Estimating Measurement Error in University Students' Work Experience Programme (SWEP) Assessment using Generalizability Theory: Implications for Transdisciplinarity in Engineering Training

Jumoke I. Oladele

Department of Science and Technology Education, Faculty of Education, University of Johannesburg, South Africa, Email: jumokeo@uj.ac.za

Received 23 March, 2022; Revised 24 May, 2022; Accepted 24 May, 2022

Available online 25 May, 2022 at www.atlas-tjes.org, doi: 10.22545/2022/00185

Test experts and other stakeholders in education are interested in the quality of students' assessments, while traditional reliability indices are giving way to Generalisability statistics encouraging transdisciplinarity in problem-solving. This study aimed at estimating measurement error in University Students' Work Experience Programme (SWEP) assessment using generalizability theory. The design adopted for the study was a one-facet nested fixed design with assessors nested within persons. The study's target population was all the 200 level undergraduate Engineering students registered for 2015/2016 academic sessions in the Faculty of Engineering in a Nigeria University and all the technicians who assessed them, with 591 students purposively selected. Their assessment scores were collated using a Proforma. Data obtained were analyzed using ANOVA statistics. Findings revealed that the assessor effect confounded the person by assessor interaction. The residual contributed more to measurement error in university engineering SWEP scores; the relative and absolute error variances were equal, with a value of 8.02. Therefore, relative and absolute interpretations cannot be distinguished. Based on these findings, educational researchers engaging in Genralizabilty theory analysis should employ the crossed facets G-study designs in assessment designs to fully harness G-theory's strength to distinguish between the absolute and relative decisions. This conclusion is so as decision-makers may be interested in one or both decisions while interpreting measurement results in defining error and Generalisability coefficients to employ synergetic efforts to ensure quality engineering education through a transdisciplinary approach to their training.

Keywords: Measurement Error, University, SWEP, Educational Assessment, Generalizability Theory, Transdisciplinarity.

1 Introduction

Any nation's technological and industrial development depends on its ability to develop its citizens towards human resources, especially in science and engineering. Engineering education is deployed non-formally through apprenticeship and formally through the universities and polytechnics to produce professionals in engineering (Idris& Rajuddin, 2012). The regulatory bodies in charge of educational sectors at tertiary levels in Nigeria constitute the National Universities Commission (NUC), the National Board for Technical Education (NBTE), and the National Commission for Colleges of Education (NCCE). These are the main organs of the Federal Ministry of Education that ensure the proper management of the entire universities, Polytechnics, Monotechnics, and Colleges of Education. The bodies are responsible for planning, organizing, managing, funding, monitoring, and supervision (Ali & Rajuddin, 2012). SWEP, as an essential multidisciplinary component of the undergraduate engineering education program in Engineering and Technology, exposes trainee engineers to on-the-job or practical experience and allows them to put what they have learned in the classroom to practice in real-life situations. This component enables them to have sound engineering skills in design, analysis, instrumentation, installation, maintenance, and other onsite activities (Faculty of Engineering and Technology SWEP Handbook (Undergraduate), 2016). Scores obtained from the SWEP assessment are used for computing students' Cumulative Grade Point Average (CGPA), which largely determines students' class of degree is a 6-credit course.

The quality of engineering graduates from universities and polytechnics in Nigeria has been the industry's major concern. Quality employable skills are required, especially in the cutting-edge innovation in training and practical know-how (Okafor et al., 2020). Oladele et al. (2020) reported that the generalizability and dependability of Engineering assessment scores were low. This lag in training was attributed to poor infrastructural facilities haunting tertiary institutions in Nigeria, coupled with large classes (Olorunfemi & Ashaolu, 2008). As a remedy to the observed lag in a multidisciplinary approach to training, employers of labor in engineering engage in several re-training of employees to enhance the required on-the-job skills. The re-training exercise can be termed as labor loss which calls for a pragmatic approach to trans-disciplinarity where all hands are placed on deck for sustainability with problem-solving (Popa et al., 2015). The pedagogical approach to teaching a subject has implications for how students learn and use acquired knowledge effectively for professional practice; how a subject is taught is integral to defining the subject (Grossman, 2009). The consensus approach to trans-disciplinarity in pedagogical designs seeks to unify knowledge. It involves teaching from a universal perspective: presenting a subject from many perspectives without any special recourse or exclusive service to a specific discipline while ensuring a bottom-top mechanism (Alonge et al., 2017).

