the estimated variance for residual by the number of assessors as shown in the formula:

$$\sigma^2_{Rel} = \sigma^2_{Abs} = \frac{\sigma^2_{apa,e}}{n'_a} \quad \text{...........Equation 4}$$

Since the residual term is confounded with the assessor effect, the variance for the assessor effect: 88.66 and number of assessors: 12 is substituted into Equation 4 and used for the computation as shown in Table 2.

Table 2: Relative and Absolute Error Variances (One-Facet Nested Design for SWEP)

Relative Error Variances	Absolute Error Variances
7.06	7.06

As shown in Table 2, the relative error variance is 7.06 while the absolute error variance is also 7.06. This shows that the relative and absolute error variances are the same.

Stage 3: Calculate the generalisability coefficient ($E\rho^2$) of University Engineering SWEP scores

The generalisability coefficient is calculated by substituting the variance components obtained into the formula in Equation 5 and results shown in Table 3.

$$E\rho^2_{Rel} = \rho^2 = \frac{\sigma^2_p}{\sigma^2_p + \sigma^2_{Rel} = \sigma^2_{Abs}} \quad \text{………. (Equation 5)}$$

Table 3: Generalisability Coefficient of University Engineering SWEP scores

	Generalisability Coefficient
$\sigma^2_{Rel} = \sigma^2_{Abs}$	7.06
$E\rho^2$	0.01

As shown in Table 3, the generalisability coefficient of university engineering SWEP scores was 0.01. This reveals that the generalisability coefficient of university engineering SWEP scores was low considering the acceptable generalisability coefficient benchmarked at 0.80.

Research Question Two: *What is the dependability index (Φ) of university engineering SWEP scores?*

To answer research question two, the dependability index is calculated by substituting the variance components obtained into the formula in Equation 6and shown in Table 4.

$$\Phi = \frac{\sigma_p^{\ 2}}{\sigma_p^2 + \sigma_{Rel}^2 = \sigma_{Abs}^2} \text{............ (Equation 6)}$$

Where: $\sigma_{Rel}^2 = \sigma_{Abs}^2 = \frac{\sigma_{apa,e}^2}{n_a^{/}}$ .... (Equation 7)

Table 4: Dependability Coefficient of University Engineering SWEP scores

	Dependability Coefficient
$\sigma_{Rel}^2 = \sigma_{Abs}^2$	7.06
Φ	0.01

As shown in Table 4, the dependability coefficient of university engineering SWEP scores was 0.01. This reveals that the dependability index of university engineering SWEP scores was also low considering the acceptable dependability index benchmarked at 0.80.

Research Question Three: *How many assessors are sufficient to get at least a generalisability coefficient and dependability index of 0.80 respectively from university engineering students' SWEP assessment?*

To determine the number of assessors sufficient to get at least a generalisability coefficient and dependability index of 0.80 respectively during university engineering students' SWEP assessment, a decision study was carried out using the results obtained for the generalizabilty study as shown in Table 5.

Table 5: Summary Table on D-study Result on the Number Assessors sufficient to get at least a $E\rho^2$ & (Φ) of 0.80 respectively

	Sample Sizes			Variances				
IDEX= UNIV.=	**$P INF.**	**A INF.**	**UNIVERSE SCORE**	**EXPECTED OBSERVED SCORE**	**LOWER CASE DELTA**	**UPPER CASE DELTA**	**MEAN**	**GEN. COEF.**
	517	24	1.54533	2.68174	1.13641	38.35770	37.22647	0.57624
	517	36	1.54533	2.30294	0.75761	25.57180	24.81865	0.67102
	517	48	1.54533	2.11353	0.56821	19.17885	18.61473	0.73116
	517	60	1.54533	1.99989	0.45457	15.34308	14.89238	0.77271

As shown on Table 5, at least 60 assessors in five assessment sessions for each of the 12 engineering assessment areas in a partially nested design are required for an acceptable generalisability coefficient and dependability index of 0.8.

Discussion of Findings

The findings of the study revealed that both generalisability and dependability coefficients were (0.01) and are low considering the acceptable generalisability coefficients of 0.80. This finding is in line with Bamidele (2015) who reported a low G coefficient of 0.54 analyzing National Examinations Council's 2014 Senior School Certificate Examinations objective tests in Electricity. This could be as a result of the students' being assessment once on each of the twelve engineering fields which makes the design uni-variate. Re-assessing the students on each of the engineering fields transforms the design to a multivariate designs which has been proved efficient in raising the G coefficients and dependability index with empirical evidence by studies carried out by Anji and Michael (1997).

