WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Using Generalizability Theory to Determine Sources of Errors in SS - II Mathematics Word Problems in Makurdi Metropolis, Benue State - Nigeria

Mr. N.A. Zaria[1], Dr. M.O. Aduloju[2], Dr. (Mrs.) C.I. Agi[3],

[1]Department of Educational Foundation and General Studies, College of Agricultural and Science Education, Federal University of Agriculture, Makurdi
[2]Department of Educational Foundation and General Studies, College of Agricultural and Science Education, Federal University of Agriculture, Makurdi
[3]Department of Educational Foundation and General Studies, College of Agricultural and Science Education, Federal University of Agriculture, Makurdi

ABSTRACT

This study was conducted to find out the use Generalizability theory to determine sources of errors in SS - II Mathematics word problems in Makurdi Metropolis. Specifically, the purpose of this study was to identify the sources of errors, estimate the magnitude of contribution of the identified sources of errors and determine the interaction effect of sources of errors in in SS - II Mathematics word problems using Generalizability Theory analysis. Three research questions were formulated to guide the study. The study adopted the two-facet/multi-facet research design where the two facet design reflecting the universe of admissible observations with two facets was of such as persons (p) x raters (r) or ratings (r') x occasion (o) was used. The population of the study was 9,364 Senior Secondary School Students from 137 Government approved public and private secondary schools in Makurdi Metropolis. The sample size used for this study comprised of 30 Senior Secondary School (SS - II) students drawn from 2 Government approved secondary schools within Makurdi Metropolis (Government College Secondary School and Community Secondary school). A multi-faceted sampling technique was adopted to draw the facets under study. The instrument that used for this study was Mathematics Word Problems Test (MWPT) with 0.83 correlation coefficient of equivalence. Test scores were analyzed using Generalizability Theory analysis under EDU-G software and ANOVA statistic. Findings made from the study revealed that raters and occasions were identified as sources of measurement error, identified sources of error jointly contributed minimally to the error variance whereas, occasion facet had the highest magnitude of error contribution on the error variance and; the difference in scoring of persons by raters was very small; the rate at which raters award scores across the two occasions was completely not different and that the ranking of persons by raters on the two occasions was average. Based on the findings made by this study and the conclusions reached, it was recommended amongst others that the occasion/situation in which a test is administered can determine whether the performance of test takers will be enhanced or not; thus, Mathematics teachers should ensure that the occasions for administering test items on Mathematics word problems are conducive.

Introduction

Measurement is one of the ways of ensuring that appropriate scores are assigned to observations with given set of rules. The process of carrying out this assignment of scores is said to constitute some degree of error. Errors are therefore said to be inevitable in measurement procedures. Since errors cannot be completely eliminated, minimizing errors have been made a priority by measurement experts. Identifying sources of measurement errors is the first step in minimization of errors. Sources of errors include respondents, raters, the instrument/items or the occasion.

Errors from respondents may be significant in distorting the validity and reliability of any measurement. Robert, Thorndike and Hagen in Kolen and Brennan (2013) stressed that on many occasions, the respondents may be reluctant to express strong negative feelings or it is just possible that they may have very little knowledge but may not admit their ignorance. This reluctance is likely to result in an interview of 'guesses.' Transient factors like fatigue, boredom, anxiety, among others may limit the ability of the respondent to respond accurately and fully. This may interfere with the validity and reliability of inference made as it may not give a true picture of the respondents' abilities.

In a similar view, Bevington, Phillip and Robinson in Advanced Instructional Systems, Incorporated (2011) submitted that situational factors (occasion) may also come in the way of correct measurement. Any condition which places a strain on the interview can have serious effects on the interview-interviewer and interviewer-respondent rapport. Certain situations surrounding the interviewer may make him to misinterpret the actual information required by the interview. Also, if someone else is present, he can distort responses by joining in or merely by being present. If the respondent feels that anonymity is not assured, he may be reluctant to express certain feelings.