Trans-disciplinarity is a relatively young concept developed by a Swiss philosopher and psychologist Jean Piaget between 1896-1980, geared toward the epistemology and the planning of future universities and educational programs (Bernstein, 2015). Transdisciplinarity employs an across-multiple domain concurrently with about four decades of practice which emerged seven centuries after disciplinarity evolved (Stroud, 2020). Employing a trans-disciplinary approach, engineering educators can be supported by assessment experts in ensuring quality engineering education. In del Cerro Santamaría's (2020) views, transdisciplinarity can also be seen as an evolution of multi and inter-disciplinarity. However, unlike the latter, transdisciplinarity does not seek to solve the paradoxes generated by the endless dissection of knowledge in smaller disciplinary units. Rather than aiming for the "unity of knowledge" by acknowledging the inherent complexity of the subject, transdisciplinarity directs to master the paradoxes. Transdisciplinarity is characterized by its focus on "wicked problems" that need creative solutions, its reliance on stakeholder involvement and engagement, and socially responsible science while providing an intriguing potential to invigorate scholarly and scientific inquiry both in and outside the academy (Bernstein, 2015). In the context of this study, transdisciplinarity connotes a hybrid knowledge among the exact sciences, the social sciences and philosophy, integration between theory and practice, ethical concerns, and the importance of experimental, designerly modes of inquiry resulting in articulations, rather than the relations, between disciplines; the whole being more than the sum of its parts with assessment as a major player in the teaching and learning process for ascertaining the extent to which learning is being achieved.

Assessment is a process by which the extent to which learning objectives are achieved is ascertained, broadly termed a test. A test is one of the many assessment instruments most appropriate for cognitive ability testing (McCulloch, 2007). According to Kizlik (2012), learner assessment is best seen as a two-way communication in which feedback is given to instructors on teaching and students as feedback on learning. For curriculum designers, carefully designed learner assessment instruments can help determine whether the pre-stated objective is achieved. Therefore, assessments in education provide quantifiable means of gauging students' learning, among other purposes. Whether implicit or explicit, an assessment should yield information relative to a pre-specified objective (Kizlik, 2012). According to Brennan (2001a), it is important to pay careful attention to assessment procedures; used for acquiring information about various learners' attributes. However, a measure of imperfection is found with the information obtained from educational assessments, like those obtained from the physical sciences depending on various measurement conditions, such as the meter rule, the measurement recorder, environmental conditions, and the like, which results in measurement error. A primary goal of educational and psychological measurement is to carry out measurements with as little measurement error as possible.

Measurement error is when the true estimate is either below or above what was obtained (Hofmann, 2005). Since the error is an indispensable aspect of observed scores, the goal is to conduct assessments with minimal measurement error. Many believe that public examination bodies that conduct standardized assessments at basic and secondary levels ascertain these qualities. These efforts should also be advanced to all educational programs at the tertiary levels and argued that most error is introduced with a poor sampling of observations most applicable with essay response formats, ability tests, self-report questionnaires, and rated observations (Tindal, Yovanoff, & Geller, 2010; Lakes & Hoyt, 2009). The systematic influences on responses and scores comprise the. The preceding reveals that items, raters, formats, and occasions are sampling spaces for any measurement that can lead to error (Tindal, Yovanoff & Geller, 2010). Therefore, the quality of measurement procedures at all levels of education is of continued interest to test experts, teachers, examination bodies and other stakeholders in education. Students' Work Experience Programme (SWEP) is a practical exercise designed to ensure engineering graduate quality.

These bring to bear the importance of getting the SWEP assessment right in terms of consistency of scores obtained, also known as reliability. Like other professionals involved in measurement, teachers are confronted with consistently assessing the behavioural outcomes of learners. Reliability indicates how consistently a score reflects students' competence; popularly approached using Classical Test Theory (CTT). Marcoulides (2000) explained that with the CTT approach, an observed score (X) consists of the true score (T) and a random error (E) component, as represented in equation 1.

$$X = T + E \tag{1}$$

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. However, the higher the T component and the smaller the E component leads to assessment accuracy in providing an individual's true score (Junker, 2012). Studies on reliability with CTT revealed that perpetual inventory used for the construction of education data per country (Portela, Alessie, & Teulings, 2006); value-added estimate for teacher-level analyses (Schochet & Chiang, 2010); self-reported BMI is subject to substantial measurement error (O'Neill & Sweetman, 2013) lead to systematic measurement error. Measurement error in CTT is undifferentiated and included in the true score, which are constituents of an observed score with a heavy reliance on the true score test model; that is, the aggregate score. A primary assumption of CTT, as pointed out by Minor (2013), is that individuals possess stable traits or characteristics. Though CTT has been useful for determining measurement reliability, only one source of measurement error can be considered at a time, while the interaction of the various sources of error cannot be estimated (Brennan, 2011). This fact is a limitation of CTT. G-theory caters for where sources of error due to the items, timing, examiners, and examinees may occur simultaneously in a measurement procedure. Therefore, an observed score is the sum of an observed true score-T and multiple sources of error, each denoted E_k.

$$X = T + E_1 + E_2 + \cdots + E_k \tag{2}$$

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.