Similarly, the findings of Mahmud (2017); though carried out as a multivariate design reported a low generalisability coefficient of 0.36 and dependability coefficient of 0.30 for undergraduate student's teaching practice scores. The researcher explained that the low dependability could be as a result of the contributions of the sources of measurement error to the variations in students' scores, the number of occasions and raters used for the programme as well as a poor coordination of assessors on scoring procedures. His findings could also be as a result of sampling error. Contrary to these findings, that of Gugiu, Gugiu, and Baldus (2012) who found a very high interrater reliability coefficient (0.929) for only two raters who received no training in how to use the four grading rubrics. Similarly, Vafaee and Yaghmaeyan (2015) revealed that the dependability is high enough to be taken as a consistent measure of the speaking ability of the test takers using scores obtained from the Columbia University placement speaking test. Both findings could be as a result of combining scores from multiple (two for the former and four for the later) analytic rubric scales to make a composite score.

On the number of assessors sufficient to get at least a generalizabilility and dependability coefficient of 0.80 respectively, five assessment sessions for each of the twelve engineering assessment areas is required to attain an acceptable generalisability coefficient and dependability index of 0.8. As such, increasing the number of assessors resulted to an increase in both generalisability coefficient and dependability index which was a positive relationship. Though divergent from the findings of Lee (2006) who submitted that it would be more efficient maximizing score dependability by increasing the number of tasks rather than the number of assessor per speech sample in a multi-task speaking measure consisting of both integrated and independent tasks, being an important component of a new version of the Test of English as Foreign Language (TOEFL), the findings of this study are in line with Shavelson, Baxter and Gao, (1993) who submitted that large numbers of assessments are needed to get a reliable measure of mathematics and science achievement at the elementary level.

The findings of this study also aligned with that of Egbulefu (2013) who carried out a study on the estimation of measurement error and score dependability in Senior Secondary Examinations. The researcher concluded that increasing the number of invigilators to ninety lead to an increase in the generalisability coefficient as well as the index of dependability which rank ordered students and classified them based on their performance absolutely. The study of Mahmud (2017) was also in congruence with this study who determined the dependability level of scores produced during teaching practice I and II exercises of undergraduate students. The study revealed that increasing the number of occasions to six and raters to five was required to obtain a dependability index of 0.70.

The findings of Kassab et al (2016) was a mix suggesting that a dependability level of 0.80 can be achieved by reducing the number of raters to three on the condition that each rater score two concept map domains or increasing the number raters to five who each score only one concept map domain. However, most of the studies pointing to an increase in the number of raters, and as such, the researcher concluded that increasing the number of the facets in generalisability theory guarantees the generalisability and dependability levels in educational assessments. This would be particularly helpful as it would allow for the introduction of the occasion facet which would transform the design from a fully nested (a:p) to partially nested p x (a:o) design. This would allow for distinguishing between all the sources of measurement error as well as distinguish between the relative and absolute error.

Conclusion

Based on the findings of this study, it was concluded that the generalizabilty and dependability coefficients of University Engineering Students' Work Experience Programme assessment scores in the sampled university are low.

Recommendations

Based on the findings of this study, as well as the drawn conclusions, the following recommendations were proffered:

1. The SWEP assessment protocol should be reviewed by increasing the number of assessments for each of the twelve assessment areas as it has been proved to provide the desirable generalisability coefficient and dependability index of students' SWEP assessment scores. To this effect, the university administration should make necessary provisions both human and financial that would accompany the increment in assessment sessions.

2. Engineering regulatory bodies such as the National Universities Commission (NUC), Council for the Regulation of Engineering in Nigeria (COREN) and Faculties of Engineering should review the frequency of SWEP assessments for University Engineering Students' Work Experience Programme so as to accommodate the increment in the number of assessments needed for the desirable generalisability coefficient and dependability index.

3. Psychometric experts should ensure that students' scores are subjected to generalisability analyses so as to estimate multiple sources of error and reduce or eliminate measurement error and hence ensure dependability as against the common use CTT techniques in order to harness the strengths of G-theory to determine the measurement conditions necessary to make reliable criterion-referenced decisions.

References

Anho, J.E. (2011). An Evaluation of the Quality and Employability of Graduates of Nigerian Universities. African Journal of Social Sciences, 1(1), 179-185.