Errors may also emanate from measurement procedures for instance, through the rater as a result of incorrect coding, faulty tabulation and/or statistical calculations, particularly in the data-analysis stage. In addition, rater errors may arise due to the personality features of the rater. This is most

true for essay test items where rater discretion is required in scoring since the testee has liberty of expression, higher expertise may be required. The inability of the rater to judge accurately and award appropriate scores for methods employed, concepts or expressions used by the testee may introduce errors. In the same way, the measuring instrument may introduce errors in measurement by means of poor printing, inadequate space for replies, response choice omissions among others and these may make the measuring instrument defective and may result in measurement errors.

In an attempt to minimize measurement errors, test experts investigated test procedures and the degree of errors introduced. This led to the emergence of a theoretical framework known as Classical Test Theory (CTT). Classical Test Theory holds that a person's observed score (X) on an item (*i*) is assumed to comprise both a true score (X_t) and an error component (X_e). This is often expressed mathematically as $X_i = X_t \pm X_e$. The true score reflects whether the examinee's amount of knowledge or ability is the true measurement of the examinee which is always contaminated by random errors. The observed score is often called a fallible score because of the error contaminant. The true score is the score that would have been obtained if there were no errors in measurement.

However, the weaknesses of CTT brought about the introduction of Generalizability Theory (G-Theory) by Cronbach in 2004 as an advancement of CTT. These weaknesses includes the fact that the error component of an observed score is said to be an undifferentiated random variable which cannot be disentangled. Also, that sources of errors could not be identified. G-theory is seen as a statistical theory that is used for the evaluation of the reliability or dependability of behavioural measurements. Being an advancement of the CTT otherwise referred to as True Score Theory, G-Theory posits that the error component is said to be random variation which can be disentangled (Webb & Shavelson, 2005). In other to have the opportunity of disentangling the error components of measurements, Cronbach, Gleser, Nanda and Rajaratnam in Gao, Brennan and Guo (2015) propounded a Psychometric Theory that is able to identify the major sources of error to behavioural measurements and pinpoints the magnitude of variability introduced by each of the sources of error in the measurement.

In the opinion of Brennan in Faleye (2018), sources of variation are referred to as facets. These facets may include persons, raters, items/forms, time, and settings among other possibilities. These facets are potential sources of error and the purpose of G-theory is to identify the sources and quantify the amount of error caused by each facet and interaction of facets. This study seeks to use G-Theory to estimate the facets of measurement errors inherent in Mathematics word problem achievement test.

Cheng (2017) viewed Mathematics word problem as a situation where a mathematical model can be applied to represent the quantities and relations present in the text and to find a solution to the given question. Many students seem to find word problems challenging due to their intellectual abilities, the teacher's capabilities as well as the conditions that surrounds the testing situation (Mukunthan, 2013). Karunanithy (2015) observed that students were very weak in solving word problems which is a pointer to the poor performances experienced in Mathematics.

This study was however carried out due to the poor performance of students in Mathematics and word problems in particular. Also, the chief Examiners' Report in West African Examination Council (WAEC) which suggested that candidates were unable to translate word problems into Mathematical Equations". The inability of candidates to translate and solve word problems accurately. Inability of behavioural measurements to determine errors emanating from measurement facets of person, rater, item and occasion in mathematics word problems. Difficulty in determining sources of errors and disentangling the error component found in Mathematics word problems also formed the reason for embarking on this study. It is in an attempt to solve the problems stated above that the researcher sought to use Generalizability Theory as a determinant of sources of errors in Mathematics word problems in Makurdi Metropolis, Nigeria.

Purpose of the Study

The purpose of this study is to use Generalizability Theory to determine sources of errors in SS - II Mathematics word problems in Makurdi Metropolis. Specifically, this study seeks to:

(i) identify the sources of errors in SS - II Mathematics word problems using Generalizability Theory.

(ii) find out the magnitude of identified sources of errors in SS - II Mathematics word problems using Generalizability Theory

1.4 Research Questions

The study is guided by the following questions:

(i) What are the sources of errors in SS - II Mathematics word problems estimated by Generalizability Theory analysis?

(ii) What is the magnitude of identified sources of errors of in SS - II Mathematics word problems estimated by Generalizability Theory analysis?