Efforts have been made to address the deficiency in using scales at the lower levels of measurement: the nominal and ordinal scales of measurement and the undifferentiated measurement error in CTT. These efforts have resulted in the birth of reliable statistical techniques such as Item Response Theory (IRT); some of the earliest authors include Lord, Hambleton, Swaminathan, and Rogers, and Rasch, Baker (2001) and Pelton (2002). IRT is premised on individual items of a test rather than on test scores as practiced with CTT. The theory is popularised by its growing use in standardized examinations, owing to its significantly improved measurement accuracy and reliability. Thornton (2002) pointed out that although IRT has provided a framework under which dichotomous and polytomous responses to items can be modeled under a specific set of assumptions, its modeling does not accommodate quantifying the relative contributions of different error sources. This fact makes it rather a uni-dimensional model, which is a divergent model from the true score to analyzing assessments based on each item. Uni-dimensional methods can only be seen as yet inconclusive as one instrument can only measure one construct or factor of ability, personality, affective, and attitude. In contrast, several factors exist concurrently in the real world that could be discovered during measurement.

The preceding arguments show that behavioral measurements are more multi-dimensional with cognitive and non-cognitive constructs. Traditional reliability indices give way to Generalisability statistics as it caters to the multi-faceted nature of measurement error in educational assessments in a single analysis. Therefore, psychometricians should involve multi-dimensional model analysis techniques like G-theory (Margono, 2015). Brandt (2008) observed that achievement tests mostly within large-scale assessments; deal with measuring abilities with sub-abilities. Having established the presence of error in measurement, with G-theory, the error can be decomposed into components related to various facets that make it multi-dimensional. This attribute stands as a remarkable difference between G-theory and CTT. Furthermore, G-theory uses Analysis of Variance statistical methods with which multiple sources of error that contribute to the undifferentiated E are disentangled (Brennan, 2001a).

G-theory also provides a trans-disciplinary approach to understanding the dependability of measures (Shavelson & Webb, 1991) and allows accurate assessment of the reliability of complex measures and measures used for either relative decisions or criterion-referenced decisions. Classical Test Theory estimates separately only one source of error at a time; that is an error due to test forms (Parallel Forms Reliability), item (Internal Consistency Reliability), rater (Inter-rater Reliability) or occasion (Test-retest Reliability), while G-theory estimates multiple sources of error separately in a single analysis. Noteworthy is the fact that whereas CTT deals with the reliabilities of relative decisions, G-theory distinguishes between relative ("norm-referenced") and absolute ("criterion" or "domain-referenced") decisions (Shavelson & Webb, 2005). Furthermore, in G-theory, the sources of measurement error associated with the entity being measured are usually termed as the objects of measurement (OoM), also known as the differentiation facet. OoM is not seen as a potential source of error arising from the testing situation, such as item, supervisor, rater, and time; all facets in G-theory contribute to error variance. With the choice and number of facets in a G-study usually up to the researcher, the object of measurement should be the objects of measurement in G-theory included as a distinct facet are usually persons. Although, some testing contexts have other entities, such as classrooms as the objects of measurement, among others (Shavelson & Wedd, 2005).

Shavelson, Baxter, and Gao (1993) studied sampling variability of performance assessments and reported the residual effect as the major source of measurement error followed by person x task interactions in performance-based assessment. Egbulefu (2013) estimated measurement error and score dependability in examinations using generalizability theory and reported that the residual contributed mostly to measurement error. Kassab et al. (2016) applied generalisability analysis to concept mapping assessment scores in a problem-based medical curriculum, Mahmud (2017) carried out a study on the dependability of teaching practice assessment, while Brown et al. (2017) compared the computer and hand administered versions of the Brown Location Test, Alonge et al. (2017) recommended a trans-disciplinary approach for teaching implementation research and practice in public health. Oladele et al. (2020) studied the generalizability and dependability of Engineering SWEP Assessment scores. The studies show that efforts have been made to estimate measurement error in assessments carried out in core professions such as education and medicine, among others; none of these studies was centered on estimating Measurement Error in University Engineering SWEP Assessment using Generalizability Theory which is the gap in literature this study is poised to fill. Specifically, this study:

i. estimate contributions of persons and assessors to measurement error of University Engineering SWEP assessment scores;
ii. estimate values for the relative error variances of University Engineering SWEP assessment scores;
iii. estimate values for the absolute error variances of University Engineering SWEP assessment scores; and
iv. determine interpretations made from the relative and absolute error variances obtained.