Anji, V. & Michael J. (1997). Assessing Reliability of Student Ratings of Advisor: A Comparison of Univariate and Multivariate Generalisability Approaches.Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28,1997).

Bamidele, S. T. (2015). Analysis of Multivariate Generalisability of National Examinations Council's 2014 Senior School Certificate Examination Objective Tests in Electricity.University of Ilorin, Ilorin: An unpublished M. Phil. Thesis.

Brennan, R.L. (2000). Performance Assessments from the Perspective of Generalisability Theory.Applied Psychological Measurement, 24(4), 339–353. Newbury Park CA: Sage Publications, Inc.

Brennan, R.L. (2001a). Generalisability Theory. New York: Springer-Verlag.

Brennan, R.L. (2001b). Manual for urGENOVA (Version 2.1).Lowa Testing Programme Occasional Papers, Number 49.Lowa: The University of Lowa.

Brennan, R.L. (2011). Generalisability Theory and Classical Test Theory.Applied Measurement in Education, 24: 1–21.Lowa: Taylor & Francis Group, LLC. DOI: 10.1080/08957347.2011.532417.

Egbulefu, C.A. (2013). Estimating Measurement Error and Score Dependability in Examinations Using Generalisability Theory.University of Nigeria Nsukka, Nsukka: An unpublished PhD Thesis.

Faculty of Engineering & Technology, University of Ilorin (2016).SWEP Handbook (Undergraduate). Ilorin: Olad Publishers.

Federal Government of Nigeria (FGN) (2013).National Policy on Education (NPE). Yaba: NERDC Press.

Gugiu, M.R., Gugiu, .C. and Baldus, R. (2012). Utilizing Generalisability Theory to Investigate the Reliability of the Grades Assigned to Undergraduate Research Papers. Journal of MultiDisciplinary Evaluation, 8(19), 26-40.

Junker, B.W. (2012). Some aspects of Classical Reliability Theory & Classical Test Theory. Retrieved from: http://www.stat.cmu.edu/~brian/PIER-methods/For%202013-02-26/Readings/CTT-formulas.pdf. Pittsburgh: Carnegie Mellon University.

Kassab,S.E., Fida, M., Radwan, A., Hassan, A.B., Abu-Hijleh, M. & O'Connor, B.P. (2016). Generalisability theory analyses of concept mapping assessment scores in a problem- based medical curriculum. MedicalEducation, (50): 730–737. doi: 10.1111/medu.13054

Mahmud, J.A. (2017). Analysis of Dependability of Undergraduate Students' Scores in Teaching Practice Courses in a Nigerian University. An Unpublished PhD Thesis: University of Ilorin.

Lee, Y. (2006). Dependability of scores for a new ESL speaking assessment consisting of integrated and independent tasks.Language testing, 23, 131-166.

Litzinger, T., Lattuca, L.R., Hadgraft, R. &Newstetter, W. (2011).Engineering education and the development of expertise.Journal of Engineering Education (JEE), 100(1), 123–150. DOI:10.1002/j.2168-9830.2011.tb00006.x

Marcoulides, G.A. (2000). Generalisability theory Handbook of Applied Multivariate Statistics and Mathematical Modeling.Ebook: Academic Press. Available at: https://www.elsevier.com/books/handbook-of-applied-multivariate-statistics-and-mathematical-modeling/tinsley/978-0-12-691360-6

National Academy of Engineering (2004).Educating the engineer of 2020: adapting engineering education to the new century/National Academy of Engineering of the National Academies. Washington DC: National Academy of Sciences.

National University Commission (NUC) (2007). Benchmark Minimum Academic Standards for Undergraduate Programmes in Nigerian Universities (Engineering and Technology). Abuja: NUC.

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

Shavelson, R. J., Baxter, G., &Gao, X. (1993). Sampling variability of performance assessments.Journal of Educational Measurement, 30, 215-232.

Shavelson R.J. & Webb N.M. (2005).Generalisability Theory: Overview. Encyclopedia of Statistics in Behavioral Science, 2, 717–719.

Shavelson R.J. & Webb N.M. (2005).Generalisability theory. 36-Green.qxd 599-612.

Vafaee, P. &Yaghmaeyan, B. (2015).Providing evidence for the Generalisability of a speaking placement test Scores.Iranian Journal of Language Testing, 5(2), 78-95.