Research Method

The study adopted the two-facet/multi-facet research design where the two facet design reflecting the universe of admissible observations with two facets was of such as persons (*p*) x raters (*r*) or ratings (*r*') x occasion (*o*) was used. The population of the study was 9,364 Senior Secondary School Students from 137 Government approved public and private secondary schools in Makurdi Metropolis (***Source:*** Benue State Ministry of Education, 2020). The

sample size used for this study comprised of 30 Senior Secondary School (SS - II) students drawn from 2 Government approved secondary schools within Makurdi Metropolis (Government College Secondary School and Community Secondary school). A multi-faceted sampling technique was adopted to draw the facets under study. The instrument that used for this study was Mathematics Word Problems Test (MWPT) with 0.83 correlation coefficient of equivalence. Test scores were analyzed using Generalizability Theory analysis under EDU-G software and ANOVA statistic.

Analysis and Results

Analysis of data for the study was done using mean and standard deviation for answering of the respective research questions.

Research Question One: What are the sources of errors in SS - II Mathematics word problems estimated by Generalizability Theory analysis?

Table 4.1: Generalizability Analysis of the Sources of Errors in SS - II Mathematics Word Problems in Makurdi Metropolis

				Components				
Source of Variance	**SS**	**Df**	**MS**	**Random**	**Mixed**	**Corrected**	**%**	**SE**
Persons (p)	1037.30000	29	35.76897	2.85057	8.80460	8.80460	40.6	2.74776
Occasions (o)	45.63333	1	45.63333	0.35862	0.35862	0.17931	0.8	0.62950
Raters (r)	0.53333	1	0.53333	0.00345	-0.00029	-0.00029	0.3	0.00893
Persons X Occasions (po)	705.86667	29	24.34023	11.90805	11.90805	11.90805	51.9	3.09193
Persons X Raters (pr)	15.96667	29	0.55057	0.01322	0.27529	0.27529	1.3	0.09654
Occasions X Raters (or)	0.30000	1	0.30000	-0.00747	-0.00747	-0.00747	2.7	0.00929
Persons X Occasions X Raters (por)	15.20000	29	0.52414	0.52414	0.52414	0.52414	2.4	0.13313
Total	**1820.80000**	**119**					**100%**	

Table 4.1 reveals the Generalizability analysis of the sources of measurement errors estimated in SS - II Mathematics word problems in Makurdi Metropolis. Result from the table indicates that persons (p) had 40.6%, occasions (o) had 0.8% and raters (r) had 0.3%, persons and occasions had 51.9%, persons and raters had 1.3%, Occasion and raters had 2.7% and persons, occasions and raters jointly had 2.4%. Thus, persons and occasions were identified as main sources of measurement error (where raters and occasions have been randomly selected). Persons (p) were not considered a source of measurement error (source of variance) since it was considered the object of measurement. This is because, variance arising from the object of measurement taking part in the research is the desired condition showing the differences inherent in persons. As such, it is not taken as a "source of error" (facet) in G theory as it is in CTT. The table also indicates that interaction between persons and occasions (po), persons and raters (pr), raters and occasions (ro) and persons, raters and occasions (pro) were automatically identified as sources of error. This is because, each probable source of error lying outside the object of measurement and having similarities is defined as a facet. It therefore implies that raters, occasions and the interaction between persons and occasions, persons and raters, raters and occasions and persons, raters and occasions were identified as sources of measurement error among SS - II Mathematics test takers in Makurdi Metropolis.

4.1.2 Research Question Two:

What is the magnitude of error contribution of identified sources of errors of in SS - II Mathematics word problems estimated by Generalizability Theory analysis?