The study was tailored to answer the following research questions:

1. What are the contributions of persons and assessors to the measurement error of University Engineering SWEP assessment scores?
2. What is the relative error variance of University Engineering SWEP assessment scores?
3. What is the absolute error variance of University Engineering SWEP assessment scores?
4. What interpretations can be made from the relative and absolute error variances?

2 Conceptual Framework

Transdisciplinarity applied to conceptual change was adapted for this study (McGregor, 2014). The author extracted how conceptual change theory can contribute to deeper understanding of what is conceptually involved when people attempt (or succeed) to transition from multi- and interdisciplinarity to transdisciplinarity, sharing a comparison of multi-, inter-, and transdisciplinarity. These comprised three axioms, which are the overt multiple Levels of Reality and knowledge as complex, emergent, and embodied; and the covert Logic of the Inclusive occupying the middle place as shown in Figure 1

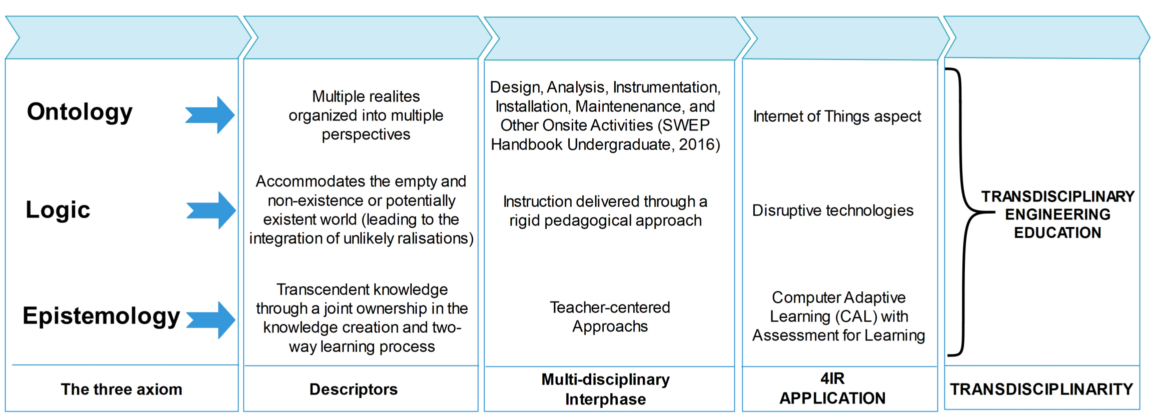

Figure 1: *The axioms of transdisciplinary research framework (Adapted from McGregor, 2014).*

As shown in Figure 1, the original model carefully explained ontology as the integration of multiple perspectives as mediated by logic which represents a fertile space in constant flux where new knowledge can be co-created to elucidate that what appears to be contradictory can temporarily be joined for trans-disciplinary insights and integrated knowledge. The epistemology aspect of the model clearly elucidates trans-disciplinary as alive, open, and always informative leading to an intellectual fusion and integrative strategy. This framework is seen as simplistic and relevant for this study as it has been clearly outlined which aids its expandability and multi-field applicability. The original model terminated at the descriptors level while the resulting model has been extended by adding new layers of multi-disciplinary interphase and 4IR applications as it translates to transdisciplinarity with regards to the students' work experience program aspect of engineering education. More succinctly, the model was adapted in the following respects:

1. Ontology: this concept relates to the conscious integration of multiple perspectives, reported as happening in the quantum space and imminent potentials and possibilities. The openness of this concept makes it adaptable to the field of Engineering education while attempting to move from the disciplinary to multidisciplinary to the transdisciplinary space in the Nigerian context. This submission is at par with the current realities of the 21 Century while leveraging the Internet of Things aspect of the fourth Industrial Revolution (41R) (Stroud, 2020).
2. Logic: this is described as the covert link that provides a fertile space that can evolve rapidly and make room for new co-creation of knowledge during the transition from learning to knowledge following state of the Art technologies. Various fields of studies are witnessing "a genuine transformation" where chaos, process, meaning, complexity, and self-organization are slowly replacing the classic concepts of structure, static, combinatory, and universal (del Cerro Santamaría, 2020). Leveraging disruptive technologies as an aspect of the 4IR is a departure from a rigid pedagogical approach resulting in a transdisciplinary approach in the training of University Engineering graduates.
3. Epistemology: this concept speaks to a system of beliefs encompassing the views and ideologies held by the teachers in terms of what makes up knowledge, how it should be taught (pedagogy), and how it should be learned (Ndlovu, 2019). This aspect of the conceptual framework allows teachers to critically how best to teach the curriculum content to ensure mastery which largely requires teaching and assessment professionalism related to the Computer Adaptive Learning (CAL) with Assessment for Learning Connectivism Framework related to the intelligent tutoring systems deployed via learning management systems (Oladele et al., In Press). This approach is a departure from rote learning to internalizing facts that covertly impact graduates' professional practice.

The forgoing shows a leading relationship that positions 4IR technologies as a major player in the transition from multidisciplinarity to a transdisciplinary domain, making the need for change largely inevitable (Stroud, 2020). With this approach to engineering education, students can be exposed to other disciplines such as project management, cost-benefit analysis and concept mapping, among others directly related to the world of work being is an advantage that transdisciplinarity offers.

3 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, are not a facet leaving the assessor as the only facet in the study (Shavelson & Webb, 1991). There are twelve assessment areas during the SWEP engineering programme. This design is nested as a unique assessor rates persons on each of the twelve assessment areas (Shavelson & Webb, 1991). 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 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.

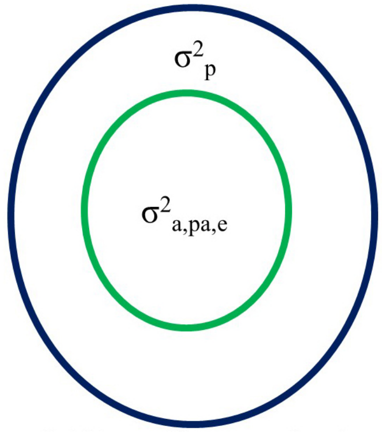

Figure 2: *Venn diagram for (a:p) design.*

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 $\sigma^2_{(a:p)}$ are given in Equation 3.

$$\sigma^2 X_{pa} = \sigma^2_p + \sigma^2_{a,pa,e} \tag{3}$$

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 2.

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; thus, assessor and residual are confounded. 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 the 2015/2016 academic session; a total of 608 and all the technologists who took part in each of the 12 assessments. A total of 591 students assessed in each of the twelve engineering fields were purposely sampled for the study. Purposive sampling ensured that students who had complete scores in the 12 engineering assessment fields participated in the study. Assessment scores were collated using a Proforma, a collation rate of 97%. The data obtained were analyzed using the Analysis of Variance (ANOVA) option for obtaining Variance Components using the GENOVA Programme (VERSION 3.1).

4 Results

Research Question One: *What are the contributions of persons and assessors to the measurement error of University Engineering SWEP assessment scores?*

To answer research question one, determining the contributions of the identified sources to measurement error in university engineering SWEP assessment was based on the estimated variance components and

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)	0.421	607	43.1733	0.04422	0.1
Assessor (a,pa,e)	1.943	6688	493107.8766	96.2506	99.9
Total	**2.364**	**7295**			**100**

percentage of total variance produced by the person (p) and assessor (a, pa,e). The analysis result as given by GENOVA output is shown in Table 1.

As shown in Table 1, the estimated variance component for persons (σ_p^2) is 0.04422, which accounted for 0.1% of the total variance in students' SWEP scores, while the estimated variance component for assessor, confounded with a person by assessor interaction and the residual ($\sigma_{a,pa,e}^2$) is 96.2506 accounted for 99.9% of the total variance. This result shows that the assessor effect confounded with a person by assessor interaction, and the residual ($\sigma_{a,pa,e}^2$) contributed more to measurement error in university engineering SWEP scores.

Research Question Two: *What is the relative error variance of University Engineering SWEP assessment scores?*

The relative 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 Equation 4.

$$\sigma_{Rel}^2 = \sigma_{Abs}^2 = \frac{\sigma_{apa,e}^2}{n_a} \tag{4}$$

Since the residual term is confounded with the assessor effect, the variance for the assessor effect: 96.2506 and the number of assessors: 12 is substituted into Equation 4 and used for the computation as shown in Table 2.