Table 4.2: Generalizability Analysis of the Magnitude of error Contribution of Identified Sources of Error in SS - II Mathematics Word Problems in Makurdi Metropolis

				Components				
Source of Variance	**SS**	**df**	**MS**	**Random**	**Mixed**	**Corrected**	**%**	**SE**
Persons (p)	1037.30000	29	35.76897	2.85057	8.80460	8.80460	40.6	2.74776
Occasions (o)	45.63333	1	45.63333	0.35862	0.35862	0.17931	0.8	0.62950
Raters (r)	0.53333	1	0.53333	0.00345	-0.00029	-0.00029	0.0	0.00893

Persons X Occasions (po)	705.86667	29	24.34023	11.90805	11.90805	11.90805	54.9	3.09193
Persons X Raters (pr)	15.96667	29	0.55057	0.01322	0.27529	0.27529	1.3	0.09654
Occasions X Raters (or)	0.30000	1	0.30000	-0.00747	-0.00747	-0.00747	0.0	0.00929
Persons X Occasions X Raters (por)	15.20000	29	0.52414	0.52414	0.52414	0.52414	2.4	0.13313
Total	**1820.80000**	**119**					**100%**	

Table 4.2 reveals the Generalizability Analysis of the magnitude of errors contributed by identified sources of error in SS - II Mathematics word problems in Makurdi Metropolis. Result from the table under the column "%" reveals the proportion of the variance of individual scores (estimated as the sum of the corrected components) that is attributed to each variance source. This shows that persons source of variance had 40.6%, occasions source of variance had 0.8%, raters source of variance had 0.0%, persons and occasions' interaction yielded 54.9%, persons and raters' interaction produced 1.3%, occasions and raters' interaction had 0.0% while persons, occasions and raters' interaction yielded 2.4%.This shows that the major source of variance (error) was the occasion facet with the highest percentage of 0.8% and the interaction between persons and occasions yielded the highest degree of contribution. Although, since Chronbach advised that these percentages should not be interpreted as directly reflecting the relative importance of each variance source (as real life decisions are generally made on the basis of total scores or mean scores, and not on the basis of non-summarized data points), the Generalizability Study Table would be deployed to estimate the relative importance and contribution of each source of error in the total error variance.

Table 4.3: Generalizability Study Table of the Magnitude of Error Contribution of Identified Sources of Error in the Total Error Variance

Source of Variance	**Differentiation Variance**	**Source of Variance**	**Relative Error Variance**	**% Relative**	**Absolute Error Variance**	**% Absolute**
P	8.80460		.....		.....	
	.....	o	.....		(0.00000)	0.0
	.....	r	.....		(0.00000)	0.0
	.....	po	(0.00000)	0.0	(0.00000)	0.0
	.....	pr	0.13764	100.0	0.13764	100.0
	.....	ro	.....		(0.00000)	0.0
	.....	Pro	(0.00000)	0.0	(0.00000)	0.0
Sum of Variances	8.80460		0.13764	100%	0.13764	100%
Standard Deviation	2.96725		Relative SE: 0.37100		Absolute SE: 0.37100	
Coef_G relative	0.98					
Coef_G absolute	0.98					

Table 4.3 shows the Generalizability Study Table of the magnitude of error contribution of identified sources of error in the total error variance. The table reveals how the two columns entitled "Source of Variance" display how the sources of variance are divided according to the measurement design: they may contribute to the "true" (i.e. differentiation) variance or to the (relative or absolute) "error" variance. The magnitude of contribution of Sources of variance shown on row of column totals "Sum of Variances" reveals that the variance of differentiation was 8.80460 while the magnitude of contribution of each source of variance on the variance of relative errors was 100% and on the variance of absolute errors was 100%. Therefore, the total magnitude of contribution of the sources of variance on "true variance" (person) was 8.80460 and the magnitude of contribution of each source of variance on "error variance" was 100%. This implies that facets of occasions, raters, persons and occasions, persons and raters, occasions and raters; and persons, raters and occasions all together had 100% contribution to the error variance.

Discussion of Findings

Based on the findings of this study, the following discussions were made:

Raters and occasions were identified as sources of measurement error among SS - II Mathematics test takers in Makurdi Metropolis. This finding supports an earlier finding by Faleye (2018) who identified the sources of error to include raters and occasions in a study that sought to obtain the variance components of G and D coefficients using the $p \times r \times o$ (person by raters by occasion) design. This current finding is supported by the finding of Gao et al (2015) who researched on modeling measurement facets and assessing Generalizability in a Large-Scale Writing Assessment. Gao et al (2015) identified rater, item and occasion errors as sources of variation in the measurement process. On the contrary, this finding negates that of Boyd et al (2008) who carried out a study on overview of measuring effect sizes: the effect of measurement error and identified no facet in particular as a source of error. Their study rather established that the source of variation in measurement were attributed to measurement error without identifying the actual sources of error.