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

Relative Error Variances
8.02

As shown in Table 2, the relative error variance is 8.02.

Research Question Three: *What is the absolute error variance of University Engineering SWEP assessment scores?*

The 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 Equation 4 results the absolute error variance of 8.02.

Research Question Four: *What interpretations can be made from the relative and absolute error variances?*

For relative decisions, variance components that influence the relative standing of individuals contribute to error, while for absolute decisions, all variance components except the object of measurement contribute to measurement error. The relative and absolute variances are shown in Table 3.

As shown in Table 3, the relative error variance is 8.02, while the absolute error variance is 8.02. This result shows that the relative and absolute error variances are the same, and therefore, relative interpretations cannot be distinguished from absolute interpretations.

Table 3: Relative and Absolute Error Variances (One-Facet Nested Design for SWEP)

Relative Error Variances	**Absolute Error Variances**
8.02	8.02

5 Discussion

For this study, the universe of admissible observation for the one-facet nested design employed in the G-study consisted of the variance for two effects (person: σ^2_p and assessor confounded with the person by assessor interaction as well as the residual: $\sigma^2_{a,pa,e}$). The result revealed that the highest contribution to measurement error in university SWEP assessment scores was the residual effect of having the assessor confounded with the person by assessor interaction ($\sigma^2_{a,pa,e}$), with a value of 96.25 accounting for 99% of the total variance. This finding showed that a larger proportion of the variance was due to the interaction of the assessor with the person by the assessor and residual (other unsystematic sources of variance) not measured in this study. The person effect contributed only 0.04, which accounted for 1% of the total variance in University Engineering students' SWEP scores. This finding suggests that the residual effect was the major contributor to measurement error in University Engineering students' SWEP assessment scores. This finding could have resulted from any unmeasured sources of systematic error in the study, strengthening the need to identify all possible sources of measurement error in educational assessments. With this finding, it could be inferred that impediments to quality engineers may not be the curricular but other overlooked aspects such as the synergy between industry and academia, which may require further questioning of values, background assumptions and normative orientations Popa et al., 2015).

This finding aligns with Shavelson, Baxter and Gao (1993), who also reported the residual effect as the major source of measurement error followed by person x task interactions in performance-based assessment. The findings also supported Shavelson and Webb (1991), whose study found the residual as the largest contributor to measurement error. This study was also supported by Egbulefu's (2013) findings, who reported that the residual also made the highest contribution to measurement error. The results of Mahmud's (2017) study equally reported that the largest variance was accounted for by the residual is, as the findings of this study. These findings show that some hidden facets affect the students' scores obtained during the assessment and the need to estimate as many sources of error as feasible in any measurement situation. Despite the twelve unique assessment areas, Shavelson and Webb (2005) noted that it is impossible to ascertain the reliability of the assessments in the one-facet nested design.

This finding deviates from Kassab, Fida, Radwan, Hassan, Abu-Hijleh, and O'Connor (2016), who reported that students' concept map scores (universe scores) accounted for the largest proportion of total variance. Students' performance could have been a result of the remaining unmeasured sources of variation such as an unhealthy person appearing at the period of assessment, limited industrial exposure in training, non-exposure to the current edge technology machines during the programs, poor infrastructural facilities haunting tertiary institutions in Nigerian, large classes among others (Olorunfemi & Ashaolu, 2008; Idris & Rajuddin,2012). Similarly, Brown et al. (2017) reported that the largest variance component in each case was for the trial being rational as each new learning trial should increase from the prior trial due to learning effects among healthy participants. Both studies were cross designs, which adequately explained the results obtained. This finding revealed that the one-facet nested design leaves important questions unanswered, which should be avoided as the goal of any educational assessment is to ensure that scores obtained represent the true performance of the students.

Results also show that the relative and absolute error variances in a one-facet design were the same, typical of nested designs. This finding was so as the error term, also referred to as residual, was indistinguishable in the design, and as such, it was lumped with the assessor effect. This outcome plays down the power of Generalisability statistics of estimating as many sources of variability as possible in measurement (Shavelson & Webb, 2005) and making both relative and absolute decisions (Brennan, 2001a). Considering that the relative and absolute variances are undifferentiated, this design exhibit characteristics of the CTT approach to a measurement error which considers an observed score to include true score and undifferentiated error.

6 Conclusion

Based on the findings of this study, the residual effect was the major contributor to measurement error in University Engineering students' SWEP assessment scores. Other factors aside from the curriculum should be examined to improve the quality of university engineering graduates in Nigeria who are adequately equipped for the world of work.

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

1. The National Universities Commission, a regulatory body in charge of educational sectors at tertiary levels in Nigeria, should map employ a transdisciplinary approach to engineering education to create a synergy between industry and academia to boost university engineering training to ensure the industrial relevance of the various disciplines. With the transdisciplinary approach, an egalitarian, inclusive, and cooperative environment is provided for every team member for quality University engineering education.
2. Generalisability analysis should be deployed by psychometric experts using crossed designs as an innovative technique for identifying all relevant stakeholders for training engineers to enhance a trans-disciplinary approach to improving the quality of university engineering graduates.

Funding: This research received no specific grant from any funding agency.

Conflicts of Interest: The author declares no conflict of interest.

References

[1] Alonge, O., Frattaroli, S., Davey-Rothwell, M., & Baral, S. (2016). A Transdisciplinary Approach for teaching implementation research and practice in public health. *Pedagogy in health promotion,* 2(2), 127-136. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5082840/pdf/nihms785916.pdf

[2] Bernstein, J. H. (2015). Transdisciplinarity: A review of its origins, development, and current issues. *Journal of Research Practice,* 11, (1), 1-20. https://jrp.icaap.org/index.php/jrp/article/view/510

[3] Brandt, S. (2008). Estimation of a Rasch Model Including Sub-dimensions. In M. von Davier and D. Hastedt (eds.), *Issues and Methodologies in Large-Scale Assessments, IEA- ETS Research Institute, Hamburg,* 1, 53-71.

[4] Brennan, R.L. (2001a). **Generalisability Theory.** Springer-Verlag.

[5] Brennan, R.L. (2011). Generalisability Theory and Classical Test Theory. *Applied Measurement in Education,* 24: 1–21. Taylor & Francis Group, LLC. http://dx.doi.org/10.1080/08957347.2011.532417

[6] Brown, F.C., O'Connor, B.P., Vitelli, K.M., Heinly, M., Rommel, G.C. & Davis, R.N. (2017). Comparison of the Computer and Hand Administered Versions of the Brown Location Test (BLT). *Archives of Clinical Neuropsychology;* 1–10.

[7] del Cerro Santamaría, G. (2020). Complexity and transdisciplinarity: The case of iconic urban megaprojects. *Transdisciplinary Journal of Engineering & Science,* 11. http://dx.doi.org/10.22545/2020/0131

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

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

[10] Grossman, P. Research on pedagogical approaches in teacher education. In: Cochran-Smith, M.; Zeichner, K., (Eds.). *Studying teacher education: the report of the AERA panel on research and teacher education.* London: Taylor and Francis; 2009. p. 425-450.

[11] Hofmann, D. (2005). Common Sources of Errors in Measurement Systems. Reproduced from *the Handbook of Measuring System Design.* John Wiley & Sons, Ltd.

[12] Idris, A., & Rajuddin, M. (2012). The trend of engineering education in Nigerian tertiary institutions of learning towards achieving technological development. *Procedia-Social and Behavioral Sciences,* 56, 730-736. https://bit.ly/3KO45T8

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

[14] 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. *Medical Education,* 50, 730–737. http://dx.doi.org/10.1111/medu.13054

[15] Kizlik, B. (2012). Measurement, Assessment, and Evaluation in Education. Available at: http://www.adprima.com/measurement.htm

[16] Lakes, K.D. & Hoyt, W.T. (2009). Applications of Generalisability theory to clinical child and adolescent psychology research. *Journal of Clinical Child and Adolescent Psychology,* 38(1), 144–165. http://dx.doi.org/10.1080/15374410802575461

[17] 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

[18] Margono, G. (2015). "Multi-dimensional Reliability of Instrument for Measuring Students' Attitudes toward Statistics by using Semantic Differential Scale". *American Journal of Educational Research,* 3(1), 49-53.

[19] McCulloch, M. (2007). An Introduction to Assessment. Learning and Teaching Centre, University of Glasgow. Available at: http://www.gla.ac.uk/services/learn

[20] Minor, L.L. (2013). Generalisability and Dependability of Brief Behavior Rating Scales for Social Skills. *An Unpublished PhD Thesis:* Louisiana State University and Agricultural and Mechanical College.