Also, the findings revealed that the major source of variance (error) was the occasion facet with the highest percentage of 0.8% and the interaction between persons and occasions yielded the highest degree of contribution. Further analysis reveals that the identified sources of error (facets of occasions, raters, persons and occasions, persons and raters, occasions and raters; and persons, raters and occasions) jointly contributed to the error variance whereas, occasion facet had the highest magnitude of error contribution on the error variance. This finding agrees with that of Faleye (2018) whose study revealed that the person source of error majorly contributed to the error component. That is to say, the person facet had the highest Estimated Error Component (EVC) and also, the magnitude of error contribution between persons and occasions was ranked high. This finding is also supported by that of Kreiter, et al (2016) who found that the variance component (VC) related to occasion was small relative to the VC associated with rater. The current finding disagreed with that of Gao et al (2015) whose finding revealed that the variance in student test scores was attributable to test measurement error estimate, not stating in particular the degree of contribution of the sources of errors.

Conclusion

Based on the findings, it was concluded that measurement is not perfect or devoid of error. These errors arise from occasions and raters which jointly contribute to the error component. Therefore, measurement experts and teachers should ensure that the occasions in which tests are administered are conducive for test takers. Also, essay items should be subjected to multiple ratings (r'') in order to help control rater errors.

Recommendations

Based on the findings made by this study and the conclusions reached, the following recommendations were made:

(i) The occasion/situation in which a test is administered can determine whether the performance of test takers will be enhanced or not. As such, Mathematics teachers should ensure that the occasions for administering test items of Mathematics word problems are conducive.

(ii) Test administrators/teachers should endeavor to prepare the minds of students before they are administered any test. This is because the occasion surrounding impromptu test may give rise to errors in the measurement process.

(iii) Students should endeavor to liaise with teachers and test administrators for a conducive occasion for tests to be takes since minimization of errors depends largely on the occasion in which the test is taken.

References

Advanced Instructional Systems, Inc. (2011). *Measurement and Error Analysis*. University of North Carolina.

Cheng, L.P. (2017). *Error Analysis for Arithmetic Word Problems – A Case Study of Primary Three Students in One Singapore School.* Singapore: Nanyang Walk.

Cronbach, L.J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, 64, 391–418.

Faleye, B. A. (2018). Generalizability Theory: Being a Conference Workshop Notes at the 2018 Annual Conference of the Association for Educational Researchers and Evaluators of Nigeria (ASSEREN) held at the Federal College of Education, Zuba, Abuja from Monday 9th July to Friday 13th July, 2018.

Gao, X., Brennan, R.L. & Guo, F. (2015). Modeling Measurement Facets and Assessing Generalizability in a Large-Scale Writing Assessment. *Graduate Management Admission Council (GMAC) Research Reports*: RR-15-01

Kolen, M.J., & Brennan, R.L. (2013). *Statistics for Social and Behavioural Sciences: Test Equating, Scaling and Linking (Methods and Practices-Third Edition).* Springer: University of Iowa, Iowa City, IA USA.

Kreiter, C.D. Wilson, A.B., Humbert, A.J. & Wade, P.A. (2016). Examining rater and occasion influences in observational assessments obtained from within the clinical environment; *Medical Education Online Journal*, 21(1): 1-5, DOI: 10.3402/meo.v21.29279

Karunanithy, M (2015). National Assessment of Achievement of Grade 04 Students in Sri Lanka, National Education Research and Evaluation Center (NEREC) of the Faculty of Education, University of Colombo, Sri Lanka

Mukunthan, T. (2013). A study on students' errors on word Problem. *International Journal of Management, IT and Engineering*: 3(10); 205-214

Webb, N.M. & Shavelson, R.J. (2005). Generalizability Theory: Overview. In Everitt, B.S. & Howell, D.C.(Eds.)(717-719); *Encyclopedia of Statistics in Behavioral Science.* Chichester: John Wiley & Sons, Ltd.