[21] Ndlovu, M. (2019). PGCE in Further Education and Training Teaching: SciTechEd MODULE: Teaching Methodology and Practicum for Further Education and Training (FET) Mathematics *Unpublished Learning Guide, University of Johannesburg.*

[22] Okafor, I. P., Manafa, F., Adeniji, A. O., & Nasiru, I. Innovation and Employability of National Youth Service Corps Members for Sustainable Development in South-East, Nigeria. *The New Educational Review*, 61(3), 131-144. https://czasopisma.marszalek.com.pl/images/pliki/tner/202003/tner6111.pdf

[23] Oladele, J.I., Owolabi, H.O. & O' Connor, B. P. (2020). Generalisability and Dependability of Students' Work Experience Programme Assessment for Quality University Engineering Graduates in North-Central Nigeria. *Journal of Evaluation,* 5, (1) 23-35. http://journal.theasseren.org.ng/index.php/joe/issue/view/15/oladele

[24] Oladele, J. I., Ndlovu, M. and Ayanwale, M. A. (InPress). Implementing Assessment as Learning with Computer Adaptive Learning for Enhancing STEM Education in Africa: A Fourth Industrial Revolution Possibility. *In The Fourth Industrial Revolution (4IR) and Mathematics Education in Africa.* Springer.

[25] Olorunfemi, A. I. & Ashaolu, M. O. (2008). A Pragmatic Approach in Engineering Education Teaching Methods and Industry Partnership. European Society for Engineering Education SEFI Annual Conference, PP1-6, Aalborg, Denmark. https://bit.ly/3vpglTO

[26] O'Neill, D. & Sweetman, O. (2013). The consequences of measurement error when estimating the impact of obesity on income. *IZA Journal of Labor Economics,* 2(3), 21-20. Maynooth: Springer. http://dx.doi.org/10.1186/2193-8997-2-3

[27] Pelton, T.W. (2002). The Accuracy of Unidimensional Measurement Models in the Presence of Deviations from the Underlying Assumptions. *An unpublished doctoral dissertation:* Brigham Young University.

[28] Popa, F., Guillermin, M., & Dedeurwaerdere, T. (2015). A pragmatist approach to transdisciplinarity in sustainability research: From complex systems theory to reflexive science. Futures, 65, 45-56. http://dx.doi.org/10.1016/j.futures.2014.02.002

[29] Portela, M., Alessie, R. & Teulings, C. (2006). *Measurement Error in Education and Growth Regressions.* Germany: IZA discussion paper Series, No. 1165.

[30] McGregor L. T. (2014) Transdisciplinarity and Conceptual Change. World Futures. *The Journal of New Paradigm Research,* 70(3-4), 200-232. http://dx.doi.org/10.1080/02604027.2014.934635

[31] Schochet, P. Z. & Chiang, H. S. (2010). *Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains (NCEE 2010-4004).* National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

[32] Shavelson, R.J. & Webb, N.M. (1991). *Generalisability theory: A primer.* Sage.

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

[34] Stroud, R. (2020). Is Transdisciplinary Education Engaging the 21st Century Engineering Student?. *Transdisciplinary Journal of Engineering & Science,* 11. https://doi.org/10.22545/2020/00138

[35] Tindal, G., Yovanoff, P., &Geller, J.P. (2010). Generalisability theory applied to reading assessments for students with significant cognitive disabilities. *Journal of Special Education,* 44, 1 3-17. http://dx.doi.org/10.1177/0022466908323008

[36] Thornton, A. (2002).*A Primer on 2 and 3-Parameter item response theory models.* Paper presented at the annual meeting of the Educational Research Exchange, University of North Texas: Denton.

About the Author

Jumoke I. Oladele holds a permanent teaching position with the University of Ilorin, Nigeria, and is currently on her postdoctoral research fellowship with the University of Johannesburg, South Africa. Her specialization is in educational evaluation and psychometrics. Her research interests are in item response theory-based psychometrics in educational assessments with computer adaptive testing, applicable to a wide range of subjects focusing on improving standardized assessment in Africa, educational evaluation, and innovation studies. So far, her research has generated a good number of publications in national/international journals. She is the initiator of the Association of Computer Adaptive Testing in Africa, of which she is a member and serves as the research/membership director. She is also a member of other learned societies such as the Teacher's Registration Council of Nigeria, the Association of Educational Researchers and Evaluators of Nigeria, and the International Association for Computerized Adaptive Testing. For her, transdisciplinarity is an evolving area of research interest driven by the need for more holistic views of problem-solving as applicable to pedagogical and assessment practices in higher education. Her contact email is jumokeo@uj.ac.za